Metatranscriptomic analysis of metal contaminated soils for isolation of metal tolerant genes

A Thesis

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

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

By:
Bharti Thakur
(901300010)

Under the supervision of
Dr. M. Sudhakara Reddy
(Professor)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Department of Biotechnology
Thapar Institute of Engineering & Technology, Patiala-147004
Punjab (India)
February, 2021

CERTIFICATE

Certified that the thesis "**Metatranscriptomic analysis of metal contaminated soils for isolation of metal tolerant genes**" which is submitted by **Mrs. Bharti Thakur**, in fulfilment 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 her 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.

M Sudhakar R

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

DECLARATION

I hereby declare that the work presented in the thesis entitled "**Metatranscriptomic analysis of metal contaminated soils for isolation of metal tolerant genes**" in the fulfilment 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 2014 to June 2020, under the supervision of **Dr. M. Sudhakara Reddy**, Professor, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India. 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.

Bharti Thakur

Place: Patiala
Date:

Bharti Thakur
Department of Biotechnology
Thapar Institute of Engineering and Technology
Patiala, Punjab, India.

Dedicated to my parents

for raising me to believe that nothing is impossible

And to my husband & Yashi

for making everything possible……

ACKNOWLEDGEMENT

This thesis is the culmination of my journey of Ph.D which was just like climbing a high peak step by step accompanied with encouragement, hardship, trust, and frustration. When I found myself at top experiencing the feeling of fulfillment, I realized though only my name appears on the cover of this dissertation, a great many people including my family members, well-wishers, my friends, colleagues and various institutions have contributed to accomplish this huge task.

First and foremost, my utmost gratitude towards my esteemed supervisor ***Dr. M. Sudhakara Reddy, Professor & Head,*** *Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala. I appreciate his contributions of time and ideas to make my work productive and stimulating. His valuable suggestions, comments and guidance encourage me to learn more day by day. I am also indebted towards him for his generosity, selfless support and especially for the excellent example and patience that he has provided to me during my Ph.D. Big thanks once again goes to him because without him, this work would have never seen the light as it is today.*

I am extremely grateful to ***Dr. Prakash Gopalan, Director****, Thapar Institute of Engineering and Technology, Patiala for providing the facilities for completion of my research work. I express my gratitude to* ***Prof. Rafat Siddique,*** *Dean (Research and Sponsored Projects), Thapar Institute of Engineering and Technology, Patiala for his support and encouragement during my research work.*

It is my privilege to express my sincere gratitude to ***Prof Anil Kumar Dutta,*** *Coordinator, TIFAC-CORE, Thapar Institute of Engineering and Technology, Patiala, for his eternal support and advices. I an extremely thankful to my Ph. D. committee members* ***Prof Moushami Ghosh and Asst. Prof. Vikas Handa,*** *Department of Biotechnology and* ***Prof. Amjadi Ali****, Department of Chemistry, Thapar Institute of Engineering and Technology, Patiala for provided their valuable suggestions to improve my research. My sincere regards to all the faculty members of* *the Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala.*

I, especially thank ***Dr. Laurence Frassinet-Tachet*** *and* ***Dr. Roland Marmeisse*** *University Lyon1, France, for their valuable time, support and guidance which made my thesis work possible. My sincere gratitude is reserved for* ***Dr Prashanth Suravajhala,*** *BISR, Jaipur and* ***Dr. M. Vasundhara Reddy*** *for their invaluable insights and suggestions. I really appreciate their willingness to help me whenever I need some clarification.*

I am also indebted to my friends ***Arkadeep, Saloni, Sumit, Shikha, Richa and Tanveer,*** *not only for all their help and useful "chai time" suggestions but also for being there to listen when I needed an ear. I will miss those random scooter rides to markets and new places in the city. With you all, attending a conference was so much fun. Many thanks to all the research scholars especially* ***Navneet****,* ***Fatima, Eetika and Anu*** *for their kind help and support. I wish to thank* ***Anku*** *and* ***Heena*** *for their help during thesis writing part.*

I, also thank all the departmental staff and laboratory staff of TIFAC-CORE and Department of Biotechnology for helping me during these years of my study both academically and officially. My sincere thanks to ***Soni ji*** *for his cooperation and help throughout my Ph.D.*

Finally, I acknowledge the people who mean a lot to me, my parents, ***Dr. Hans Raj Mehta*** *and* ***Mrs. Ompati Mehta*** *for showing faith in me and giving me liberty to choose what I desired. I salute you both for the selfless love, care, pain and sacrifice you did to shape my life. I would never be able to pay back the love and affection showered upon me and my daughter* ***Yashaswi*** *by my parents. Also I, express my thanks to my brother* ***Lakhwinder Singh*** *for always driving me to the university and for his warm hugs whenever I felt low and sisters* ***Preeti*** *and* ***Litesh*** *for their valuable prayers, selfless love, care, dedicated efforts and motivations which contributed a lot for completion of my thesis.*

I owe thanks to a very special person, my husband, ***Ravish*** *for his continued and unfailing love, support and understanding during my pursuit of Ph.D degree that* *made the completion of thesis possible. You were always around at times I thought that it is impossible to continue, you helped me to keep things in perspective. I, greatly value his contribution and deeply appreciate his belief in me. I appreciate my baby, my little girl* ***Yashaswi*** *for abiding my ignorance and the patience she showed during my thesis writing. Words would never say how grateful I am to both of you. I consider myself the luckiest in the world to have such a lovely and caring family, standing beside me with their love and unconditional support.*

My heartfelt regard goes to my father in law ***Sh. Desh Raj Thakur****, mother in law* ***Smt. Alkesh*** *for their love and moral support. Although you hardly understood what I researched on, you were willing to support any decision I made. I also thank my sisters in law* ***Mrs*****.** ***Deepika*** *and* ***Dr*****.** ***Sandeepika*** *for their valuable support.*

I thank the Almighty for giving me the strength and patience to work through all these years so that today I can stand proudly with my head held high.

Bharti Thakur

LIST OF PUBLICATIONS

The following publications in peer reviewed journals are the outcome of the present research work:

- **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. *Journal of microbiological methods*, *152*, 119-125.
- **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. *Science of The Total Environment*, *661*, 432-440.
- **Thakur, B**., Yadav, R., Mukherjee, A., Fraissinet-Tachet, L., Marmeisse, R., & Reddy, M. S. (2021). Protection from metal toxicity by Hsp40-like protein isolated from contaminated soil using functional metagenomic approach. *Environmental Science Pollution Research.*

Conferences

- **Thakur, B.** & Reddy, M. S. (2016). A multi-metal resistant eukaryotic environmental transcript *ycf*1Vit39 from contaminated site. National Conference on Fungal Biotechnology & 43rd Annual Meeting of The Mycological Society of

India. November 16-18, 2016, BISR, Jaipur, India. (M.J. Thirumalachar Young Scientist Award).

Thakur, B. & Reddy, M. S. (2017). Soil metatranscriptomics unveils ubiquitin as multi-metal resistant environmental eukaryotic transcript from contaminated site. International Conference on “Emerging Trends in Biotechnology for Waste Conversion (ETBWC – 2017)” to be held on 8th -10th October 2017, XIV Annual Convention of Biotech Research Society India at CSIR-NEERI, Nagpur.

CONTENTS

Contents	**Page No.**
List of Figures	i-x
List of Tables	xi-xii
Abbreviations and symbols	xiii-xiv
Abstract	xv-xvi
Chapter 1: Introduction	1-9
1.1. General Introduction	1-8
1.2. Aim & Objectives of the study	8-9
Chapter 2: Review of Literature	10-73
2.1. Soil	10-11
2.1.1. Soil Habitat and organisms	11-15
2.2. Metals: Classification and potentially toxic members	15-19
2.2.1. Mode of action of potentially toxic metals	19-22
2.2.2. Sources of metal pollution in soil	23-31
2.2.2.1. Fertilizers and agro-chemicals	24-26
2.2.2.2. Municipal wastes, animal and agro-food wastes, sewages sludge and effluents	26-28
2.2.2.3. Application of manures and other animal wastes	28-29
2.2.2.4. Soil irrigation with natural freshwaters and wastewaters	29-30
2.2.2.5.Atmospheric deposition	30-31
2.2.3. Cellular mechanisms involved in metal tolerance	31-42
2.2.3.1. Organic acids and amino acids	32-33
2.2.3.2. Metallothioneins	33-36
2.2.3.3. Phytochelatins	36-38
2.2.3.4. Heat shock proteins	38-41
2.2.3.5. Cell wall and the root exudates	41-42
2.3. Approaches used to study soil microbiome	42-54
2.3.1 Culture dependent approaches	42-43
2.3.2 Culture independent approaches	43-54

2.3.2.1. Genetic fingerprinting 43-45
2.3.2.2. Metagenomics 45-49
2.3.2.3. Metatranscriptomics 49-54
2.4. Characterization of a metatranscriptome 54-62
2.4.1. Isolation and processing of mRNA 54-55
2.4.2. Computational analysis of metatranscriptomic data 55-56
2.5. Metatranscriptomic approach to explore eukaryotic lineages 57-62
2.6. Methodological challenges in metatranscriptomics 63-66
2.6.1. Efficient mRNA extraction protocol and mRNA instability 63-64
2.6.2. Enrichment of mRNA 64-65
2.6.3. cDNA synthesis and amplification 65-66
2.7. Ecologically significant genes procured via metatranscriptomic approach 66-69
2.7.1. Organic matter decomposing enzymes 66-67
2.7.2. Aromatic and xenobiotic degrading genes 67-68
2.7.3. Heavy metal tolerant genes 69
2.8. Metatranscriptomics of soil 69-71
2.9. Summary of literature review 71-73
Chapter 3: Materials and Methods 74-114
3.1. Biological materials and culture conditions 74-75
3.2. Soil sampling 75-76
3.3 Physicochemical properties of metal polluted soil 76-82
3.3.1 Total organic carbon (Walkley, 1947) 76-78
3.3.2 Total nitrogen (Kjeldahl method) (Piper, 1966) 78-79
3.3.3 Total phosphorous (Kitson and Mellon, 1944) and metal estimation 79-81
3.3.4. Available Phosphorous (Olsen, 1954) 81-82
3.4. Total RNA extraction 83-84
3.5. Metagenomic DNA isolation 84-85
3.6. Qualitative and quantitative analysis of nucleic acids 85-86
3.6.1. Non denaturing agarose gel electrophoresis 85
3.6.2. Formamide agarose gel electrophoresis 85-86
3.6.3. Quantification of nucleic acids 86
3.7. cDNA synthesis and size fractionation 86-87

3.8. Size fractionation and validations of cDNA fractions 87-90

3.9. cDNA cloning 90-92

3.10. Plasmid isolation 92-93

3.11. Colony PCR 93-94

3.12. Yeast Transformation 94-97

3.13. 5-Flouroorotic acid (5-FOA) test 97-98

3.14. Yeast complementation assay 98-99

  3.14.1. Drop out assay 98-99

3.15. Yeast Plasmid isolation 100-101

3.16. Bacterial plasmid Isolation 101

3.17. Bioinformatic analysis 102

3.18. Growth assay 103

3.19. Quantitative reverse transcription PCR 104-105

3.20. Inductively coupled plasma mass spectrophotometry 105

3.21. Functional diversity 107-112

  3.21.1. Nucleic acid extraction and library preparation 107-110

  3.21.2. NGS Analysis 110-112

  3.21.3. Analysing diversity and phylotypes 112

3.22. Statistical analysis 112

Chapter 4: Results and Discussion 115-253

4.1.Construction of cDNA libraries from RNA isolated from metal polluted soil 115-130

  4.1.1. Soil sampling and analysis 115-116

  4.1.2. RNA isolation and cDNA synthesis 117-120

  4.1.3. Synthesis of metatranscriptomic cDNA libraries 121-130

    4.1.3.1. Synthesis of cDNA and size fractionation 121-125

    4.1.3.2. cDNA library synthesis 125-130

4.2. Heterologous expression of metal tolerant genes by yeast functional complementation 130-154

  4.2.1. Drop Assay 132-135

  4.2.2. 5-Flouroorotic acid Test 136-140

  4.2.3. Yeast Colony PCR 140-142

  4.2.4. Plasmid extraction and sequence analysis 149-154

4.3. Characterization of cadmium tolerant cDNAs 155-229
4.3.1. PLBe1/Ubiquitin fusion protein 155-174
4.3.1.1 Sequence analysis 156-162
4.3.1.2 Relative quantification of cDNA PLBe1 163-164
4.2.1.3. Multi-metal tolerance of PLBe1 164-169
4.3.1.4. Yeast metal uptake by cDNA PLBe1 169-174
4.3.2. PLCe10/ von Willebrand factor type D Domain 174-193
4.3.2.1. Sequence analysis 174-181
4.3.2.2. qPCR analysis 182-183
4.3.2.3. Multi-metal tolerance of PLCe10 183-188
4.3.2.4. Yeast metal uptake of PLCe10 188-193
4.3.3. PLCc43/Heat shock protein type I like protein 193-212
4.3.3.1. Sequence analysis of PLCc43 193-201
4.3.3.2. Relative quantification of cDNA PLCc43 201-203
4.3.3.3. Tolerance to metal toxicity and metal content 203-207
4.3.3.4. Yeast metal uptake by cDNA PLCc43 207-212
4.3.4. PLCd43/ γ-Glutamylcysteinyl transferase 212-229
4.3.4.1. Sequence analysis 212-219
4.3.4.2. Induction of cDNA PLCd43 220-221
4.3.4.3. Multi-metal tolerance profiling of PLCd43 221-225
4.3.4.4. Yeast metal uptake by cDNA PLCd43 225-229
4.4. Functional eukaryotic diversity from metal contaminated soil 229-250
4.4.1.Amplification of V4 region of 18S rRNA/DNA 230-231
4.4.2.Data analysis 231-250
4.4.2.1. Quality analysis of Reads: FastQC 231-241
4.4.2.2. Data analysis via Mothur 242-250
4.5.Conclusion 250-253
Chapter 5: Summary 254-261
References 262-317
Appendix I 318
Appendix II 319-320
Appendix III 321-328
Publications

LIST OF FIGURES

Fig. No.	Figure Title	Page
2.1	Metals of periodic table classified as Class A (black), Class B (dark grey) and borderline (light grey) where Fe may belongs to either Class A metals or borderline depending upon its oxidation state i.e. Fe^{3+} or Fe^{2+} while Cu and Pb may be Class B or borderline depending upon whether Cu^{+} or Cu^{2+} in case of Cu and Pb^{2+} or Pb^{4+}. s-block, p-block, d-block, lanthanide and actinide series of periodic table are also shown.	16
2.2	Dose-response curves for essential and non-essential metals. Curve for essential metals showed deficiency at low and sub optimal concentration, tolerance at optimal concentration and toxicity at higher concentration while for non essential metals severe toxicity was observed after a threshold concentration of metals (critical concentration).	21
2.3	Point sources of metal contamination in the soil: Natural as well as anthropogenic factors or activities such as weathering of rocks, atmospheric deposition, industrial effluents, municipal waste etc., contaminate the soil environment with potentially toxic metals	24
2.4	Cellular homeostasis maintained by metallothioneins (MTs). Potentially toxic metals (PTMs shown in colored dots) when entered into the cell resulted into production of reactive oxygen species (ROS) which further results into induction of metallothioneins (MTs). PTMs bind to sulfhydryl group of MTs to form a MT-PTM complex and maintain redox status by reducing the availability of diffusible forms of these metals within cells and therefore decreasing their toxic potential.	35
2.5	Metal detoxification by phytochelatins (PC) in a plant cell. Potentially toxic metals (PTM shown in black dots) activates phytochelatin synthase (PCS) which further results into formation of PTM-PC complexes that are consequently transported through tonoplast to the vacuole by ATP-binding-cassette (ABC) and ABC-P1B transporter. PTM also transported inside the cell through plasma membrane via ZIP and vacuolar transport of PTM is mediated by natural resistance associated macrophage protein (NRAMP).	37

2.6 Network of heat shock proteins (Hsps) active in response to abiotic stress. Adverse conditions such as metal toxicity resulted into synthesis of misfolded proteins inside a cell. As a consequence, Hsps dissociate themselves from heat shock transcription factor (HSF), released activated HSF then bind to heat shock element (HSE) and lead to up-regulation of Hsps in the cell which will aid in reforming these misfolded proteins (Rowley *et al.* 2011). 40

2.7 Schematic representation of main stages involved in single cell genomics, metagenomic and metatranscriptomic analyses of soil samples Single cell genomics retrieved the sequence information from individual cells while metagenomics provides a snapshot of the genetic composition of microbial communities in an environment and metatranscriptomics provides an insight into community level transcriptional profiling of an environment (Prosser, 2015). 51

3.1 Schematic outline of Mint-2 cDNA synthesis and bi-directional size fractionation. 88

3.2 Illustration of construction of recombinant vector pFL61. Restriction digestion of cDNA and vector pFL61 with SfiI enzyme and ligation. 91

3.3 Illustration of functional screening of soil metatranscriptomic cDNA library: (a) Plasmid extraction from transformed bacteria; (b) Screening of libraries by yeast complementation ; (c) Selection of transformants on Cd amended media plates; (d) Drop out assay; (e & f) Recombinant plasmid extraction from resistant clones and their sequencing; (g) Relative expression of Cd tolerant cDNAs in the presence of Cd and (h) Transformation of cDNAs into other metal sensitive yeast mutants for further characterization. 95

3.4 Illustration of impact of 5-Flouroorotic acid (5-FOA) on yeast cells carrying *URA3*$^+$ gene. 5-FOA is converted into a toxic product 5-fluoro-uracil in the presence of *URA3*$^+$ gene product which is lethal to the yeast cells and can only survive by ejecting or flushing out of plasmid. 98

3.5 Illustration of drop out assay performed to screen metatranscriptomic cDNA library for Cd tolerant cDNAs. 99

3.6 Workflow for downstream analysis of datasets: Sequenced data was analyzed using MOTHUR software (v.1.38.1) 111

4.1 Soil sampling site: Agro-forestry land in Pierrelaye (PL) in northwest of Paris, France. 115

4.2 Total RNA isolation from soil sample PL. (a) Total RNA bands on agarose gel before DNase I treatment (b) Total RNA bands on agarose gel after DNase I treatment. M denotes 100 bp marker where as PL1, PL2 and PL3 are soil sample triplicates collected from site PL. 117

4.3 Electrophoregram generated for capillary electrophoretic separation of the total RNA extracted from the soils (a) PL1, (b) PL2, (c) PL3 through Bioanalyzer 2100 (Agilent Technologies, USA). 119

4.4 Size fractionation of cDNA synthesized from total RNA. A DNA marker was run along with cDNA on two agarose gels of same consistency. (a). Unstained gel carrying cDNA was superimposed over the EtBr-stained DNA marker gel to visualize the fractionation and three gel slices containing different cDNA size fractions were cut out (b) Agarose gel showing three fractions of cDNA i.e. fraction A, fraction B and fraction C extracted from excised cDNA gel slices 123

4.5 Validation of cDNA fractions with conserved gene family primers. PCR products were loaded on the agarose gel and amplified PCR product was observed only for (a) fraction B (msrA and RibS3) and (b) fraction C (β-Tub and EF1-α). 124

4.6 Agarose gel electrophoresis for SfiI digested cDNA libraries. First well is loaded with DNA marker 1 Kb, subsequent well with Sfi I digested Lib A, Lib B and Lib C. –4 kb) and (C) SfiI digested libraries PL-A, PL-B and PL-C. For each sub library, the size range of the cDNA inserts is similar as after size fractionation. 126

4.7 Validation of insert size of cDNA libraries by colony PCR. Panels a, b and c are the gel images of separated PCR products of 42 clones from each of the three libraries i.e., Lib A (0.1 - 0.5 kb); Lib B (0.5-1 kb) and Lib C (1-3 kb) respectively. 128

4.8 Primary screening of Cd tolerant cDNAs by transforming cDNA library into Cd sensitive yeast mutant *ycf1*$^{\Delta}$. Yeast transformants *ycf1*$^{\Delta}$ carrying cDNAs were spread on (a) SD – Ura (b) SD-Ura + $CdSO_4$. SD - Ura plates without metal considered as control. 131

4.9 Secondary mode of screening of transformants carrying cDNA corresponding to Lib C for their metal tolerant phenotype conferred to mutant *ycf1*$^{\Delta}$ at Cd 40 µM. Wild strain BY4741 and *ycf1*$^{\Delta}$ are used as control and EV denotes to empty vector pFL61. 133

4.10 Secondary mode of screening of transformants carrying cDNA corresponding to Lib B for their metal tolerant phenotype conferred to mutant $ycf1^{\Delta}$ at Cd 40 µM. Wild strain BY4741 and $ycf1^{\Delta}$ are used as control and EV denotes to empty vector pFL61. 134

4.11 Secondary mode of screening of transformants carrying cDNA corresponding to Lib A for their metal tolerant phenotype conferred to mutant $ycf1^{\Delta}$ at Cd 40 µM. Wild strain BY4741 and $ycf1^{\Delta}$ are used as control and EV denotes to empty vector pFL61. 135

4.12 Drop test of yeast clones of Lib C recovered after 5-FOA screening at higher Cd concentration i.e. 60 µM $CdSO_4$ to confirm the role of metatranscriptomic cDNAs in conferring tolerance to the hypersensitive yeast mutant $ycf1^{\Delta}$. Wild strain BY4741 and $ycf1^{\Delta}$ are used as control and EV denotes to empty vector pFL61. 137

4.13 Drop test of yeast clones of Lib B recovered after 5-FOA screening at higher Cd concentration i.e. 60 µM $CdSO_4$ to confirm the role of metatranscriptomic cDNAs in conferring tolerance to the hypersensitive yeast mutant $ycf1^{\Delta}$. Wild strain BY4741 and $ycf1^{\Delta}$ are used as control and EV denotes to empty vector pFL61. 138

4.14 Drop test of yeast clones of Lib A recovered after 5-FOA screening at higher Cd concentration i.e. 60 µM $CdSO_4$ to confirm the role of metatranscriptomic cDNAs in conferring tolerance to the hypersensitive yeast mutant $ycf1^{\Delta}$. Wild strain BY4741 and $ycf1^{\Delta}$ are used as control and EV denotes to empty vector pFL61. 139

4.15 Tertiary screening of Cd tolerant yeast transformants. Agarose gel electrophoresis marked the presence of metatranscriptomic cDNA of requisite size inside the yeast transformants $ycf1^{\Delta}$ after yeast colony PCR. Panel A, B and C showing colony PCR product of transformants from three different sized cDNA libraries i.e., C (1-3 Kb), B (0.5- 1 Kb), A (0.1-0.5 Kb). 141

4.16 Agarose gel showing bands corresponding to recombinant plasmid extracted from Lib B transformants. M denotes 1 kb marker DNA, EV stands for empty pFL61 without cDNA and sample number 1 and 2 corresponds to Lib B transformants. 150

4.17 Sequence homology of cDNA PLBe1 showing 97% identity (98% positives) with ubiquitin domain containing protein (UFP) fused with AN1 type Zn finger of *Acanthameoba castellani*. 157

4.18 Translation of nucleotide sequence of cDNA PLBe1. Red highlighted region is full length ORF showing the nucleotide sequence and encoded amino acids. 158

4.19 Multiple sequence alignment of PLBe1 gene with other homologous sequences retrieved by BLASTp analysis. Alignment was performed using the Clustal omega program and manually adjusted. Fully conserved cysteine residues (highlighted in red) and histidine residues (highlighted in green) of C– terminal zinc finger domain is shown. Accession numbers are *Mucor ambiguus* (GAN02591) *Syncephalastrum racemosum* (ORY94370), *Acanthameoba castellini* (XP_004351586), *Guillardia theta* CCMP2712 (XP_005818947), *Basidiobolus meristosporus* CBS 931.73 (ORX93623). 159

4.20 Amino acid sequence of UFP encoded by cDNA PLBe1: Highly conserved lysine residue at 48 position (green highlighted) and diglycine residue in the C-terminal end of ubiquitin moiety at 75 and 76 aa (highlighted in blue). Cysteine and histidine residues of C- terminal zinc finger moiety of the fusion protein are highlighted in red. Phosphorylation sites are highlighted in grey. 160

4.21 Secondary structure of deduced amino acid sequence encoded by cDNA PLBe1 showing various spatial arrangements of amino acids. 161

4.22 Phylogenetic relation of the transcript PLBe1 with other taxa corresponding to similar proteins constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side where as GenBank accession numbers are given in parentheses. Branch lengths are proportional to evolutionary distances. 162

4.23 Fold increase in expression levels of cDNA PLBe1 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCe10 without metal treatment). Bars sharing a common letter are not significant at $P<0.05$ ($n=3$). Error bars are ±SD. 164

4.24 Drop assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLBe1 at various concentrations a) Cd 40 µM, b) Cu 150 µM, c) Zn 10 mM and d) Co 2 mM. Wild strain BY4741 and *cup1*^{s} are used as control and EV denotes to empty vector pFL61. 165

4.25 Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLBe1 in SD-Ura medium supplemented with (a) Cd 40-80 µM, (b) Cu 150-500 µM, (c) Zn 10-12 mM, (d) Co 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3). 168

4.26 Accumulation of a) Cd, b) Cu, c) Zn and d) Co in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLBe1 Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3). 171

4.27 Sequence homology of cDNA PLCe10 showing revealed 30% identity (50% positives) to von Willebrand factor type D (VWD) of vitellogenin-6 of nematode *Ascaris suum* 175

4.28 Translation of nucleotide sequence of cDNA PLCe10. Red highlighted region is an ORF of the nucleotide sequence encoding for von Willebrand factor type D domain. 176

4.29 Alignment of amino acid sequences of different organisms showing highly conserved GL/ICG motif (red highlighted) to the C-terminal ends. Accession numbers are in parenthesis *Toxocara canis* (KHN77828), *Ascaris suum* (ERG83573), *Caenorhabditis elegans* (CCD65563), *Angiostrongylus vasorum* (ABV46673), *Haemonchus contortus* (CDJ93502) and *Oesophagostomum dentatum* (KHJ90957). Cysteine residues are highlighted with green color. 178

4.30 Deduced amino acid sequence of VWD protein: Highly conserved GL/ICG motif (highlighted in green) to the C-terminal end. Cysteine residues are given in bold red color while RXXR cleavage sites at the N-terminal end are highlighted in blue box with white font. Bold and boxed amino acids are sites for O-linked glycosylation and grey highlighted are phosphorylation sites. 179

4.31 Secondary structure of deduced amino acid sequence encoded by metatranscriptomic cDNA PLCe10 showing different possible spatial arrangements among the amino acids. 180

4.32 Phylogenetic relation of the transcript PLCe10 with other taxa corresponding to similar proteins constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side with GenBank accession numbers in parentheses. Branch lengths are proportional to evolutionary 181 distances.

4.33 Fold increase in expression levels of cDNA PLCe10 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCe10 without metal treatment). Bars sharing a common letter are not significant at $P<0.05$ (n=3).Error bars are ±SD. 183

4.34 Drop assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLCe10 at various concentrations a) Cd 40 µM, b) Cu 150 µM, c) Zn 10 mM and d) Co 2 mM. Wild strain BY4741 and *cup1*^{s} are used as control and EV denotes to empty vector pFL61. 185

4.35 Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLCe10 in SD-Ura medium supplemented with (a) Cd 40-80 µM, (b) Cu 150-500 µM, (c) Zn 10-12 mM, (d) Co 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3). 188

4.36 Accumulation of a) Cd, b) Cu, c) Zn and d) Co in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCe10 Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3). 191

4.37 Sequence homology of cDNA PLCc43 showing 51% homology (65% positives) with Hsp40 Type I protein of *Acanthameoba castellanii*. 194

4.38 Translation of nucleotide sequence of cDNA PLCc43 showing the nucleotide sequence and encoded amino acids. Full length ORF is highlighted red region. 195

4.39 Deduced amino acid sequence of Hsp40 type I like protein encoded by cDNA PLCc43: Highly conserved tri-peptide HPD (green highlighted) and G/F rich region (boxes in bold) is shown. Four cysteine rich domains CXXCXGXG (yellow highlighted bold letters) are underlined in the polypeptide sequence. Putative phosphorylation sites are highlighted in red. 196

4.40 Prediction of secondary structure of deduced amino acid sequence of metatranscriptomic cDNA PLCc43 with the confidence of prediction for various spatial arrangements among amino acids 197

4.41 Alignment of sequences of Hsp40 type I like proteins. Alignment was performed using the Clustal omega program. The fully conserved tripeptide HPD and the cysteine rich domains CXXCXGXG are highlighted in yellow and red colors respectively. Hsp40 type I from *Acanthameoba castellini* (XP_004349556) and DnaJ homolog members from *Danio rerio* (NP_998658) and *Xenopus tropicalis* (NP_001004807) are shown. 198

4.42 Phylogenetic relation of the transcript PLCc43 with other taxa corresponding to similar proteins. Phylogenetic tree was constructed using neighbor-joining method and the bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side where as GenBank Accession numbers are given in parentheses. Branch lengths are proportional to evolutionary distances. 199

4.43 Fold increase in expression levels of cDNA PLCc43 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCc43 without metal treatment). Bars sharing a common letter are not significant at P<0.05 (n=3). 202

4.44 Drop assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$ expressing PLCc43 at various concentrations (a) Cd 40 µM, (b) Cu 150 µM, (c) Zn 10 mM and (d) Co 2 mM. Wild strain BY4741/*EV*, *cup1*^{S}*/EV* serves as positive controls and *ycf1*$^{\Delta}$/EV, *cup1*$^{\Delta}$*/EV, zrc1*$^{\Delta}$*/EV* and *cot1*$^{\Delta}$*/EV* as negative controls respectively. EV stands for empty vector. 204

4.45 Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLCc43 in SD-Ura medium supplemented with (a) Cd 40-80 µM, (b) Cu 150-500 µM, (c) Zn 10-12 mM, (d) Co 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at P<0.05 (n=3). 206

4.46 Accumulation of a) Cd, b) Cu, c) Zn and d) Co in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCc43. Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at P<0.05 (n=3). 210

4.47 Sequence homology of cDNA PLCd43 showing 44% identity with γ-Glutamylcysteinyl transferase (PC synthase) enzyme of *Arabidopsis thaliana.* 213

4.48 Translation of nucleotide sequence of cDNA PLCd43. Red highlighted region is full length ORF showing the nucleotide sequence and encoded amino acids. 214

4.49 Alignment of amino acid sequence of cDNA PLCd43 encoding PCS like protein. Alignment was performed using the Clustal omega program and manually adjusted. Accession numbers are *Triticum aestivum* (Q9SWW5.1), *Arabidopsis thaliana* (Q9S7Z3), *Lotus japonicus* (Q2TSC7.1). 216

4.50 Deduced amino acid sequence of PCS like protein encoded by cDNA PLCd43: Highly conserved C^{120}, H^{237} and D^{255} are highlighted in green while five conserved cysteine residues present at N-terminal of the polypeptide are boxed. Phosphorylation sites are underlined. 217

4.51 Secondary structure of deduced amino acid sequence encoded by metatranscriptomic cDNA PLCd43 showing different possible spatial arrangements among the amino acids alongwith confidence of prediction. 218

4.52 Phylogenetic relation of the transcript PLCd43 with PC synthase enzymes corresponding to other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side where as GenBank Accession numbers are given in parentheses. Branch lengths are proportional to evolutionary distances. 219

4.53 Fold increase in expression levels of cDNA PLCd43 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCd43 without metal treatment). Bars sharing a common alphabet are not significant at $P<0.05$ ($n=3$). Error bars are ±SD 221

4.54 Drop assay of metal sensitive mutant strains *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLCd43 at various concentrations (a) Cu 150 µM (b) Zn 10 mM, (c) Co 2 mM. Wild strain BY4741/*EV*, *cup1*^{s}*/EV* serves as positive controls and *cup1*$^{\Delta}$*/EV*, *zrc1*$^{\Delta}$*/EV* and *cot1*$^{\Delta}$*/EV* as negative controls respectively. EV stands for empty vector pFL61 222

4.55 Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLCd43 in SD-Ura medium supplemented with (a) Cd 40-80 µM, (b) Cu 150-500 µM, (c) Zn 10-12 mM, (d) Co 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at 225

P<0.05 (n=3).

4.56 Accumulation of a) Cd, b) Cu, c) Zn and d) Co in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCd43. Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at P<0.05 (n=3). 227

4.57 Agarose gel electrophoresis of amplified V4 region of 18S cDNA and 18s rDNA. Each well loaded with genomic DNA (gDNA) and cDNAs product of three replicates of soil samples. DNA marker of size 1 kb was run along with cDNA and gDNA amplified products. 230

4.58 FastQC report of Read 1 and Read 2. (a) Basic statistics of the run, (b) Per base sequence quality of Read 1 and (c) per base sequence quality of Read 2. Yellow box represents the median quality score at each base window; blue line is mean quality score at each base window and red line within each yellow box is median quality score at that base window. 232

4.59 FastQC report of Read 1 and Read 2. (a) Per tile sequence quality for Read 1 (b) for Read 2 234

4.60 FastQC report of Read 1 and Read 2. (a) Per sequence quality score for Read 1 (b) for Read 2 235

4.61 FastQC report of Read 1 and Read 2. (a) Per base sequence content for Read 1 (b) for Read 2 237

4.62 FastQC report of Read 1 and Read 2. (a) Per sequence GC content for Read 1 (b) for Read 2 238

4.63 FastQC report of Read 1 and Read 2. (a) Adaptor content for Read 1 (b) for Read 2 239

4.64 FastQC report of Read 1 and Read 2. (a) Kmer content for Read 1 (b) for Read 2 240

4.65 Taxonomic annotation of both 18S rDNA and cDNA sequences obtained from soil sample from PierreLaye (PL), France. All of the major eukaryotic phyla are represented. Fungi and Metazoa phyla are predominant while in HZL plots, the Ciliophora taxa are also predominant. 241

4.66 Heatmap showing eukaryotic diversity of agro-forestry soil contaminated with potentially toxic metals 245

4.67 Principal component analysis of soil sample from agro-forestry soil contaminated with potentially toxic metals 245

LIST OF TABLES

Table No.	Title	Page
2.1.	Metatranscriptomic approach applied on different environments	61-62
3.1	Bacterial and yeast strains used in the study	74
3.2	PCR primers designed for various experiments conducted in this work	106
3.3	Soil samples with two different set of primers used for amplification of V4 region of 18S rDNA and 18S rRNA	109
3.4	Different sets of primers used for amplification of V4 region of 18S rDNA and 18S rRNA	113
4.1	Physico-chemical characteristics of the soil sampling site	116
4.2	Tolerance level of yeast transformants corresponding to cDNA library A, B and C screened against Cd toxicity. Total 498 clones screened from Lib C, 78 from Lib B and 80 from Lib A were used in drop test assay in presence of 40 µM of CdSO$_4$.	143
4.3	Tolerance level of yeast transformants corresponding to cDNA library B creened against Cd toxicity. Total 78 clones screened from Lib B were tested by drop test assay in presence of 40 µM of CdSO$_4$.	146
4.4	Tolerance level of yeast transformants corresponding to cDNA library A screened against Cd toxicity. Total 80 clones screened from Lib A were tested by drop test assay in presence of 40 µM of CdSO$_4$.	147
4.5	Transformants procured after primary, secondary and tertiary screening	148
4.6	Bioinformatic analysis of metatranscriptomic cDNAs screened for cadmium tolerance	151
4.7	Effect of different cadmium concentrations on the relative expression of cDNA PLBe1	163
4.8	Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLBe1 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant	167

strain with empty vector pFL61 (EV).

4.9 Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLBe1 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV). 170

4.10 Effect of different cadmium concentrations on the relative expression of cDNA PLCe10 182

4.11 Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLCe10 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant strain with empty vector pFL61 (EV). 187

4.12 Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCe10 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV). 190

4.13 Effect of different cadmium concentrations on the relative expression of cDNA PLCc43 202

4.14 Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLCc43 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant strain with empty vector pFL61 (EV). 206

4.15 Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCc43 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV). 209

4.16 Effect of different cadmium concentrations on the relative expression of cDNA PLCd43 220

4.17 Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLCd43 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant strain with empty vector pFL61 (EV). 224

4.18 Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLBe1 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV). 226

ABBREVIATIONS & SYMBOLS

%	Percent
°C	Degree centigrade
µg	Microgram
µl	Microlitre
µM	Micromolar
ARISA	Automated ribosomal intergenic spacer analysis
As	Arsenic
bp	Base pair
Cd	Cadmium
cDNA	Complementary deoxyribonucleic acid
cm	Centimeter
Co	Cobalt
Cu	Copper
Cys	Cysteine
Da	Dalton
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleoside-5'-triphosphate
e.g.	*exempli gratia*
EDTA	Ethylenediamine-tetra acetic acid
EST	Expressed sequence tags
et al.	*et alia*
g	Gram
HSE	Heat shock element
HSF	Heat shock factor
Hsp	Heat shock protein
i.e.,	*id est*
IPTG	Isopropyl-β-thiogalactoside
kb	Kilo base pair
KDa	Kilo dalton
M	Molar
mg	Milligram
ml	Mililitre
mm	Millimeter
mM	Millimolar
MOPS	3-(N-morpholino)propane sulfonic acid
mRNA	Messenger RNA
NGS	Next Genome sequencing
Ni	Nickel
O.D.	Optical Density

ORF	Open reading frame
OTU	Operational taxonomic unit
p.I	Isoeletric point
Pb	Lead
PC synthase	Phytochelatin synthase
PC	Phytochelatin
PCR	Polymerase chain reaction
ppm	Parts per million
PTM	Potentially toxic metals
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RT	Reverse transcription
SD	Standard Deviation
SDS	Sodium dodecyl sulphate
TRFLP	5’ terminal restriction fragment length polymorphism
Tris	Tris-(hydroxymethyl-) aminomethane
U.V.	Ultraviolet
UFP	Ubiquitn fusion protein
UPS	Ubiquitinn-proteosome system
URA	Uracil
VWD	von Willebrand factor type D Domain
w/v	Weight by volume
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactoside
Zn	Zinc

ABSTRACT

Environmental pollution due to accumulation of potentially toxic metals is an imminent universal problem that uncompromisingly endangers human health, biodiversity and ecosystems. Accumulation of heavy metals into the soil environment results into deterioration of soil and water quality, thus affecting the soil environment and its biotic as well as abiotic factors. Soil is also known as a hot spot for flourishing microbial communities and for countless phenomenon and mechanisms actively combating the heavy metal toxicity. Many of members of these microbial communities cannot be easily isolated or cultured in vitro under traditional microbiological methods, hence to appreciate their true functional diversity and their activities expressed in situ in response to various adverse environments, a new omic approach i.e., metatranscriptomics is developed which involves the use of pool of mRNA directly extracted from environmental samples. Here in this work, the potential of metal polluted agro-forestry soil from Pierrelaye (PL), France, was exploited through metatranscriptomic approach to mine the novel genes responsible for metal tolerance/resistance and to know the diversity of polluted soil at the time of sampling. To achieve this, a pilot scale sequence-based and function-based metatranscriptomic analysis on soil samples from polluted site PL was done. In the present investigation, three size fractionated cDNA libraries from total RNA of metal contaminated soil was synthesized and screened for Cd tolerant genes by yeast complementation system using Cd sensitive *ycf1*$^{\Delta}$ mutant. Here, a total of 35 cDNAs identified as unique good quality sequences are reported out of which four cDNAs with full length ORFs are further characterised for bioinformatic analysis, expression studies and multi-metal tolerance profiling. Each of the four cDNAs screened for Cd tolerance

were further characterised as ubiquitin fusion protein (UFP)/ cDNA PLBe1, von Willebrand factor type D domain (VWD)/ cDNA PLCe10, heat shock protein 40 Type I like protein (Hsp40 Type I like protein)/ cDNA PLCc43 and phytochelatin synthetase (PC synthase)/ cDNA PLCd43 as through bioinformatic analysis.

Furthermore, as the focus of the study is eukaryotic community present in the site thus the amplicons of hypervariable V4 regions of total 18S rDNA and cDNA derived from total 18S rRNA were also sequenced through Illumina-MiSeq (2 x 250 bp). Here, multiple sets of degenerate primers were used to gain a complete and more understanding of eukaryotic community structure and members of the sampling site. The NGS datasets retrieved were further analysed using Mothur pipeline after quality check *via* FastQC, Analysis was performed against SILVA reference database which primarily target V3-V4 region of 18S rRNA gene. It was reported that soil samples were dominated by the Metazoa and Fungi followed by other eukaryotic phylum such as Cercozoa, Conosa, Ciliophora etc. Hence, this study reported that amplicon sequences retrieved from soil were of eukaryotic origin and inferred that soil sample had a substantial impact on the soil eukaryotic diversity. Hence, the outcomes of this study i.e., eukaryotic origin metal tolerant cDNAs as well as amplicons sequencing depicts the efficiency of work strategy where both sequence based as well as function based metatranscriptomic approaches were exploited to focus on and exploit only eukaryotic microflora for various active genes and phenomenon active in the soil at the time of sampling.

INTRODUCTION

Chapter 1 INTRODUCTION

1.1. General Introduction

Soil is upper layer of earth and wraps almost all the terrestrial part of the earth. It is considered as a product of influence of continuous interaction between various biotic and abiotic factors like organisms, climate, parent materials (organic and inorganic matter), topography and time (Gilluly *et al.* 1975). Soil is a medium for countless physical, chemical, biological processes which are also responsible for continuous but slow changes in its properties. Soils are also largest dump for hazardous elements such as potentially toxic metals (PTMs) released directly or indirectly into soil environment through number of natural as well as anthropogenic activities (Xalxo *et al.* 2020). Potentially toxic metals is a category of metals, metalloids or their non-biodegradable compounds which persists in soil for a longer period of time and their presence above the threshold values causes long term chronic effects on the biotic as well as abiotic factors of an environment (Bolan *et al.* 2003; Algreen *et al.* 2012). Trace elements such as cadmium (Cd), chromium (Cr), arsenic (As), copper (Cu), silver (Ag), lead (Pb), cobalt (Co), zinc (Zn), nickel (Ni), iron (Fe) and the platinum (Pt) group elements are categorised into the group of PTMs due to their lethal effects on an environment (Diaconu *et al.* 2020). Natural processes such as weathering, volcanic eruptions, soil erosion, leaching and seepage of metals into water bodies and soil, introduced number of elements (including PTMs) in a very low amount hence regarded as trace elements (Kabata-Pendias and Pendias, 2001; Pierzynski *et al.* 2008). Principally, introduction of PTMs into the soil environment is attributing to the negative impact of technological developments such as speedy urbanization and industrialization with poorly or unplanned waste disposal and management system (Swarup and Dwivedi, 1998). Accumulation of high amount of metals and metalloids in a soil environment is result of toxic emissions from various anthropogenic activities such as rapidly expanding industries, mining and metallurgical processes, generation of high metal e-waste, spillage of petrochemicals, leaded gasoline, sewage sludge, agricultural products such as fertilizers, coal industry and transport sector (Kaplan *et al.* 2011; Fernández *et al.* 2018). Such disturbances and hastening of slowly occurring natural geochemical cycles of metals by human activities results into the accumulation of one or more PTM above the defined values that are high enough to impose serious risks to human beings as well as other living beings such as plants and animals (D'Amore *et al.* 2005; Basta *et al.* 2005). It has been projected that emission of metals into the environment is one to three fold higher than the natural fluxes and PTMs from anthropogenic sources tend to be easily bioavailable than nature's pedogenic or lithogenic processes (Sposito and Page, 1984; Kaasalainen and Yli-Halla, 2003). Accumulation of PTMs has become serious threat to the soil environment due to (i) generation rate of PTMs via anthropogenic activities is relatively higher to natural ones, (ii) PTMs get transferred from mines to random environmental sites where the potential of direct exposure is higher, (iii) the difference in concentration of PTMs in discarded products (relatively high) in comparison to the receiving end in an environment (iv) presence of metals in a form of chemical species that is more bioavailable (D'Amore *et al.* 2005). Unlike organic contaminants, PTMs do not experience any microbial or chemical degradation and persists in an environment for a longer period of time after their introduction, however transformation of these compounds into more stable and bioavailable chemical species is possible (Kirpichtchikova *et al.* 2006; Adriano, 2003). Presence of PTMs in soil rigorously inhibits the biodegradation of organic pollutants (Maslin and Maier, 2000) and poses serious threats to health of living beings and quality of an environment by entering into the ecosystem through direct or indirect contact with polluted soil, drinking polluted ground water and food chain (McLaughlin *et al.* 2000a; McLaughlin *et al.* 2000b, Ling *et al.* 2007). Metal contamination in ecosystem affect the cellular components of organisms such as cell organelles and enzymes involved in maintaining the homeostasis inside a living organism (Wang and Shi, 2001). Potentially toxic metal ions interact with cellular machinery or biomolecules such as nuclear material and nuclear proteins which thereby led to cell cycle modulation, carcinogenesis or apopstosis (Chang *et al.* 1996; Wang and Shi, 2001; Beyersmann and Hartwig, 2008). Further, entry of these PTMs inside the cells lead to the production of reactive oxygen species (ROS) and oxidative stress which lead to toxicity inside the cells (Tchounwou *et al.* 2001; Tchounwou *et al.* 2004; Patlolla *et al.* 2009; Yedjou and Tchounwou *et al.* 2008; Sutton and Tchounwou; 2007). In response to metal toxicity, living organisms have numerous mechanisms and phenomena which aid in surviving and adapting such adverse conditions of metalliferous soil through maintenance of homeostasis (Xalxo *et al.* 2020). These mechanisms include cell wall, extracellular exudates and mycorrhiza as extracellular members while intracellular strategies involve plasma membrane which reduces metal uptake or facilitate the efflux of metals entered in cytosol, chelation through various biomolecules and vacuolar compartmentalisation (Kumar *et al.* 2016). Therefore, adverse conditions such as metal contamination which affect the soil abiotic properties also affect the diversity and functioning of microorganisms living in that environment.

Soil is considered as a richest ecosystem with a large portion of genetic diversity on earth where majority of the organisms belongs to microflora of the soil (Ponge, 2015). Many microorganisms which are the principal agent of organic mineralization in soil may also have the potential to convert potentially toxic metals in to stable and/or less toxic forms (Bell *et al.* 2014). Although microbes exist throughout the soil profile however they are most abundant in surface soils, rhizospheric soils and around the macropores (Bundt *et al.* 2001; Fierer *et al.* 2007). It was observed that one gram of soil represents 10^{10} to 10^{11} bacteria (Horner-Devine *et al.* 2003) with 6,000 to 50,000 estimated numbers of bacterial species (Curtis *et al.* 2002) and approximately 200 million of fungal species (Leake *et al.* 2004). Soil microbes such as bacteria, archaea and fungi play diverse and significant roles in maintaining the homeostasis in the ecosystem by playing crucial role as important component of biogeochemical cycling, plant productivity, enriching soil diversity through symbiosis (mycorrhizal fungi and nitrogen fixing bacteria) and mineralisation especially in poor ecosystems (Van Der Heijden *et al.* 2008). Wide metabolic diversity of soil microbes hints that their activities might have contributed towards cycling of major elements such as carbon (C), nitrogen (N), phosphorus (P) and PTMs which further affects the structural as well as functional characteristics of the soil. Both uni- and multicellular members of soil ecosystem have been explored to discover a range of tolerance mechanisms against PTMs at both molecular and cellular level which further helps in revitalizing the PTM contaminated sites (Thavamani *et al.* 2017; Teng *et al.* 2017).

Although biosphere is principally dominated by microorganisms (Whitman *et al.* 1998), yet most of them have not been explored as traditional culture techniques limits the study to only those that can be cultivated under laboratory conditions (Staley and Konopka, 1985; Hugenholtz and Pace, 1996). Recent researches in microbial ecology advocate the presence of numerous varieties of novel microorganisms in the environment in comparison to the few lab cultivated microbes (Ward *et al.* 1990; Giovannoni *et al.* 1990). This is further authenticated from DNA complexity and discovery of numerous unique 16S/18S rRNA gene sequences from various environmental sources (Torsvik *et al.* 1996; Goodman, 1997; Handelsman *et al.* 1998). It was observed that 99 to 99.9% of microflora present on the earth are unable to grow in lab conditions due to tremendous diversity in physiological capabilities of these microorganisms (Amann *et al.* 1995). A major obstacle in culturing indefinable microbial communities from soil is fast growing microbes which will out-compete the majority of slow growing microbes. It is observed that huge majority of indefinable microbes are slow growing and their growth rate is rarely at optimum in natural habitat. Though some modified culturing techniques such as using antibiotics, diluting nutrient conditions, using nutrient poor media cultivation with other microbes and longer incubation periods allowed the cultivation of few of the uncultivable microbes that are already indentified using various molecular techniques in soil (Kaeberlein, 2002; Stingl *et al.* 2007; D’Onofrio, *et al.* 2010; George *et al.* 2011; Stewart, 2012). While technique of culturing the soil microbes on selective and modified media with known nutrients sources such as carbon sources (e.g. glucose, cellulose, pectin), nitrogen sources (e.g. amino acids, alkaloids) or sulphur sources (e.g. arylsulphate) could be used to successfully isolate the functional microbes from an environment (Maimaiti *et al.* 2007). Potential of microflora to flourish in unfavorable environments such as metal polluted soils, xenobiotic compounds; saline water bodies etc., could be a screening medium to isolate novel microorganisms (Ramesh *et al.* 2007). Since, these approaches entailed physiologically accurate concentrations of these compounds to trigger the required phenotype in particular organisms, hence finding such parameters is also a huge task. Furthermore, the biggest challenge in isolation of novel microorganisms is utter diversity, for instance lifestyle of archaea and eukaryotes is quite diverse and different in comparison to bacteria and has been abandoned from efforts to culturing these microorganisms. The major obstruction in the study of eukaryotic microbial communities is intrinsic characteristics of eukaryotes such as large genome size, presence of introns and dilution of eukaryotic nuclear material in total gene pool of an environment. Thus, functional and ecological roles of individual eukaryotic taxa is still unexplored and unknown as majority of the eukaryotic microbes are considered as a part of super-organisms where each microbial group possess special features responsible for numerous biochemical transformations and require more sophisticated methods for their isolation (Prosser, 2007; Fierer *et al.* 2007).

To explore the true diversity of unexplored microbiota of an environment, various culture independent approaches such as metagenomics have been designed. Metagenomic approach involves direct isolation of nucleic acid i.e. DNA from an environmental sample and has been proven to be a promising approach to compare and explore the soil ecology (Biddle *et al.* 2008). Although, metagenomics carry more vast genetic information in comparison to cultivable host but it has not laid any dramatic impact on improving cultivation of unknown microflora as it involves study of metagenome comprises of total genomes present in an environment (Tyson and Banfield, 2005). Furthermore, high throughput sequencing techniques in conjunction with metagenomics have provided an extraordinary insight into the world of uncultured microorganisms but culturing an uncultured microbe through these techniques is rarely observed as it does not provide any information on actual functional characteristics possessed by these microbes in an environment (Qin *et al.* 2010; Tyson and Banfield, 2005). Therefore to explore the activities of microbial communities flourishing in a specific environment such as metal polluted soil, another approach i.e. metatranscriptomics is developed in which gene expression profiling of functional and taxonomic marker genes can be studied through community mRNA (Handelsman *et al.* 2007). Metatranscriptomics involves the use of high throughput sequencing techniques for community level transcriptional profiling of reverse transcribed RNA providing a direct access to the community transcriptional profiles (Carvalhais *et al.* 2013). Next generation sequencing techniques such as Illumina’s HiSeq and MiSeq platforms is known to be providing insight into the various metabolic interactions within microbial communities by directly sequencing the metatranscriptomes extracted from various environments (Bentley *et al.* 2008; Frias-Lopez *et al.* 2008; Gilbert, *et al.* 2008; Urich, *et al.* 2008; Turnbaugh, *et al.* 2009; Turnbaugh, *et al.* 2010). Furthermore, amplicon sequencing of a variable region of the 18S rRNA gene from various environmental samples has showed the way towards better understanding of eukaryotic microbial diversity in polluted soil environments (Marmeisse *et al.* 2017). Next generation techniques in conjunction with functional metatranscriptomic approach where mRNA derived cDNA libraries are screened for novel genes and microbes by transforming into suitable eukaryotic host is a promising approach. This approach connects the community based structural and functional information in a single experimental analysis and provides us information beyond the genomic potential of a community as in case of DNA based methods. This advancement approves a new approach to direct the designing of culture conditions based on *in situ* physiology which further helps in isolation of active microflora especially eukaryotes from an environment. Hence, ecological significance of microbes to all the living beings present in our environment compel the field of molecular microbial ecology to access the wealth of information present within the total microflora with the help of new omic techniques such as metatranscriptomics (Handelsman *et al.* 1998).

1.2. Aim of the study

Eukaryotic microbiota represents a hidden and under explored source of numerous biocatalysts and metabolites that have been procured from a partial number of taxa thus does not represent whole eukaryotic realm with a remarkable significance. This study aims to explore the hidden and unexplored domains of soil eukaryotic microflora responsible for numerous processes and mechanisms that can be significant in bioremediating the metal polluted sites. This was achieved by exploring the functional diversity of soil environment by targeting V4 region of 18S rRNA/rDNA through next generation technique Illumina Miseq along with screening of novel metal tolerant gene and microbes through functional metatranscriptomic approach where mRNA derived cDNA libraries were screened by transforming into metal sensitive yeast strains. In this study, the RNA content (gene expression) of the soil sample was explored through metatranscriptomic approach which gave an insight into microbial activity and their response towards environmental metal stress and rRNA to rDNA ratio indicated the relative microbial activity in metal polluted soil. This is the first study reported in the field of microbial ecology where both amplicons sequencing through high throughput sequencing technique and functional approach of metatranscriptomics have been exploited altogether to explore metal polluted soil site. This approach connects the community based structural and functional information in a single experimental analysis and provides us information beyond the genomic potential of a community as in case of DNA based methods. Henceforth, to explore our soil sample by these approaches, our study was divided into three main objectives.

Objectives

- Construction of cDNA libraries from RNA isolated from metal polluted soil
- Heterologous expression of metal tolerant genes by yeast functional complementation
- Documentation of functional eukaryotic diversity from metal contaminated soil

REVIEW OF LITERATURE

Chapter 2 REVIEW OF LITERATURE

2.1. Soil

Soil is one of the most complex biosystem on the earth, functioning at the interface of the lithosphere, atmosphere, hydrosphere and biosphere (Young and Crawford, 2004). Soil is a prime element of the terrestrial ecosystem, comprised of organic matter, minerals, gases, liquids and living organisms such as microorganisms, plants, animals and all these affiliates of soil together support life on this planet (Szabolcs, 1994). Soil organic matter (SOM) and minerals in together constitute solid phase of the soil while porous phase holds gases and water content of the soil. Soil is known to have significant role in four different ways: (i) acts as medium for the growth of small and big members of kingdom plantae, (ii) acts as means of water storage, supply and purification, (iii) centre for occurrence of numerous phenomena and nutrient cycles (iv) habitat for numerous microbes and other organisms. All these functions either occurring inside the soil or supported by the soil, also impact other properties of the soil such as texture, pH, porosity, water retention, hence, modify Earth's atmosphere (Voroney, 2007). In 1973, Lovelock and Margulis, proposed the Gaia hypothesis which stated that the physico-chemical properties of earth are strongly related to the activities of living organisms present on it and theorized earth as a "super-organism" with an inherent ability to control its climate and chemistry in order to maintain an environment favorable for the life (Lovelock and Margulis, 1973). Due to its significant role in controlling and maintaining the earth's environment hence the sustainability of life on the earth, soil is referred as a "critical zone" of the earth and also given a special status among other components of the planet (Voroney, 2007). Thus, to study this component of the ecosystem, one should be aware of process of its genesis and its biotic as well as abiotic components.

2.1.1. Soil habitat and organisms

Soil habitat is a term given to the entirety of the living organisms i.e., microbes, flora and fauna inhabiting the soil environment and its abiotic factors (Lavalle, 2012). The characteristics of habitat of a community are determined by the geology and climate of that terrain and plant vegetation. Interaction among various kind of parent material with various other abiotic factors such as sunlight, wind, temperature, rainfall, latitude elevation etc. over a wide geographical region with similar environmental parameters and biotic communities has evolved into present day terrestrial biomes (Aislabie and Deslippe, 2013). Soils provide a wide range of habitats to the living biota hence harbour most of the earth’s genetic diversity. Soil is an abundant ecosystem on the earth for various organisms especially microorganisms and majority of them are unknown and undescribed (Torsvik and Ovreas, 2002). It has been estimated that one gram of soil accommodate around one million of microorganisms with more than 10^6 of bacterial cells and kilometres of fungal hyphae while the total number and diversity of species vary from type of soils (Roesch *et al.* 2007). Prokaryotic and eukaryotic microorganisms such as bacteria, fungi and mycorrhiza play a significant role in numerous biological processes and acts as reservoir of nutrients in biologically active layer of soil i.e., rhizosphere (Aislabie and Deslippe, 2013). Soil organisms break down the complex organic matter into simpler forms to be easily assimilated by plants and other organisms. Soil living organisms help in maintaining the soil structure and prevent the nutrient loss by storing them into their bodies (Young and Rietz, 2000). Soil biological component includes megafauna (rabbits, rodents) which is size 20 mm upward, macrofauna (2 -20 mm size) such as woodlice, earthworms, beetles, centipedes, ants etc, mesofauna (100 μm to 2 mm) such as springtails, microfauna and microflora (1-100 μm) such as bacteria, yeast, fungi, protozoa and rotifers (Artz *et al.* 2010). Out of these biological components, bacteria and fungi plays a significant role in maintaining soil by acting as decomposers to organic matter to produce detritus and other simple compounds (Fierer *et al.* 2007; Eilers *et al.* 2010). Soil bacteria residing in soil helps in transforming complex forms into simpler forms or vice versa through various processes such as nitrification, nitrogen fixation and denitrification (Goldfarb *et al.* 2011; Yadav *et al.* 2018). Soil fungi share both parasitic and beneficial relation with other living organisms. Most of the fungal species are saprophytic in their behavior as they flourish on dead and decaying organic matter beginning with the use of sugars and starches followed by utilizing cellulose and lignin (Barr and Aust, 1994; Thwaites *et al.* 2006). Symbiotic association of fungi with plants is called mycorrhiza in which mycelium flourishes partly in the soil and partly in the roots obtaining carbohydrates from the roots and in return providing nutrients and moisture to the plants (Deslippe *et al.* 2012). Further, root exudates from various plants acts as growth stimulators for microbial growth thereby increasing the growth of predators like amoeba hence increases the mineralisation rate. Most of these microbes in the bulk soil form mucilaginous micro-aggregates of size 20-250 µm approximately, on which clay particles are stick to provide protection against various adverse conditions like dessication and predators’ attack by members of soil micro-fauna such as protozoa and nematodes (Ritz *et al.* 2003, Davinic *et al.* 2012).

Soil macrofauna and mesofauna facilitate the process of bioturbation which involves movement of moisture and gases into the soil (Meysman *et al.* 2006). Micro-zones with good aeration may be only extended up to few millimetres from areas with poor aeration. Likewise, the areas close to the surface of soil may be rich in decomposing organic matter and nutrients while subsoil may be deprived in nutrients; the soil solution in pores may be highly acidic or more basic depending upon the biological activities and mineralogy of the respective soil (Brady and Weil, 2002). Activities of earthworms, termites, ants, millipedes and beetles (tenebrionid) affect the soil formation by facilitating the mixing of soil layers due to burrowing and ingesting soil particles and organic matter hence convert into simpler forms easy to be assimilated by plants (Lee and Foster, 1991; Scheu, 2003). These are categorised as “ecosystem engineers” as they exert a primary effect on soil quality by transporting soil particles from one layer to another, increasing the soil porosity, aeration and infiltration of water into soil easily (Jouquet *et al.* 2006). Similarly, soil detritivores like earthworms ingest detritus and other breakdown products while saprotrophs extract soluble nutrients from detritus products. Animals like ants, termites can also impede the soil development by burrowing and denuding large soil areas around their habitation (De Bruyn and Conacher, 1990). Higher animals like gophers, moles and dogs dig into the lower soil horizons bringing materials to the surface and their tunnels are always open to the surface encouraging the movement of water into the lower layers (Kinlaw, 2006). Mixing of soil through such animal activities are called pedoturbation which create distinct soil horizons by mixing of lower and upper soil horizons by creating and refilling such underground habitats (Bohlen *et al.* 2004).

Similarly, vegetation of a region plays an important role in soil development and structure improvement via numerous ways such as break down of complex minerals by plant secretions, decomposition of dead plants and their parts increases the organic matter content of the soil hence the growth of other plants and microbes thus initiating soil formation. Flora of a region penetrate deep into the soil, making nutrients from deeper soil profiles available to upper layers by absorption through plant roots and increases porosity of the soil profile with open channels after decomposition (Lynch, 1995; William and Weil, 2004). Plant root exudates contains organic compounds like sugars, organic acids, mucigel, marsh off cells which are easy to decompose by various microbes hence increases the organic matter and fertility of the soil (Nguyen, 2003). Plants also prevents the soil erosion through strong hold of roots (Gyssels *et al.* 2005), balance the soil moisture content with slow rate of evaporation or transpiration thereby preventing soil from desiccation during dry seasons and excessive moisture during moist months due to highly variable relationships between moisture loss and leaf area index (Heck *et al.*1999). Diversity of vegetation growing on a terrain depends upon certain soil factors such as chemistry, pH, temperature, density, depth and moisture content (Vogiatzakis *et al.* 2003). Anthropogenic activities like different agricultural practices, industrialization, transportation, civilizations influence the soil formation, its quality and development by affecting various factors like soil texture, soil composition, moisture, mineral contents and organic matter content of the soil (Dudal, 2005). Agricultural practices such as setting fire to the large part of land (Prairie grasslands), use of fertilizers to increase the crop yield and pesticides to control the pests affect the soil quality both positively and negatively (Anderson, 2006; Lisetskii, 2016). Similarly addition of industrial effluents to the water bodies and soil abruptly disrupt the soil quality as these effluents are rich in xenobiotic compounds and various potentially toxic metals (PTMs) which are tough to decompose. This modern era of civilization has degraded the soil quality by modifying the soil formation processes hence results into more polluted and less fertile land (Schon *et al.* 2011).

2.2. Metals: classification and potentially toxic members

Metals are members of periodic table with high electrical conductivity, malleability, ductility and luster that are directly proportional to another physical property i.e., temperature (Müller, 2007; Housecroft and Sharpe, 2008). Such temperature dependent properties make it easy to define metals and to distinguish them from non-metals and metalloids. The properties of these chemical species can be determined by their respective positions in the periodic table (Figure 2.1). Metallic characteristic of elements decreases when we move from left to right and from bottom to top of the periodic table, the diagonal border between metals and non-metals corresponds to the elements with intermediate properties between metals and non-metals called metalloids. Further, metal elements of periodic table can also be grouped depending upon the last electronic sub-shell present in its atom such as s-block elements, d-block elements, f-block elements and p-block elements. The s-block elements are further subdivided into two main groups: alkaline elements and alkaline earth elements among which except hydrogen, all elements present are metals. The d-block elements also known as transition elements are all metals and have capability to form complexes with other elements by attaining different valence states on losing their valence shell electrons. The members of f-block also known as rare

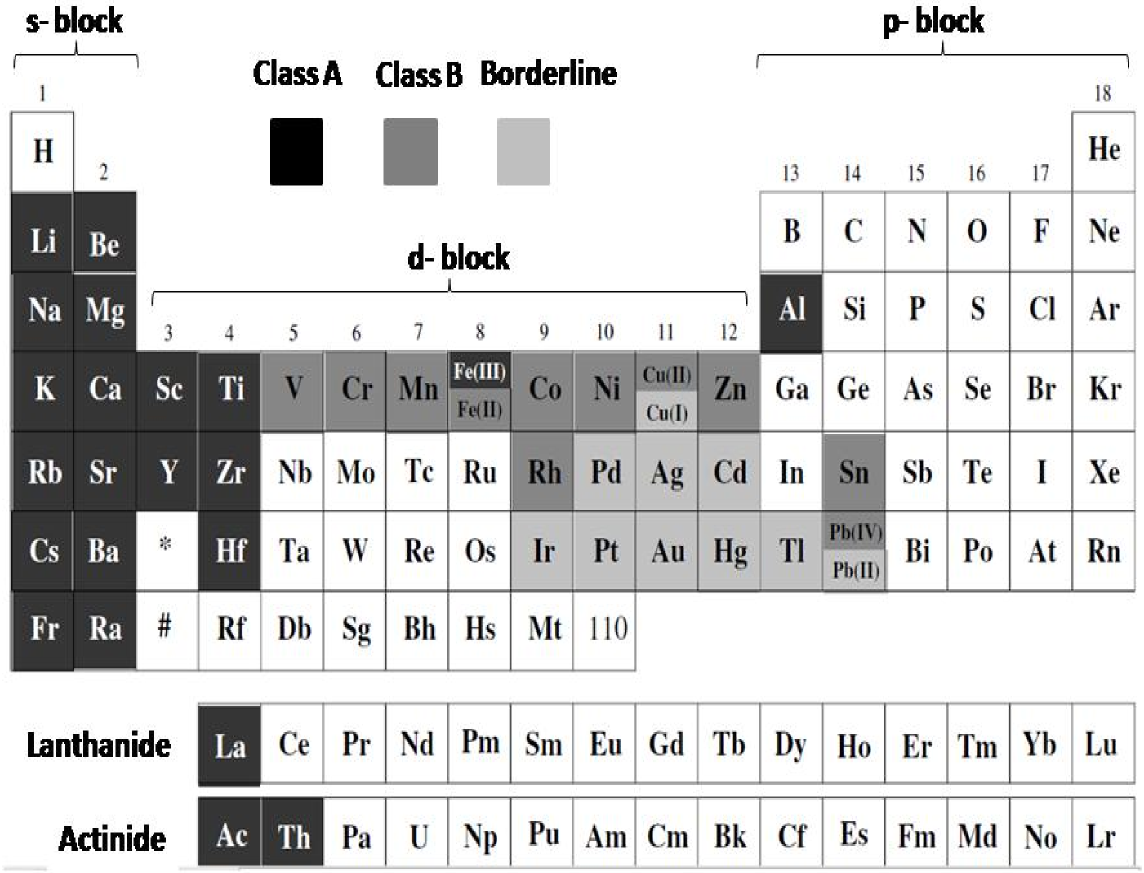


Figure 2.1: Metals of periodic table classified as Class A (black), Class B (dark grey) and borderline (light grey) where Fe may belongs to either Class A metals or borderline depending upon its oxidation state i.e., Fe^{3+} or Fe^{2+} while Cu and Pb may be Class B or borderline depending upon whether Cu^{+} or Cu^{2+} in case of Cu and Pb^{2+} or Pb^{4+}. s-block, p-block, d-block, lanthanide and actinide series of periodic table are also shown (Duffus, 2002).

earth metals further subdivided into lanthanides series and actinides series. The next group positioned at the right side of the periodic table corresponds to p-block elements which consist of metals, metalloids and non-metals. This includes third to seventh group of periodic table excluding eighth group of noble gases. Metallic members in p-block consist of Pb, In, Po, Tl, Cb, Bi, Al, Sn, Te and Ga and form amphoteric oxides except Bi. Similarly, other elements of the p-block i.e., Te, Ge, As and Si are categorised as metalloids; in which B and Sb are also included occasionally. Nieboer and Richardson (1980) classified metals and metalloids in three different classes i.e., class A, the borderline and class B on the basis of ionic and covalent indices where class A has ability to react with oxygen carrying ligands e.g. Mg^{2+}, Ca^{2+}, Al^{3+}, As^{3+} and the affiliates of class B forming bonds with S- and N-containing ligands (Figure 2.1). Generally, Class B group members such as Ag^{+}, Tl^{+}, Hg^{2+}, Cd^{2+} are very toxic in nature and their existence as thiol or imidazo groups in proteins led to the toxicity among living beings. An alternative term for class A ions or O_2 donating ligands is hard acids and for S donating ligands or class B is soft acids with some intermediate between the two classified on the basis of stability constants of the complexes formed (Pearson, 1968).

Metals in soils are present in different forms such as free metal ions, exchangeable metal ions, soluble or insoluble metal complexes, organically bound metals, precipitated or insoluble compounds such as in the form of oxides, carbonates and hydroxides, or as indigenous soil content in the form of silicate minerals. Although metals exist naturally in the soil environment, environmental contamination occurs due to various anthropogenic activities such as mining, smelting, industrial processes, domestic and agricultural use of compounds containing potentially toxic metals (PTMs) in traces (Shallari *et al.* 1998; Goyer, 2001; He *et al.* 2005; Herawati *et al.* 2000). Potentially toxic metals (PTMs) is a category of metals, metalloids or their compounds possessing high density and high relative atomic weight with metallic properties such as ductility, malleability, conductivity, ligand specificity (Algreen *et al.* 2012). Various natural processes such as weathering, volcanic eruptions, soil erosion, leaching and seepage of PTMs into ground water and soil through water resources results into environmental contamination (Nriagu *et al.* 1989). Coal burning in power plants, petroleum combustion, nuclear power stations, metal processing in refineries, plastics, textiles, microelectronics, wood preservation and paper processing plants are main industrial sources of metal contamination in the environment (Arruti *et al.* 2010; Strater *et al.* 2010; Pacyna, 1996). Toxicity of a metal in soil depends on its bioavailability i.e., their ability to be transferred from soil to a living organism (Morel, 1996). According to Hamelink *et al.* (1994), PTMs are present in environment in trace amounts i.e., less than 10 ppm and their bioavailability is affected by physicochemical factors such as pH, organic matter etc., that affect thermodynamic equilibrium and kinetics of a reaction, n-octanol/water partition coefficients (K_{ow}) and lipid solubility (Hamelink *et al.* 1994). Various biological aspects for instance species characteristics, biochemical or physiological adaptation, trophic level interactions also have significant role in the availability of PTMs to the living organisms (Verkleji *et al.* 1993). Worldwide, the estimate of PTMs released annually from all the sources was about 22000 metric ton of Cd; 1,350,000 of Zn; 939,000 of Cu and 783,000 of Pb (Oves *et al.* 2016). As the PTMs show high and diverse binding affinity towards soil constituents hence soil is a major reservoir media for potentially toxic elements. Some PTMs such as Co, Cu, Cr, Zn, Fe, Mn, Mo, Ni, Se and Mg are necessary in very low concentrations to carry out various biochemical and physiological functions among the living beings. These elements are called essential elements and acts as cofactors in majority of the enzymes and played significant role in a range of oxidation and reduction reactions occurring in the living organisms (WHO, 1996). However, inadequate supply of these micronutrients results into various other malformations or defects such as growth retardation or diseases while the concentrations higher than the threshold concentrations have adverse effects on living beings and the environment (Underwood, 1971; Mitchell and Burridge, 1979). For instance, Cu is an essential element in numerous enzymatic reactions such as dopamine β-monoamine oxygenase, superoxide dismutase, ferroxidases, peroxidase, cytochrome oxidase and monoamine oxidase (Festa and Thiele, 2011). The ability of Cu to exist as oxidised Cu (II) state and reduced Cu (I) state classified it as an essential element used in key biological processes such as carbohydrate metabolism, biosynthesis of catecholamine and cross linking of fibrous proteins such as collagen, elastin etc (Agency for Toxic Substances and Disease Registry, 2002; Harvey and McArdle, 2008; Stern, 2010). However, this property of copper also makes it potential toxic element as the transition of Cu (II) and Cu (I) ions results into oxidative stress due to generation of superoxide and hydroxyl radicals (Tchounwou *et al.* 2008). Similarly, presence of excess amount of other essential elements in an environment can potentially result into adverse effects due to cellular and tissue damage among living beings. Some of the elements such as Cr and Cu have very small range of concentration between beneficial and lethal effects. Other category of elements which have no biological significance are called non-essential elements and are mostly toxic to the living beings in their relative ionic state such as Sr^{2+}, Tl^{+}, U^{4+}, As^{3-}, Ba^{2+}, Be^{2+}, Bi^{3+}, Cd^{2+}, Al^{3+}, Sb^{3+}, In^{3+}, Pb^{2+}, Li^{+}, Hg^{2+}, Ni^{2+}, Pt^{2+}, Ga^{3+}, Ge^{4+}, Au^{+}, Sn^{2+}, Ti^{+}, V^{3+} and Ag^{+} (Chang *et al.* 1996; Tchounwou *et al.* 2008).

2.2.1. Mode of action of potentially toxic metals

To classify a metal element into a toxic substance, two facts should be considered: (a) the effect of that metal element on living organisms is always dependent upon its bioavailability, hence toxicity of a metal is dose response figure (b) Evaluation of relationship between dose of the PTM and response of the living organisms e.g. a bell shaped dose response relationships predicts that a particular metal is crucial at low concentrations while toxic at high concentrations (Marschner, 1995). Potentially toxic metals affect cellular components such as cell membrane, mitochondria, endoplasmic reticulum, nuclei and enzymes involved in metabolism, detoxification and damage repair and in maintaining the homeostasis inside a living organism (Wang and Shi, 2001). These metal ions interact with cellular machinery or biomolecules such as nuclear material and nuclear proteins which led to cell cycle modulation, carcinogenesis or apopstosis (Chang *et al.* 1996; Wang and Shi, 2001; Beyersmann and Hartwig, 2008). Such PTMs have been reported to cause toxicity inside the cells by four means: (1) production of reactive oxygen species (ROS) to elicit oxidative stress; (2) damaging the structure of DNA molecules and/or inactivating DNA repair mechanisms; (3) interference with membrane functions and nutrient assimilation; and (4) to disturb protein functions and activity (Lemire *et al.* 2013; Beyersmann and Hartwig, 2008; Wysocki and Tamás; 2010). While at molecular level, Naumann *et al.* (2007) reported that there is huge number of potential target proteins for a specific PTM and if a PTM is modified or its concentration is changed to compare its toxic effects, there is a possible probability to observe the change in dominant target proteins due to decreased or increased lewis acid strength. This would affect number of other proteins which are part of different metabolic pathways inside the affected cells (Appenroth, 2010). These PTMs interfere with protein folding *in vivo*, refolding of denatured proteins *in vitro* by three possible ways: (1) binding to functional groups such as thiols (2) displacement of essential metals especially cofactors in metalloproteins; or (3) catalyze the oxidation of amino acid side chains (Lemire *et al.* 2013; Beyersmann and Hartwig, 2008; Sharma *et al.* 2011). Hence, all these modes of metal action that target proteins results into aggregation of denatured and nascent proteins which further affect the cell homeostasis and viability (Sharma *et al.* 2008).

In 1981, Berry and Wallace has described the dose response curves for essential metals and observed deficiency at sub-optimal concentration, tolerance at optimal concentration and toxicity at higher concentration (Figure 2.2). Essential metals such as Zn^{2+} and Cr^{3+} when present in high concentrations can cause toxicity due to oxidative stress even in the most biologically active form (Sharma *et al.* 2009; Nickens *et al.* 2010).

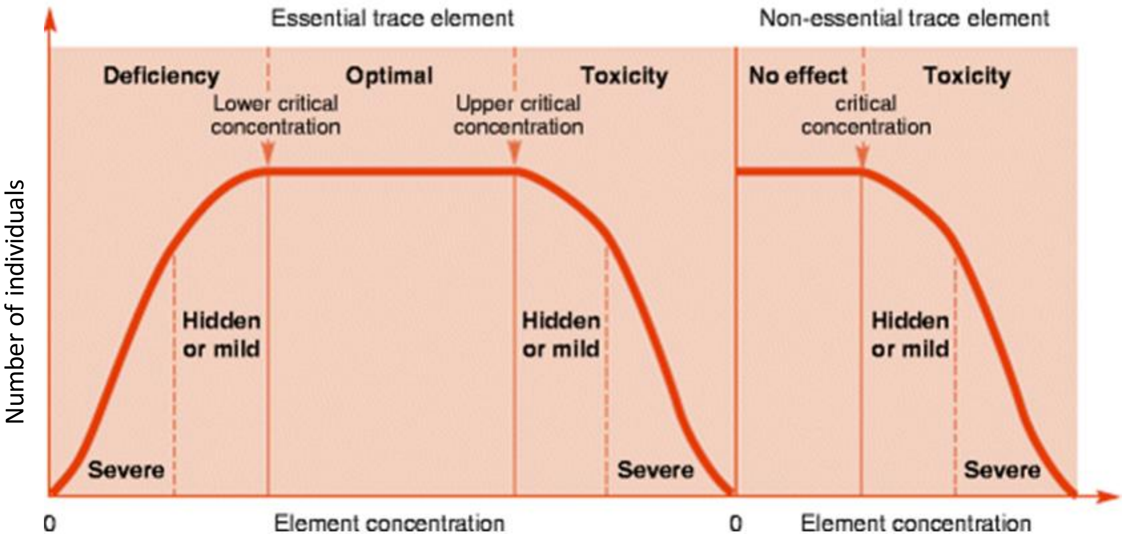


Figure 2.2: Dose-response curves for essential and non-essential metals. Curve for essential metals showed deficiency at low and sub optimal concentration, tolerance at optimal concentration and toxicity at higher concentration while for non-essential metals severe toxicity was observed after a threshold concentration of metals (critical concentration) (Alloway, 2008).

In high concentration, Cr^{3+} start producing 8- hydroxydeoxyguanosine which further breaks the DNA strands hence causing carcinogenicity (Qi *et al.* 2000). Similarly, Cu^{2+} metal when present in higher concentration participates in Fenton's reaction resulting into production of highly reactive hydroxyl radicals (Valko *et al.* 2005). There is one little known fact that certain non-essential metal elements when applied at low concentrations have a stimulating and inducing effect on living cells, hence, called low concentration stressors. For instance, Cd^{2+} grant stimulating effects in barley seedlings whereas metals such as Pb^{2+} and Ti^{2+} resulted into detached leaves in barley plant (Kovacs *et al.* 2009; Nyitrai *et al.* 2007). It has been observed that Cd^{2+} ions displace the essential metal ions i.e., Zn^{2+} and Fe^{2+} in proteins hence results into oxidative stress especially due to lipid peroxidation (Bertin *et al.* 2006). It also inhibits the mismatch repair pathway by binding to low sequence specificity DNA inducing single strand breaks in DNA strands (McMurray and Tainer, 2003). Although Cd^{2+} metal itself does not catalyse the generation of ROS directly but it promotes the oxidative stress by continuously depleting the antioxidant defences inside the cell (Adamis *et al.* 2004). Similarly, Ni^{2+} metal a known carcinogen, cause toxicity due to production of DNA damaging ROS species and inhibiting the DNA repair mechanisms inside the cells (Kasprzak *et al.* 2003). Thus, it has been concluded that the main biochemical mechanism responsible for toxicity inside the living organisms is oxidative damage to cellular macromolecules like DNA and proteins (Beyersmann and Hartwig, 2008). Thus, all cells and organisms maintain metal homeostasis within physiological or sub-toxic levels by utilizing their metal detoxification mechanisms respectively.

2.2.2. Sources of metal pollution in soil

Over the last century, research on PTMs in the field of agriculture, environment and life sciences has been done due to increased social apprehension over presence of excess amounts of these elements in the environment. In pre-technological era, the cycling of these elements has reached a stable state and a rigorous natural balance was maintained on distribution of these elements in the environment (Adriano *et al.* 2004). In modern era of urbanization and industrialization, the consumption of these elements has been increased due to ever-expanding technologies (Anirudhan and Sreekumari, 2011). Soils are considered as sink for the PTMs released from various anthropogenic activities such as industrial effluents, automobiles, mining and agricultural practices (Figure 2.3). During past decades, the level of PTMs have increased significantly in the soils and distributed so widely that increased level of PTMs from anthropogenic origin has also been reported from remote areas such as glaciers (Fernández *et al.* 2018). Potentially toxic metals enter in the soil system by number of pathways and their behavior and fate differ depending upon their chemical state and sources (Silva *et al.* 2020). The most common sources of anthropogenic contamination of PTMs in soil environment involves use of commercial fertilizers and other agrochemicals, industrial effluents, sewage sludge, irrigation water, depositions from urban and automobile emissions and various chemical used as soil amendments (Shallari *et al.* 1998; Goyer, 2001; He *et al.* 2005; Herawati *et al.* 2000).

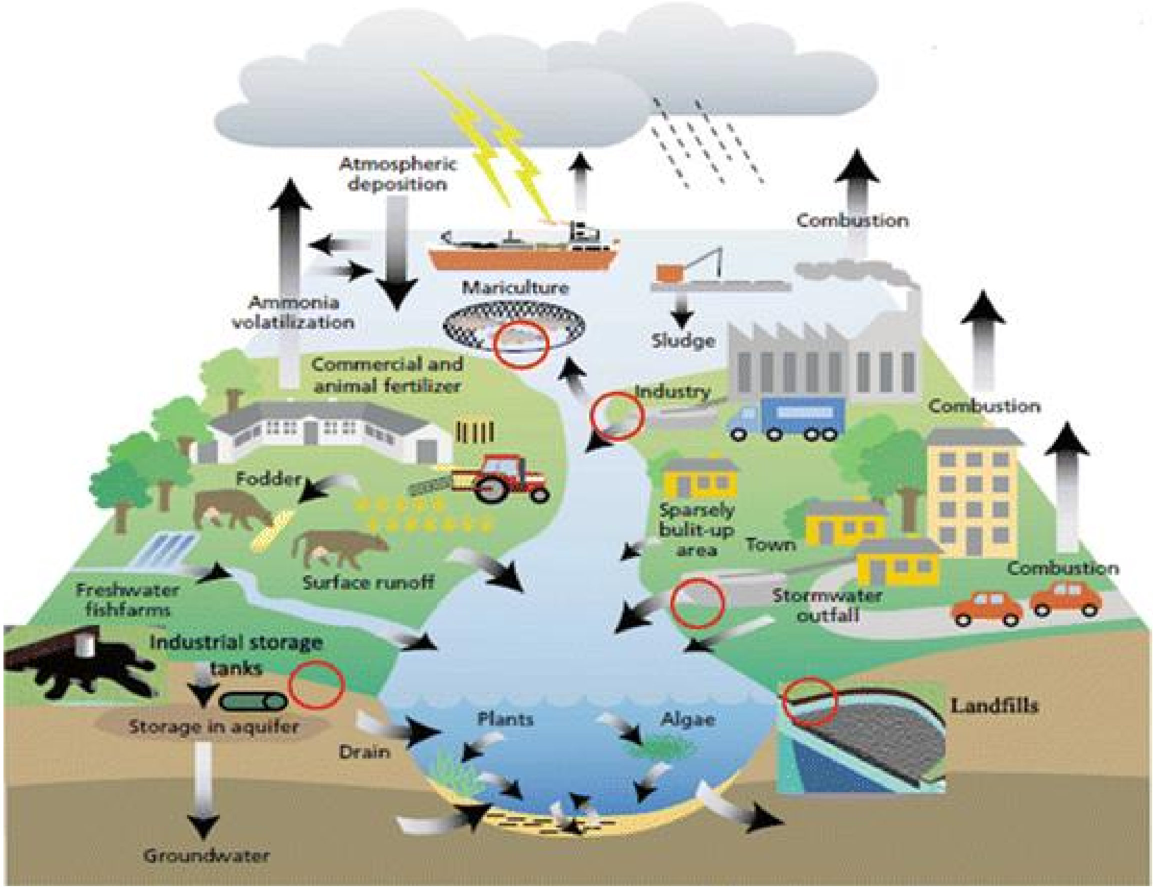


Figure 2.3: Point sources of metal contamination in the soil: Natural as well as anthropogenic factors or activities such as weathering of rocks, atmospheric deposition, industrial effluents, municipal waste etc., contaminate the soil environment with potentially toxic metals (Adapted from Armon and Starosvetsky, 2014)

2.2.2.1. Fertilizers and agro-chemicals

Commercial agro-chemicals and fertilizers are the most common source of metal contamination in the soils. Application of agro-chemicals and fertilizers containing small amount of PTMs as impurities, which may have come directly from parent rock, catalysts or reagents used, corrosion of equipment’s or from the materials such as gypsum, kaolin, limestone etc., which were added as fillers, conditioners and coatings during commercial preparations. Though, these ingredients may have proved beneficial to the metal deficient soils for plant growth and nutrition but excessive use of these agricultural products for better yield resulted into soil metal toxicity (Marrugo-Negrete *et al.* 2017; Liu *et al.* 2020). It was observed rock phosphate based fertilizers contained highest amount of PTMs followed by carbonate rock based fertilizers while nitrogen based fertilizers and potassium sulphate contained the lowest amount of various PTMs (Chen *et al.* 2007; Stacey *et al.* 2010; Vogel *et al.* 2015). Numerous studies reported the comprehensive information on the PTM content of several commercial fertilizers and the most abundant PTMs determined were Zn, Cd, Cr, Co, Mn, Mo, Ni, Be, Rb, Pb, Sr, Ti, V, As, and Sn while Bi, Sc and Hg are present in the minute quantity (Adriano, 1986; Gabe and Rodella, 1999; McBride, and Spiers, 2001; Molina *et al.* 2009; Uprety *et al.* 2009; Jiao *et al.* 2012). Such trace elements are commonly supplied by all the fertilizers and affected the soil health and composition after few applications, thus listed among the potential source of soil contamination and deterioration.

Agrochemicals, mainly pesticides including weedicides, fungicides, herbicides, insecticides etc., are extensively used in agricultural fields to control weeds, fungal infections, insects and diseases among various crops (Tirado *et al.* 2008; López *et al.* 2012). These agrochemicals mainly inorganic and organometallic compounds used in fruit orchards and vegetable fields, contained high amount of PTMs such as Cd, As, Pb, Cu, Co, Zn, Mn, Hg and Sn and contribute several kilograms of PTMs to the soil environment in few applications (Fenik *et al.* 2011; Fantke *et al.* 2012). For instance, copper sprays on some vegetables used for pathogen control enriched Cu metal in the soil by 11.2 to 16.8 kg/ha per year and led to various disease among humans (Adriano, 1986;

Stern, 2010; Laughton *et al.* 2019). Similarly, a single application of lead arsenate add up to 2.3 kg/ha of Pb and 0.5 kg/ha of As into the soil environment (Purves, 1977; Codling *et al.* 2015). The use of mercurial fungicides also results in contamination of soils with excessive Hg (Lin *et al.* 2012).

2.2.2.2. Municipal wastes, animal and agro-food wastes, sewages sludge and effluents

Rapid urbanization and industrialization resulted into increased amount of waste produced from increased urban activities such as household effluents, automobile related effluents such as service stations, fuel, atmospheric deposition, industrial effluents etc (US EPA, 1986, Sorme and Lagerkvist, 2002). Due to their richness in organic matter and plant micro and macro nutrients, these wastes were used as soil amendments depending upon their fertilizer value, type and concentration of trace elements present in that particular waste. In recent years, application of these wastes as soil amendments has become very popular because of necessity to manage them due to their everyday increasing amounts, costly treatment alternatives, their potential soil fertilizing capacity and protection of atmosphere and water bodies from pollution (Yuvaraj *et al.* 2020). In spite of various merits, soils amended with such wastes become highly contaminated with PTMs which further rendered the soil unsuitable for agriculture due to excessive phytotoxicity, lesser yield and degraded food products which in turn increased the dietary exposure of these PTMs to other members of food web specially animals and human beings (da Silva Oliveira *et al.* 2007). After effects of applications of sewage sludge and urban waste was also reported in soil microbial biomass and their activities (Parat *et al.*

2005). For instance, an increase in the total and viable bacterial population was observed initially in the sludge amended soils in comparison to the arable soils which further declined with the course of time (Mitchell *et al.* 1978). The concentration of PTMs in treated and composted urban waste depends upon the quality of the parent material and the technique used in treatment and composting which is chosen purely based on physical and chemical characteristics of the raw material constituents (Smith, 1992). The load of PTMs in sewage sludge and municipal waste also depend upon the point sources of the contamination present in a particular region, for instance PTM load in sewage waste from industrial area would be quite high in comparison to the PTM load from household waste (Paradelo *et al.* 2011). In addition to these point sources of contamination, wastes such as e-waste, paints, ceramics, and plastics also contributed to the metal load of the sewage sludge. Presence of high level of Cu, Pb, Zn and Cd was noted in sewage sludge and wastewater which could also be due to corrosion of canalization (Stofella and Mateo Box, 2004). It has been observed that wastes from municipal solid refuses also showed the presence of high concentrations of Zn^{2+}, Pb^{2+}, Cu^{2+}, Mn^{2+} and Cr^{2+} followed by lower amounts of Ni^{2+}, B^{2+} and Mo^{2+} and the lowest concentrations of As^{2+}, Co^{2+}, Cd^{2+} and Hg^{2+} (Qui *et al.* 2010). Although, the trend of PTM contents present in composted municipal refuses may resembles that of sewage sludge, still a much higher range of variability and much higher maximum levels were observed in sewage sludge and wastewater in comparison to the municipal refuses due to different point source of contamination (Díaz-Cruz *et al.* 2009; Qui *et al.* 2010). Thus, elevated dressings and irrigation of rural and urban soils with sewage sludge, municipal waste and waste water for a prolonged time led to a pronounced effect on PTM content and potential contamination of the soil site.

2.2.2.3. Application of manures and other animal wastes

Cattle and cow dungs manures from farmlands have been used traditionally on the agricultural sites due to their fertilizer value, presence of high contents of organic matter and macro and micronutrients. Despite the beneficial effects of using manures and other animal waste in improving soil fertility and soil characteristics, presence of excessive amount of PTMs such as Zn, Pb and Cu in these materials affected the soil fertility in a negative way (Achiba *et al.* 2010). The major source of these PTMs in the manure and other agricultural products are cattle, poultry and swine feeds which are often fortified with various trace metals in order to carry various physiological processes and to avoid any health disorders among animals (Sistani and Novak, 2006). Various types of PTMs found in manures and animal wastes are As, B, Ba, Be**,** Cd, Co, Cr, Cu, F, Hg, Mn, Mo, Ni, Pb, Rb, Sb, Sc, Se, Sn, Sr, Ti, V and Zn (Senesil *et al.* 1999). These PTMs enter into the soils through animal waste based manures or waste water. Despite contributing a very low concentration of PTMs to the soil, it can lead to toxicity among plants and soil contamination. For instance amendments derived from wastes of swine and poultry fed with Cu^{2+} and Zn^{2+} rich diets would add 3 to 6 kg Cu/ha per year to the soil (Berryman, 1965). Generally, plants have a tendency to assimilate a little amount of PTMs but when the PTMs application rate in a crop field exceeds the plant's removal rates then this led to accumulation of these metals in the soil (Ko *et al.* 2008). It was reported in a research that 20% of cattle grazing in animal manure amended fields had liver Cu^{2+} concentrations exceeding the potential toxic concentration of Cu^{2+} i.e., 150 mg/kg fresh weight (Sistani and Novak, 2006). Such soils with such high concentrations of PTMs can also be a point source of contamination for ground water and surface water bodies and cause phytotoxicity (Bolan *et al.* 2004).

2.2.2.4. Soil irrigation with natural freshwaters and wastewaters

Worldwide over exploitation of groundwater resources and regular dumping of industrial effluents into soil and surface water resulted into an anomalous distribution of PTMs in the water resources (Ramesh *et al.* 1995). The level and occurrence of PTMs in a water resource depends upon the nature and composition of source water and various activities occurring in its proximity e.g. industries, mining or agriculture. Various anthropogenic activities such as manufacturing and processing of consumer goods, use of fertilizers and agrochemicals, corrosion of the storage and conveyance amenities in a community and inputs from domestic activities like everyday consumed food items and goods and other human wastes significantly raised the level of PTMs in a water resource (Sistani and Novak, 2006). A consistence high concentration of PTMs in wastewaters is a strong indication of discharging industrial waste into municipal wastewater treatment plants Surface and ground waters resources used for irrigation of agricultural fields contained variable concentrations of several PTMs and excessive concentration results into phytotoxicity (Chudaeva *et al.* 2011). The actual level of PTMs in wastewater may also vary depending upon the treatment techniques and time, sometimes even within a particular treatment plant (Sistani and Novak, 2006). Levins and Gosk (2007) reported the increased level of PTMs such as Cu, Co and U in water table aquifers under the agricultural lands of Latvia which indicated the contamination of ground water with PTMs due to day by day increasing anthropogenic activities (Levins and Gosk (2007). Similarly, accumulation of huge amount of Cu^{2+}, Fe^{2+}, Pb^{2+} and Zn^{2+} was observed in due to prolonged irrigation of agricultural soil with sewage waste water in Egypt (Abdel Sabour *et al.* 1995). Potentially toxic metals and their associated elements formed complexes and precipitates or concentrated at several places depending upon the environmental parameters thus increased the level of metal contaminants in the soil environment (Ramesh *et al.* 1995).

2.2.2.5. Atmospheric deposition

Atmospheric deposition of PTMs takes place through both natural phenomena as well as anthropogenic activities (Lazor *et al.* 2020). Among the natural sources, ashes from volcanic eruptions, weathering of rocks and agitation of earth surface by air turbulence are important sources of atmospheric deposition of PTMs. Atmospheric deposition through anthropogenic activities is quite high in comparison to the contribution from natural ways and introduced new and previously non-existent impurities to the atmosphere (Page and Chang, 1979). The principal anthropogenic sources responsible for atmospheric deposition of PTMs were gases and aerosols entering the atmosphere due to combustion of fossil fuels and woods, mining and smelting of metal ores, cement industry, incineration of refuses such as municipal refuses and sewage sludge, automobiles exhausts (Loppi *et al.* 2000; Rauch *et al.* 2005; Saha *et al.* 2019). All these processes either involved fuel combustion process or high temperature conversion industrial processes involving PTMs which further led to release of volatile substances carrying PTMs or their oxides into the atmosphere (Hernández-Pellón *et al.* 2019). Potentially toxic metals such as As, Be, Cd, Mn, Cr, Ni, Pb, Sb, Ti, V and Zn were reported to be present in exhausted combustion flue gases. These airborne PTMs were transported to the long distances through winds and deposits on land and water sources and distance of transport varies depending upon various factors such as emitting source, physical and chemical properties of these PTMs compounds at the time of transport, and meteorological conditions of the atmosphere etc (Sistani and Novak, 2006). Consequently PTMs enriched fine particulates which could not be easily captured by emission control devices would be emitted and probably remain suspended in the atmosphere for long time.

2.2.3. Cellular mechanisms involved in metal tolerance

Elevated concentrations of both essential and non-essential metals may lead to toxicity among living organisms and these toxicity symptoms may be due to various cellular or molecular levels of interactions within the living organisms. Most probable reasons for metal toxicity inside a cell is binding of PTMs to the sulphydryl group of proteins leading to inhibition of protein activity and displacement of essential elements by these toxic metals further results into disruption of structure (Van Assche and Clijsters, 1990) or formation of free radicals or ROS resulting into oxidative stress (Dietz *et al.* 1999). However, some of the living organisms have developed different mechanisms and phenomena which aid in surviving and adapting such adverse conditions of metalliferous soil through maintenance of homeostasis. Several potential mechanisms and strategies have been reported to be involved in detoxification and tolerance towards metal stress by avoiding the accumulation of toxic concentrations of metals within the cell hence preventing the harmful effects of such PTMs. Such strategies include extracellular networks such as cell wall; extracellular exudates while intracellular strategies involving plasma membrane which reduces metal uptake or facilitate the efflux of metals entered in cytosol. Several repair mechanisms are also present inside the cell for repairing the damage caused due to toxicity and is preferred over synthesis of specific proteins with a potential to resist metal toxicity (de Vos *et al.* 1991; Dietz *et al.* 1999). Different mechanisms such as chelation through organic acids, amino acids, compartmentalisation of these toxic ions into vacuoles to avoid disturbance in normal cell functioning or synthesis biomolecules such as heat shock proteins (Hsps), metallothioneins (MT) also occur inside the cell (Kumar *et al.* 2016).

2.2.3.1 Organic acids and amino acids

Organic acids such as carboxylic acids (especially mono and dicarboxylic acids) and amino acids such as citric acid, maleic acid and histidine are potential ligands for metals and reported to play a role in providing tolerance towards toxicity caused by PTMs (Clemens, 2001, Galal and Shehata, 2015). Organic acids exudates from plants may provide a proton source for solubilisation of metals and result into acidification of soil (Devêvre *et al.* 1996; Fomina *et al.* 2005). For instance, solubility of Cd^{2+} in the soil increases with the formation of metal-carboxylic acid complexes like acetic acid and succinic acid present in wheat rhizospheric soil results into accumulation of Cd^{2+} ions. However, absence of strong correlation between the exposure of metals and the amount of organic acid exudates secreted, does not support it as general tolerance mechanisms among the living organisms (Meharg, 2003). For example, in *Triticum caerulescens,* no relation was observed between hyperaccumulation of PTMs and root exudates whereas some studies have observed that microbial interactions in the roots of *T. caerulescens* changes the soil conditions such as pH, results into increase in the solubility of Zn^{2+} metal hence hyperaccumulation of metal ions in the plant (McGrath *et al.* 1997; Zhao *et al.* 2001).

Similarly metal chelation by amino acids especially proline and histidine, have been reported in cells and xylem sap (Rai, 2002; Sharma and Dietz, 2006). Proline was reported to accumulate in microalgae on exposure to Cd metal and works by reducing the formation of the free radicals and enhancing GSH level to maintain the reducing environment (Siripornadulsil *et al.* 2002). Similarly, increase in histidine content was observed in Ni^{2+} hyper-accumulating plant *Alyssum lesbiacum* in order to combat the Ni^{2+} metal toxicity and reported to increase tolerance of non-accumulating plant species grown in its proximity towards Ni^{2+} metal as well (Persans *et al.* 1999). Another example is aminocarboxylate which is an amino acid derivative synthesized from S-adenosyl-L-methionine and known to be involved in translocation of minerals (Stephan and Scholz, 1993). Apart from its chelating action, this amino acid derivative has been reported to be a precursor of phytosiderophore mugineic acid which is involved in binding metal ions like Zn^{2+}, Cu^{2+} and Fe^{2+} (Treeby *et al.* 1989).

2.2.3.2. Metallothioneins

Metallothioneins (MTs) are small cysteine rich metal binding peptides that are ubiquitous in eukaryotes and cyanobacteria. Divergent MTs have been reported among the living organisms such as blue green algae, algae, fungi, higher plants and animals. Metallothioneins are categorised into two different classes: Class I MTs and Class II MTs. Class I MTs are known to possess cysteine residues aligning with an equine (mammalian) MT while class II MTs have similar cysteine residues but not aligning with class I MTs (de miranda *et al.* 1990; Robinson *et al.* 1993; Prasad, 1999). Various species have been reported to possess wide range of MT gene families and more than one class of MTs. For instance, in *Arabidopsis thaliana*, MT3 and MT4 type of MTs have been reported along with Class I and Class II MTs (Goldbrough, 2000). Metallothioneins are known to protect living organisms from metals both noxious like Cd, Hg and nutrient elements like Cu, Zn (if present in more than the threshold limit) because of their metal binding ability and induction properties (Hall, 2002). Along with metal ion chelation, MTs are also known to provide antioxidant protection mechanisms as well as plasma membrane repair to the native organisms (Figure 2.4) (Hamer, 1986). All sulfhydryl groups present in MTs aid in metal binding with the capacity of 7 gram-atoms/mole protein (Klaassen and Lehman-Mckeeman, 1989) and maintains a redox potential by actively binding and releasing metals under physiological conditions (Wang *et al.* 2016). For example, a 67 amino acid MT isolated from a marine brown algae *Fucus vesiculosus* reported to possess the property of tolerating toxic concentration of both Cu^{2+} and Cd^{2+} metal *in vitro* (Morris *et al.* 1999). In lower plants such as *A. thaliana*, MT2a gene from wild type strain was reported to be strongly induced in presence of Cu^{2+} in Cu sensitive *Arabidopsis* mutants and accumulated high concentration of Cu^{2+} inside the plants (van Vliet *et al.* 1995). Similar observations were made when MT gene from *Brassica juncea* were conferring tolerance in *A. thaliana* towards Cd and Cu toxicity.

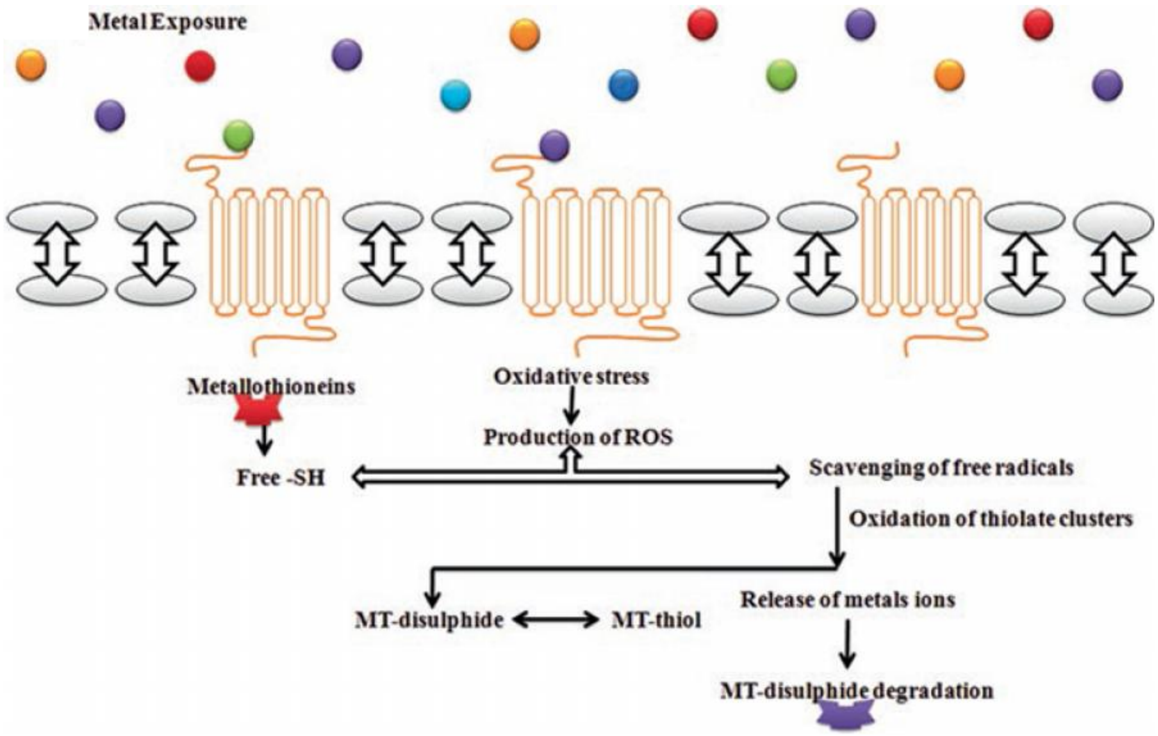


Figure 2.4: Cellular homeostasis maintained by metallothioneins (MTs). Potentially toxic metals (PTMs shown in colored dots) when entered into the cell resulted into production of reactive oxygen species (ROS) which further results into induction of metallothioneins (MTs). PTMs bind to sulfhydryl group of MTs to form a MT-PTM complex and maintain redox status by reducing the availability of diffusible forms of these metals within cells and therefore decreasing their toxic potential (Chaudhary *et al.* 2018).

Expression of MT genes varies during different developmental stages of plant as well as under varying environmental conditions. In higher plants such as *Bactris campestris* and *Nicotiana tabacum,* metal tolerance was reported to be induced by transforming these plants with a mammalian MT which further conferred Cd tolerance by minimizing the Cd translocation from roots to shoots through MT-metal complex formation (Hall, 2002; Pal and Rai, 2009). In 2008, Ramesh et al. reported two MTs from ECM fungi *Hebeloma cylindrosporum* which showed tolerance towards a range of PTM ions such as Cd^{2+}, Ni^{2+},

Pb^{2+}, Zn^{2+} and Cu^{2+}. Similarly, Ziller et al. (2017) reported five cysteine rich proteins classified as MTs like proteins isolated from soil through metatranscriptomics and reported to play significant role in maintaining the homeostasis during metal toxicity. Henceforth, these studies signified the role of these cysteine rich MT proteins in providing metal tolerance to almost all the living organisms in contaminated environment (Bellion *et al.* 2006; Kalsotra *et al.* 2019)

2.2.3.3 Phytochelatins

Phytochelatins (PCs) are enzymatically synthesized proteins possessing a general structure (c-Glu Cys)$_n$-Gly where n ranges from 2 to 11 and belongs to the family of metal complexing peptides induced by the presence of PTMs (Rauser, 1995; Hall, 2002). PCs are not only present in plants; they are also reported in other living organisms such as fungi and other organisms (Gekeler *et al.* 1988; Piechalak *et al.* 2002). PCs are synthesized by transpeptidation of Glu-Cys moiety of glutathione molecule and the reaction is carried out by enzyme PC synthase which is activated in the presence of metal ions (Cobbett, 2000). The genes encoding enzyme PC synthase has been reported in *Arabidopsis* and yeast (Clemens *et al.* 1999; Vatamaniuk *et al.* 1999). PCs have potential to bind different metals including Cu, Zn, Cd or As through sulfhydryl and carboxyl residues and its biosynthesis is regulated by glutathione molecule which is a precursor in PC synthesis (Olsson *et al.* 2015). It is established fact that metals induce PC synthesis both *in-vivo* and *in-vitro* continuously until the complete chelation of activated metal ions present (Loeffler *et al.* 1989). Chelation of metal ions by PC involves complex mechanisms where PC synthase gets activated by metal ions results into PC biosynthesis followed by formation of complex and sequestration in vacuole and lastly more complex formation with sulfide or organic acid in the vacuole hence detoxification (Figure 2.5). PCs are also known to play significant role in maintain the metal ion homeostasis hence regulate the metal ion availability among plant cells (Guo *et al.* 2008).

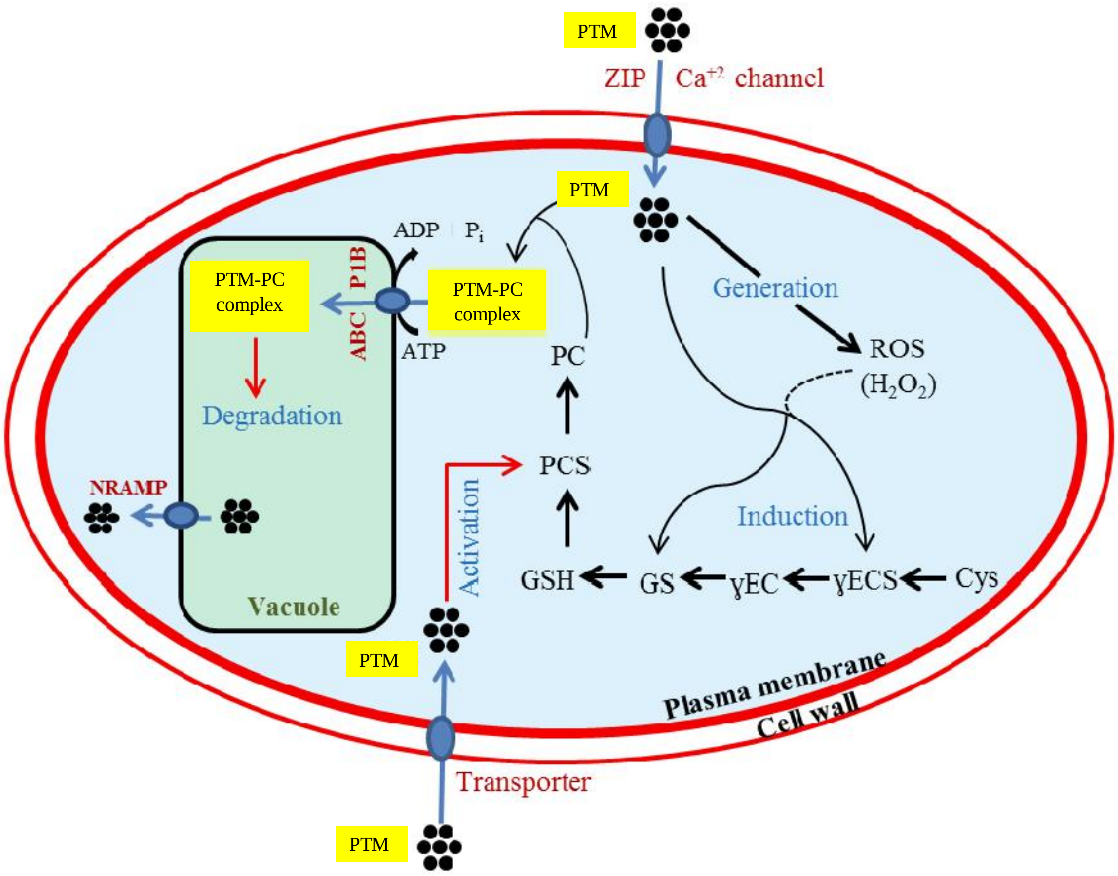


Figure 2.5: Metal detoxification by phytochelatins (PC) in a plant cell. Potentially toxic metals (PTM shown in black dots) activates phytochelatin synthase (PCS) which further results into formation of PTM-PC complexes that are consequently transported through tonoplast to the vacuole by ATP-binding-cassette (ABC) and ABC-P1B transporter. PTM also transported inside the cell through plasma membrane via ZIP and vacuolar transport of PTM is mediated by natural resistance associated macrophage protein (NRAMP). (Figure is adapted and modified from Anjum *et al.* 2015).

Over-expression of PC synthase genes in some lower organisms like bacteria and fungi significantly increases their potential towards metal tolerance whereas in higher plants the metal tolerance is not always contributed by PC synthase over-expression. Hence, it is presumed that detoxification of PTMs through PCs might have first evolved as defence mechanism in lower plants and then in higher plants. Similarly, Xiang and Oliver, (1998) observed that the genes involved in glutathione biosynthesis show tremendous increase in the rate of transcription in the presence of PTMs specifically for those which are supposed to be detoxified by PCs such as Cu and Cd. In 1995, Howden *et al.* reported the correlation between the amounts of PCs accumulated inside Cd- sensitive mutants of *Arabidopsis* and the degree of sensitivity of the plant towards Cd metal. Inhibition of PCs due to biosynthesis of glutathione further confirmed the active role of PCs in detoxification of PTMs among algae, fungi and higher plants (Yang and Chu, 2011). Furthermore, Zhu et al. (1999) reported the increased level of glutathione and PCs in response to Cd metal toxicity due to over-expression of gene γ-glutamylcysteine synthetase from *E. coli* in *Brassica juncea*. It was observed that the level of PCs synthesized in response to Cd accumulation is sufficient to chelate all the Cd metal absorbed and protect the process of photosynthesis in the plant (Haag-Kerwer *et al.* 1999; Zhenyan *et al.* 2005).

2.2.3.4. Heat shock proteins

Heat shock proteins (Hsps) belong to the class of molecular chaperons generally known to be functional in folding and assembly of proteins. It has also been reported in numerous researches that Hsps are also involved in protecting the living cells by preventing the denaturation of proteins or activating the repair mechanisms and protection of other proteins under stress conditions such as high or low temperatures, insecticides, dessication and PTMs (Figure 2.6) (Vierling, 1991; Steinert, and Pickwell, 1993; Lewis *et al.* 2001; Burges *et al.* 2015). Hsps were first discovered in 1962 as chromosome puffs in *Drosophila* after exposure to high temperature conditions hence named as heat shock (Ritossa, 1962). Hsps are found in all living organisms and named on the basis of molecular size such as Hsp70 is named after its molecular weight ranges between 66-78 kDa. Chief families of Hsps are Hsp100, Hsp90, Hsp70, Hsp60, Hsp40 and the small Hsps which are below 30 kDa size. Among these, Hsp70 is believed to be chief Hsp family with characteristics of solely inducible, constitutive and inducible, and solely constitutive proteins. It has been reported in numerous researches that Hsps like Hsp70, Hsp80, Hsp40 and Hsp25 are activated in presence of PTMs (Neumann *et al.* 1994; Zhen *et al.* 2007; Lin *et al.* 2014; Ireland *et al.* 2004). Increase in metal stress or heat stress, increases the expression level of mRNA of Hsps as observed in rice (low Hsps 16-20 kDa) (Tseng *et al.* 1993) and in *Armeria maritima* (Hsp17) for Cu metal stress (Neumann *et al.* 1995). Cell culture of *Silene vulgaris* and *Lycopersicon peruvianum* also showed increase level of mRNA expression of low molecular weight Hsps such as Hsp17 in response to a range of toxic metals (Wollgiehn and Neumann, 1999), however when observed in plants growing on metalliferous soils very low or no Hsps were found in the plants proposing that Hsps plays no role in heritable metal tolerance of *Silene* on metalliferous soil. Hence, it is suggested that Hsps are important part of response towards sudden exposure of living organisms to extreme stress which is of great ecological relevance with respect to less severe but regular incidence of stress.

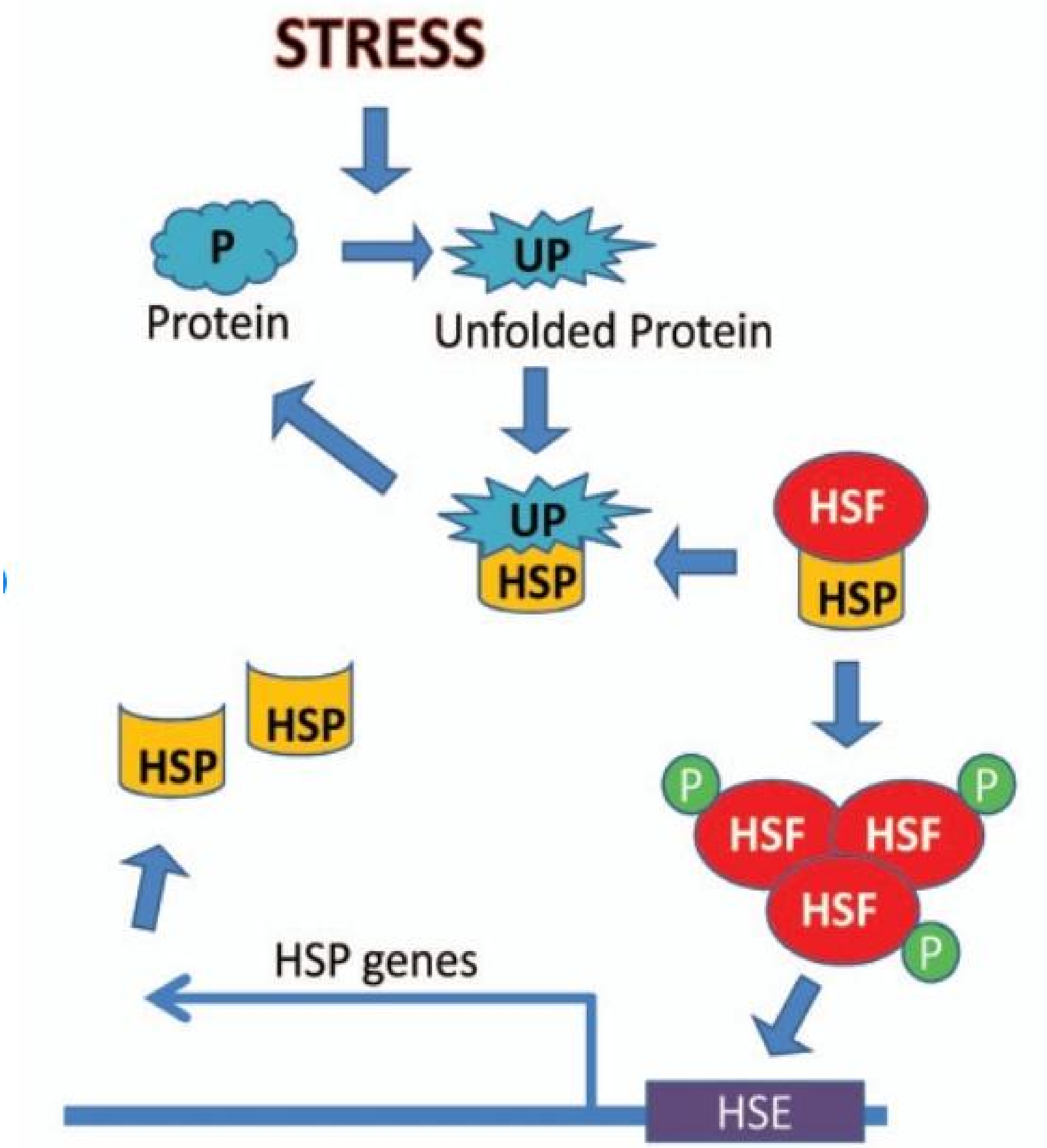


Figure 2.6: Network of heat shock proteins (Hsps) active in response to abiotic stress. Adverse conditions such as metal toxicity resulted into synthesis of misfolded proteins inside a cell. As a consequence, Hsps dissociate themselves from heat shock transcription factor (HSF), released activated HSF then bind to heat shock element (HSE) and lead to up-regulation of Hsps in the cell which will aid in reforming these misfolded proteins (Rowley *et al.* 2011).

Hsps are also used as biomarkers to detect the disturbances caused in the environment due to stress conditions and the advantage of using Hsp as biomarker is that the disturbances can be detected way before in comparison to the biomarkers based on growth rate, mortality or fertility (Werner and Nagel, 1997). Induction of Hsp as biomarkers for cellular stress has been investigated in various soil and marine organisms (Köhler *et al.* 1992; Köhler and Eckwart, 1997).

2.2.3.5. Cell wall and the root exudates

Although the role of cell wall and metal binding properties in metal tolerance is not clearly defined still few reports have mentioned the role of cell wall in providing tolerance towards metal toxicity as cell wall of root cells is in direct contact with the metals present in the soil and the absorption of these metals onto cell wall is of very limited quantity because of impermeable nature of cell wall due to peptidoglycan. It has been reported in *Silene vulgaris* that metals get accumulated, either bound to proteins or as silicates in the cell wall of epidermal cells (Bringezu *et al.* 1999). Another part of roots that is known to play significant role in metal tolerance is root exudates which acts as metal chelator to trap certain metals (Marschner, 1995). The presence of Ni^{2+} chelating histidine and citrate in the root exudates of Ni^{2+} non-hyperaccumulating plants helps the plants to reduce Ni^{2+} adsorption on to the roots cells hence played a role in Ni detoxification (Salt *et al.* 2000). As root exudates have variety of metal chelating compounds that could make it more significant to confer tolerance to various other metals present in the soil solution. For instance, the roots of buckwheat plant secretes oxalic acid in response to Al^{3+} metal present in the soil and accumulates Al^{3+} in the form of non-toxic Al oxalate complex in the leaves (Ma *et al.* 1998). In 2001, similar findings were reported for maize and wheat plants and inferred that such root exudates are mediated by Al^{3+} activated anion channels in the plasma membrane thus detoxifying both internally and externally (Ma *et al.* 2001). Similarly, enzymatic secretions from the roots such as root reductases help in maintaining ionic content inside the plant cells by reducing metal ions such as Fe^{3+} and Cu^{2+} for increasing uptake of Fe^{3+}, Mn^{2+}, Cu^{2+}, and Mg^{2+} metal ion (Welch *et al.* 1993). For example ferric reductases from *Arabidopsis* and maize plants maintain Fe^{3+} content inside the plants (Yang and Roemheld, 1999).

2.3. Approaches used to study soil microbiome

2.3.1 Culture dependent approaches

Culture dependent approaches involve isolation and cultivating microbes present in a particular habitat or environment using specific growth media and growth parameters depending upon target organisms in the study. Classic microbiology requires pure culture of an organism to study the characteristics of that organism such as genetics and physiology in detail hence culture dependent approaches overlook majority of microbial diversity flourishing in an environment. Around 0.1% to 1% of total microbial species present in soil are cultivable under any condition (Torsvik *et al.* 1996; Torsvik and Ovreas, 2002). One major limitation of culture dependent approaches is dominance of fast growing microbial taxa present in soil environment on a nutrient rich media which will automatically exclude majority of other slow growing but significant species. Other techniques such as single cell sequencing and mini-metagenome approaches tried to bridge the gap left between classic culture dependent approaches and advanced culture independent approaches but these practices are not yet fully developed (Hutchison and Venter, 2006; McLean *et al.* 2013). Similarly using “great plate count anomaly” as an isolation technique give two different observations, one is observed under microscope and another one is those which are able to grow on media (Staley and Konopka, 1985). Though by providing nutrient limited media and longer incubation time, cultivation of some novel soil microbes already discovered using molecular techniques, have been reported such as *Acidobacteria, Verrucomicrobia* and others (George *et al.* 2011, da Rocha *et al.* 2010; Davis *et al.* 2011; Stewart, 2012). Bomar et al. (2011) successfully cultured the dominant gut microbiota from nematode by determining the key carbon sources metabolized by them after numerous failed attempts of culturing.

Another key factor in the survival of microorganisms is temperature which influences both microbial viability and growth. Microorganisms tolerate an extensive range of temperature in their natural habitat while classic incubation temperatures for isolation in the lab are between 27°C to 37°C hence providing slightly lesser or higher temperatures in comparison to the isolation site can help in finding of new microbial species. Environmental sites with extremely high or low temperatures such as hot water springs and glaciers are habitats for extremophiles and providing such incubation conditions to isolate such extremophiles through traditional methods is difficult. Hence, isolation of novel microbes and determination of total diversity present in a complex environmental sample such as metal polluted soil need more sophisticated and accurate methods.

2.3.2. Culture Independent Approaches

2.3.2.1. Genetic fingerprinting

Genetic fingerprinting is a technique that identifies organisms at different taxonomic groups on the basis of variation in target DNA sequences. This technique involves amplification of marker gene from the environmental DNA sample by polymerase chain reaction (PCR) followed by restriction digestion of amplified product to screen out the novel amplified fragments. Digested product is then separated on agarose gel and different bands observed under UV illuminator give an insightful image of community structure present in an environmental sample. Denaturing gradient gel electrophoresis (DGGE) and 5' terminal restriction fragment length polymorphism (TRFLP) are the most used techniques of genetic fingerprinting (Liu *et al.* 1997). Another technique to interpret the diversity of an environmental site is to target variable region of DNA sequence for amplification such as automated ribosomal intergenic spacer analysis (ARISA) involves amplification of internally transcribed spacer (ITS) region that lies between 16S and 23S rRNA genes in prokaryotes and 18S and 5.8S rRNA genes in eukaryotes (Garcia-Martinez *et al.* 1999). Advantage of ARISA is that it sidesteps the need of additional treatments for amplified product and allows the measurement of fragment size and abundance which can further help to generate graphs on the basis of principal component analysis (PCA) or multidimensional scaling (MDS) to predict community arrangement or differences between the communities present in an environment. Fragmented bands which we assume to be different between the communities are sequenced after gel extraction to identify the respective organism. These genetic fingerprinting techniques have broadly used to study various environments such as rhizosphere with 16S/18S rRNA and ITS gene or the cDNA derived from 16S/18S rRNA or ITS gene (Lueders *et al.* 2004; Costa *et al.* 2006; Murase and Frenzel, 2007; Haichar *et al.* 2008; Fechnar *et al.*

2010; Tkacz *et al.* 2013; Konstantinidis and Tiedje, 2007). Several works have also reported the use of other phylogenetic markers and functional genes to study the community structure of an environment (Haichar *et al.* 2012). Though, genetic fingerprinting is robust platform for various environmental samples but at the same time it comes with major drawbacks such as PCR biasness due to primers and amplify genetic material irrespective of viability of that living being and hindrance in PCR amplification due to various impurities present in samples such as humic acids in case of soil (Opel *et al.* 2010; Venter *et al.* 2006; Fechnar *et al.* 2010).

2.3.2.2. Metagenomics

Metagenome is term applied to the total genome content of both prokaryotes and eukaryotes present in an environment and carry wide genetic information in comparison to the culturable subset (Handelsman *et al.* 1998). Metagenome of an environment is considered as a rich source of various metabolically active genes, identification and characterisation of which can revolutionise the field of biotechnology. Given the profound efficacy and significance of microorganisms present in an environment, an approach is needed to explore the hidden treasure of information within the metagenome (Sharma *et al.* 2005). Hence, metagenomic approach which involves genomic analysis of microbial communities provides a glance into the wide variety of unexplored life present in an environment thus offering a new insight into the structural and functional aspects of ecosystem as diverse as soils and oceans (Handelsman, 1998, Marmeisse *et al.* 2017). Metagenomics was first and foremost approach conceived in 1990s with various goals in mind to know about the contributions made by unknown community members of an environment (Stein *et al.* 1996; Handelsman *et al.* 1998). Metagenomics have demonstrated as an influential approach for exploring the ecology (Biddle *et al.* 2008) and metabolic profiling of complex environmental microbial communities (DeLong *et al.* 2006; Tringe *et al.* 2005) as well as for identifying the novel biomolecules by screening a library constructed from isolated environmental DNA (Daniel, 2005; Handelsman, 2004; Ferrer *et al.* 2009; Steele *et al.* 2009; Simon and Daniel, 2010). Metagenomic approach involves direct isolation of DNA from environmental samples thus sidesteps the need of isolation and cultivation of microorganisms (Figure 2.7) (Rodriguez and Konstantinidis, 2014). In 1985, Pace *et al.* were first to suggest the direct cloning of environmental DNA into a suitable vector and host to unveil numerous hidden treasures in an environment. Further, in 1991, Schmidt *et al.* used this technique of direct cloning of environmental DNA to analyse the marine picoplankton community by cloning 16S rRNA gene sequences into phage vector. First and foremost step in construction of these metagenomic libraries is isolation of high quality DNA covering the genome content from majority of the microbial diversity present in the original sample. DNA isolation from some of the environmental samples such as soils (Hårdeman and Sjöling, 2007; Pathak *et al.* 2009), hot springs (Rhee *et al.* 2005), Antarctic desert soil (Heath *et al.* 2009) and buffalo rumens (Duan *et al.* 2009) is quite daunting task as many microorganisms like eukaryotes are quite reluctant to cell lysis which results into development of numerous protocols for isolation. High quality DNA is further used to synthesize libraries which are classified into two groups on the basis of average insert size: small insert libraries carrying insert size less than 15 kb into plasmid vector whereas large insert libraries insert size up to 40 kb with cosmid and fosmid vector while large inserts of size greater than 40 kb with BAC vector. The choice of vector system to be used in the study depends upon quality of isolated DNA, insert size, vector copy number required, host and screening strategy to be used in the study (Daniel, 2005). In 1999, Rondon *et al.* analyse the genes from *Bacillus cereus* by synthesising a library and expressing them in *E. coli* through bacterial artificial chromosome (BAC) vector. BAC vector possess the ability to accommodate large sized DNA inserts even greater than 100 kb hence facilitate the cloning of large DNA fragments (Shizuya *et al.* 1992). Shuttle vectors such as cosmids and BAC can be used to transfer libraries from bacterial host such as *E. coli* to other hosts such as *Streptomyces* (Courtois *et al.* 2003; Martinez *et al.* 2004). Sequencing of these vectors carrying inserts helps in identifying the characteristics of insert and gives us knowledge about the taxonomic origin of organism and identification of organism. It has been reported that such studies identified a wide range of microbial taxa, functional genes as well as numerous gene products with industrial or ecological significance such as antibiotics and enzymes (Rondon *et al.* 2000; Rani *et al.* 2008; Jiang *et al.* 2009, Donato *et al.* 2010).

Metagenomic approach has been explored in two different ways to analyse the genomic diversity thus the metabolic potential possessed by the microbial community: (i) Sequence based and (ii) function based approach. Both these approaches involve direct isolation of environmental DNA, its ligation into suitable vector. Sequence based metagenomic approach involves the random shotgun sequencing followed by analysis through bioinformatics tools. Sequence based metagenomics has led to the discovery of more than 600 novel putative glycosyl hydrolases from hindgut of termites with the help of bioinformatics tools (Warnecke *et al.* 2007). Identifying novel proteins based on sequence homology has an advantage that the initial identification step does not need successful expression or cellular machineries to facilitate the correct folding of recombinant proteins. Function based approach of metagenomics involves screening of metagenomic library on the basis of specific functions or characteristics they possess and enable the screening of several clones on a single selective media agar plate. The function-driven approach of metagenomics was first demonstrated by Healy *et al.* through screening of cellulose encoding functional genes from metagenomic libraries synthesized from various microbial communities flourishing in thermophilic anaerobic digester (Healy *et al.* 1995). This approach has led to the discovery of several novel enzymes such as esterases, lipases and alcohol dehydrogenases (Ferrer *et al.* 2009; Rashamuse *et al.* 2009). Function based approach of metagenomics has been proved to be the more efficient, more targeted and cost effective approach in comparison to the random shotgun DNA sequencing. Functional based approach has its own barriers such as no or less availability of suitable activity assays for high throughput screening. The main advantage of functional screening of metagenomic library is it allows the discovery of novel genes which would be missed by purely bioinformatics analysis of sequence based approach.

During initials years, metagenomics principally explored to mine novel biomolecules from the environment but as molecular biology proceeded towards the next generation sequencing techniques and other methods; it gave large scale analysis of microbial assemblages such as comparative community metagenomics, metatranscriptomics and metaproteomics (Chistoserdova, 2010; Sjöling and Cowan, 2008). The comprehensive data sets derived from these approaches with environmental parameters allows us to unravel ecological significance of complex ecosystem functions of microbial communities. Thereby, this new scientific approach provides admittance to the world of organisms which are recalcitrant to the traditional culturing methods and offers a first peek into the wide variety of life that has never been explored earlier thus giving a new insight into the structural and functional aspects of ecosystem as diverse as soils and oceans (Daniel, 2005; Handelsman, 2007).

2.3.2.3. Metatranscriptomics

DNA based metagenomics become a standard approach to study microbial community structure by sequencing DNA from random community present in environmental samples and subsequently enlighten their taxonomic origin and protein coding gene diversity. However, to investigate the activities of microbial communities in response to the rapid environmental changes such as metal pollution require the analysis of changes in the abundance and composition of active fraction of microbial communities. To elucidate how these microbial communities respond during adverse conditions, another approach i.e., metatranscriptomics is developed in which gene expression profiling of functional and taxonomic marker genes can be studied through community mRNA (Handelsman *et al.* 2007). Metagenomics provides a snapshot of the genetic composition of microbial communities in an environment at any given time interval whereas metatranscriptomics provides an insight into the potential expression of genes in environment at the time of sampling (Shi *et al.* 2011; Vorobev *et al.* 2020). Metatranscriptomics involves the use of high throughput sequencing techniques for community level transcriptional profiling of reverse transcribed RNA which overcomes a number of problems associated with microarrays by providing a direct access to the community transcriptional profiles (Carvalhais *et al.* 2013). Advances in high throughput sequencing techniques such as 454 FLX GS, Illumina Hiseq 2000, Illumina MiSeq platforms generate more magnitude of data in comparison to the conventional Sanger's sequencing techniques at a fraction of cost of older technologies (Bradley *et al.* 2016; Caporaso *et al.* 2012; Auffret *et al.* 2016; Khatri *et al.* 2017). Illumina MiSeq can process a number of samples (~400 samples) in a single run depending upon degree of automation available with a yield of 25 million 2 x 300 bp paired end reads per turn. This advanced high throughput technique has helped in investigating several microbial environments such as soils (Affuret *et al.* 2016), oceans (Fuhrman, 2009), human microbiota (Yatsunenko *et al.* 2012) and wastewater treatment system (Vanwolterghem *et al.* 2016). The data generated from this technique is further analyzed by typical computational steps i.e., filtering out poor quality sequence reads, trimming off sequencing adapters and barcodes, merging of paired end reads into single sequence on the basis of overlapping and assignment of sequences to the samples via using barcodes. Identical sequences are further clustered using an identity cut off and each cluster constitutes an "operational taxonomic unit (OTU)" which is further assigned to a taxon using a database and a classification algorithm. As the primary aim of this approach is to identify over and underrepresented genes present in the cDNA library synthesized from reverse transcribed RNA therefore require a reliable public database which can facilitate the comparison of relative frequencies of expressed genes (Dong *et al.* 2017).

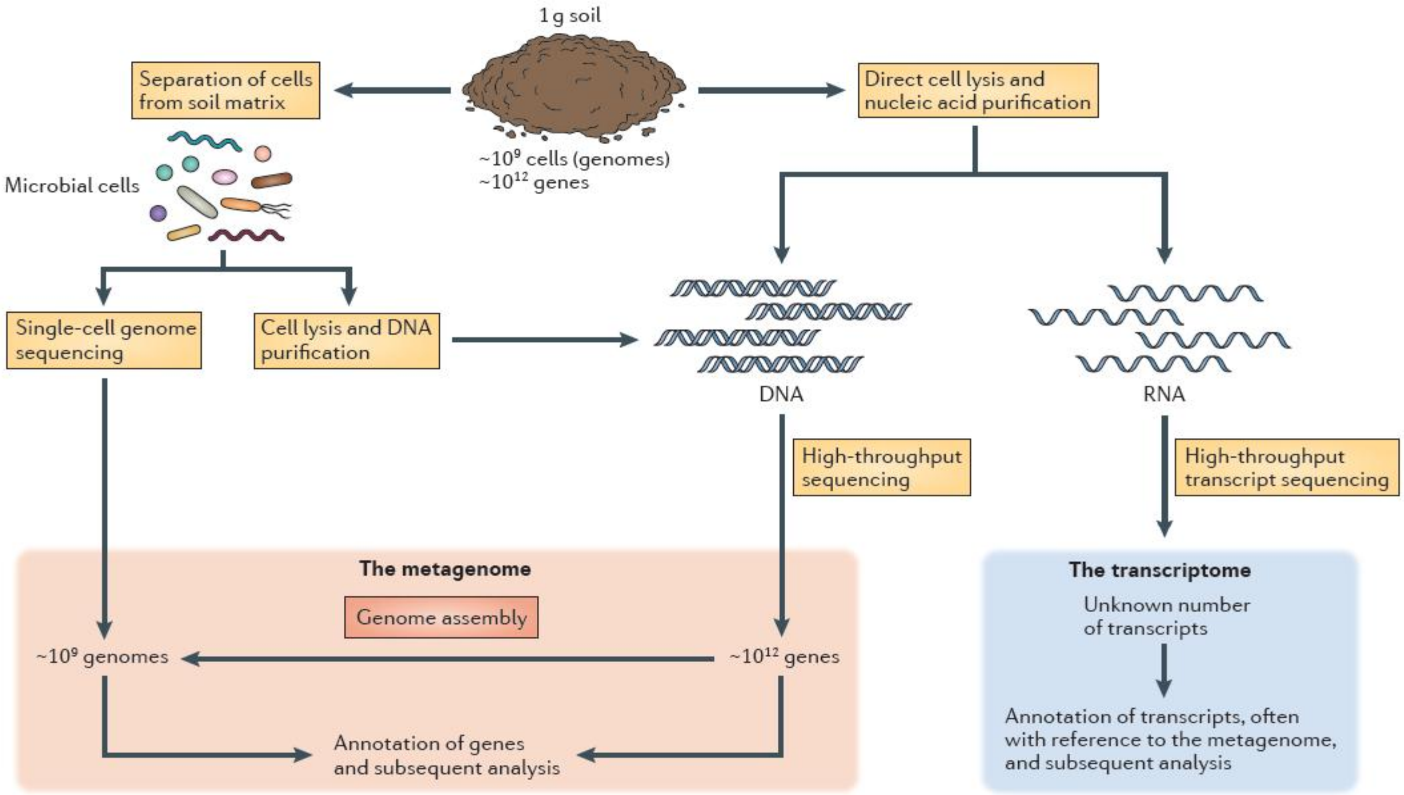


Figure 2.7: Schematic representation of main stages involved in single cell genomics, metagenomic and metatranscriptomic analyses of soil samples Single cell genomics retrieved the sequence information from individual cells while metagenomics provides a snapshot of the genetic composition of microbial communities in an environment and metatranscriptomics provides an insight into community level transcriptional profiling of an environment (Prosser, 2015).

A number of public databases such as NCBI non-redundant protein database (NCBI-nr) (Shi *et al.* 2011), SILVA database (Quast *et al.* 2013), Greengenes (McDonald *et al.* 2012), Ribosomal Database Project (RDP) (Cole *et al.* 2014), SEED subsystems and KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Frias-Lopez *et al.* 2008; McCarren *et al.* 2010; Shi *et al.* 2011) are available to BLAST the sequence data generated through high throughput sequencing and provide an insight into various phenomena and processes occurring inside the environment. Finally, indexes of α-diversity is calculated for each sample, β-diversity is computed across the samples, distribution of OTUs as a function of specific environmental parameters, analysis of co-variance between OTUs and other hypothesis testing relevant to the sample under study (Zhou *et al.* 2015). Another important factor to be considered while designing a metatranscriptomic experiment is accessibility of metagenomic sequences corresponding to the microbial communities to be investigated. Availability of a coupled metagenome provides an expression ratio i.e., cDNA:DNA calculated through the mapping of number of cDNA and DNA reads to each gene present in the metagenome. Expression ratios are considered more enlightening in comparison to simple relative frequencies as they provide information about expression of a gene and are standardized to the occurrence of genes within the metagenome whereas relative frequency consider abundance of a transcript only.

As metatranscriptomic approach involves direct sequencing of cDNA libraries synthesized from reverse transcribed RNA with no prior knowledge about the sequences present in a sample thus prove to be a useful approach to detect novel genes and phenomena active in the environment (Shi *et al.* 2009). In 2011, Gottesman and Storz detected novel small bacterial RNAs typically 50-500 bp in length that can bind to mRNA targets or proteins and have the ability to modify the respective gene expression levels or functions such as virulence, quorum sensing and stress response (Podkaminski and Vogel, 2010; Gottesman and Storz, 2011; Richards and Vanderpool, 2011).

Another approach of metatranscriptomics involves exploration of eukaryotic gene pool by focusing particularly on eukaryotic messenger ribonucleic acid (mRNA) enriched from the metatranscriptome isolated directly from environmental sample (Grant *et al.* 2006). Eukaryotic mRNA possesses 3’ poly-deoxyadenosine (poly-dA) tail which facilitates the reverse transcription of mRNA into cDNA through priming with the help of poly-dT oligonucleotide. Full length cDNAs synthesized from reverse transcription are further amplified by PCR and cloned downstream of a promoter region in a specific shuttle vector. This promoter helps in transcription of the cDNA insert and its successful expression in the form of protein inside the bacterial or eukaryotic host (Bailly *et al.* 2007). Extraction and purification of RNA especially from eukaryotes was also a daunting task but availability of various commercial kits and dedicated protocols published in the market can help us to circumvent this problem and yield a good quantity of cDNA (mg) from small amount of environmental RNA (ng) (Yadav *et al.* 2014, 2016). A good metatranscriptomic cDNA library constructed from highly purified mRNA encompasses more than 10^6 independent inserts (Bailly *et al.* 2007, Damon *et al.* 2011, Lehembre *et al.* 2013). This approach circumvents the problems related to dilution of eukaryotic DNA as in most of the metagenomes, presence of introns and taxon specific gene transcriptional signals especially specificity among promoters (Marmeisse *et al.* 2017).

Functional metatranscriptomics approach looks similar to the one described as functional metagenomics in the terms of work flow still differ in several points: (a) this approach involves eukaryotic genetic material hence unlikely to have screening of bacterial genes. (b) This approach only targets gene expressed in the environment at the time of sampling and do not involve non-expressed genes present in corresponding metagenome. (c) Transcription of gene in the final host involves an expression plasmid carrying compatible promoter hence, eradicate the chances of transcription and expression failure due to use of distantly associated promoter sequences.

2.4. Characterization of a metatranscriptome

Advent of massive parallel sequencing and RNA sequencing has opened up new and exciting opportunities with dynamic range in the field of transcriptome analysis. Metatranscriptomics enables to study the gene expression of complex environments and demonstrate the characteristics and functions of microbial communities actually active at a given context of time. Such functional data further helps in identifying the active metabolic pathways associated with particular environmental conditions. Metatranscriptomic analyses involves: (i) isolation and processing of microbiome mRNA into cDNA libraries, (ii) computational analyses of metatranscriptomic data

2.4.1. Isolation and processing of mRNA

Typically, metatranscriptomic study involves the isolation of total RNA from the environmental sample. While working on eukaryotes, messenger RNA can be enriched by using cDNA kit provided with oligo-d(T) primers which can bind to the poly-A tail of the mRNA. In case of prokaryotes, where mRNA is without poly-A tail and makes only 1-5% of total RNA, hence makes its selection during cDNA synthesis unsuitable. Removal of rRNA involves use of the probes that can anneal to the target sequences i.e. rRNA followed by their removal with the help of magnetic beads (Sultan *et al.* 2014). As metatranscriptomics involves RNA, the next challenge is avoiding RNA degradation due to ribonulcease contamination. Maintenance of good laboratory practices and use of RNase inhibitors in the procedure can be an effective strategy. Massive parallel sequence analysis involves isolation of RNA and their fractionation to synthesize cDNAs by using adaptors which are ligated to cDNA ends. These cDNAs with adaptor sequences are used to synthesize cDNA library which will be further amplified and screened for functionally active genes (Giannoukos *et al.* 2012).

2.4.2. Computational analysis of metatranscriptomic data

As metatranscriptomic experiments generate huge data and analysis of this data is an important and challenging task. With consistent increase in data size and number, an automated and efficient high throughput analysis of this data is essentiality of the field of molecular biology to infer the biological meaning from these datasets (Gosalbes *et al.* 2011; Korf, 2013). Various analytical suites such as HUManN, MG-RAST, have been developed to get end to end solutions and applied alongside the combinations of specialised bioinformatics tools such as BOWTIE and GEM (Abubucker *et al.* 2012; Glass *et al.* 2010; Langmead and Salzberg, 2012; Marco-Sola *et al.* 2012). Before this analysis, few steps common to all metatranscriptomic data are applied and consist of filtering of host reads and non mRNA reads and filtering or trimming of low quality reads and nucleotides (Wang *et al.* 2009). After initial quality check and filtering, reads are assembled into contigs (which is an optional step) followed by mapping these contigs against the available reference genome. The assembly of reads is quite challenging step computationally and requires high quality sequencing data. This step holds high potential to discover the hidden knowledge about expression of genes in a sample. To enable the assembly step, deeper sequencing is required hence only highly abundant regions can be assembled from a larger set of reads (Morgan and Huttenhower, 2014). Assembly step is much needed where reference genome and subsequent gene annotations are not available such as in RNA-seq of non-model organisms. In case of unavailability of reference genomes for a particular data sets, the annotations for the sequenced transcripts are obtained by finding sequence similarity to the sequenced and annotated proteins i.e., by aligning them against large annotated protein databases such as SILVA, Blast2GO (Conesa *et al.* 2005) and if highly similar proteins are found, a similar biological function is inferred.

Another important issue in the metatranscriptomic analysis is combining the RNA-seq data and whole genome data analysis. Analyzing the metagenomic and RNA-seq data simultaneously for a given sample enables to infer the actual expressed genes out of potentially existing genes (Bikel *et al.* 2015). Further, a summary of data is converted into relative gene expression values and can then be further analyzed accordingly such as gene expression level within a sample, richness within samples, and similarity between samples.

2.5. Metatranscriptomic approach to explore eukaryotic lineages

Eukaryotic microflora from various environments covers a large number of taxa out of which many are still unknown. It was observed that eukaryotic microflora is highly diverse in terms of their taxonomic as well as functional diversity and secondly a large portion of eukaryotic microflora is still unexplored and or uncultivable hence represents the so called dark matter (Solden *et al.* 2016). Hence, by direct analysis of environmental RNA through metatranscriptomic approach gives a novel technique to have a rapid access to the hundreds of taxonomically as well as functionally diverse eukaryotes in a habitat including the uncultured and unknown ones at the time of sampling (Marmeisse *et al.* 2017). It has been observed that eukaryotic microbial communities from various environments such as soils, aquatic environments (fresh and marine) and sediments embrace the representative species of almost all the major eukaryotic lineages though the comparative analysis of these environments concluded that each of them is characterized by the presence of specific eukaryotic phyla (Grossmann *et al.* 2016; Groussman *et al.* 2020). For instance, soils are characterized by the abundance of filamentous fungi and amoebozoa in comparison to the environments like lakes which are characterized mainly by the dominance of ciliates and chytrids (Tedersoo *et al.* 2014). Worldwide analysis of eukaryotic microbial communities by meta-barcoding also drawn attention to incomplete understanding of eukaryotic diversity and underpin the point that eukaryotic diversity is still not well described. For, instance, global survey of protist diversity in oceans observed that ten times more OTUs in comparison to the number of already described marine planktonic species (deVargas *et al.* 2015). Such global studies have also confirmed that majority of the eukaryotic microbes are not omnipresent and variations in distribution of species illustrated a strong bio-geographical outline at large geographical scales (deVargas *et al.* 2015, Talbot *et al.* 2014). For instance, saprophytic fungi present in soils of tropical biomes showed more species richness and a remarkable taxonomic difference from those present in soils of temperate biomes (Tedersoo *et al.* 2014). Environmental factors such as availability of water, nutrients and soils can also affect the taxonomic distribution among the eukaryotic microbial communities such as protists (Bates *et al.* 2013). These data and factors helps in understanding the environmental diversity and are essential to wisely collect the environmental sample covering different species and microbial communities of an environment. Second reason to focus on eukaryotic communities is availability of lesser number of genome sequences which is a crucial source for the precise taxonomic as well as functional annotations of environment derived eukaryotic sequences and in executing novel gene mining practices on various environmental samples. Although, it is tough to provide the exact number of accessible eukaryotic genome sequences, various works have signified that in year 2014, genome sequences corresponding to one species as a minimum was accessible for about half of hundred identified eukaryotic lineages (Keeling *et al.* 2014; del Campo *et al.* 2014). These observations highlighted huge differences among the various lineages as a few of them are represented by numerous genome sequences while a large number of them are still represented by only a single genome sequence. For example, there are thousands of known genome sequences for fungi due to ongoing wide range of genome sequencing projects hence a provides a good resource for novel fungal genome sequences retrieved from various environmental samples (Grigoriev *et al.* 2014). Such huge differences among the availability of genome sequences for various eukaryotic lineages laid stress on exploring eukaryotic microbial communities especially the unknown and under explored ones. Henceforth, these data and observations from various studies have illustrated that analysis of metatranscriptomes represents an efficient practice to access taxonomic as well as functional diversity of eukaryotic communities flourishing in different environments.

Thirdly, the members of eukaryotic communities depicts significant role in natural phenomena and industrial processes. It has been observed that various biomolecules of eukaryotic origin plays a dominating role in various ecological processes such as organic matter degradation (Damon *et al.* 2011), biotransformation of xenobiotic compounds (Singh *et al.* 2018) as well as large scale industrial processes such as biomass treatment, fermentation industry and textile and paper industry (Haon *et al.* 2015; Spohner *et al.* 2015). Eukaryotic microbiota represents a hidden and under explored source of numerous biocatalysts and metabolites. Eukaryotic microorganisms mainly fungi are also among the dedicated major producers of primary metabolites such as organic acids and secondary metabolites such as immunosuppressors and antibiotics (Hofrichter and Esser, 2010). It has been observed that these metabolites have been procured from a few of the fungal taxa and does not represents entire eukaryotic realm which itself urge to explore the under mined domains or taxa of eukaryotic origin with remarkable significance in modern era industry. Another reason to be limited to only few fungal species with known industrial applications as many of them are recognized as "Generally Recognized as Safe" (GRAS) microbial species such as numerous fermented products and metabolites have a common producer i.e., *Aspergillus* (Marmeisse *et al.* 2017). Besides from their significance in the metabolite production, eukaryotes are also known to be source of several biotechnological significant products which are either lacking or not expressed in bacteria and archaea. For example enzymes like Class II heme peroxidases especially manganese and lignin peroxidases for hydrolysis of lignin (Lundell *et al.* 2014) and several other unspecific peroxygenases which catalyses the oxyfunctionalisation of various organic molecules hence functions in biomass treatment or organic matter degradation (Hofrichter and Ulrich, 2014). Similarly, enzymes such as lytic polysaccharide monooxygenases are found to be catalyzing the oxidative breakdown of polysaccharides such as hemicelluloses and cellulose (Quinlan *et al.* 2012). Metatranscriptomic approach introduced a new family of glycoside hydrolases consists of β- glucanases act on number of polysaccharides and are responsible for biomass treatment and (Agger *et al.* 2014; Lafond *et al.* 2012). Henceforth, with the discovery of several novel classes of enzymes of eukaryotic origin in recent studies depicts that there is lot to be explored among eukaryotic taxonomic and functional diversity and is still incomplete.

Table 2.1: Metatranscriptomic approach applied on different environments

Environment samples	**Sequencing platform**	**Size of data sets**	**Significant findings**	**Reference**
Forest Soil	Sanger sequencing	119 clone sequences	Eukaryotic diversity, novel enzymes	Bailly *et al.* (2007)
Nutrient-poor sandy soil	Pyrosequencing on Roche GS20 sequencer	258,411 high-quality RNA-tags.	CO_2 fixation and ammonia oxidizers enzymes	Urich *et al.* (2008)
Sugar maple forest soil	Screening by auxotrophy complementing genes in yeast	10 transformed yeast sequenced	Acid phosphatase, Imidazole glycerolphosphate dehydratase	Kellner *et al.* (2011)
Spruce forest	Screening by auxotrophy complementing genes in yeast	7.7×10^5 screened transformants and 25 putative genes	Novel fungal oligopeptide transporter	Damon *et al.* (2011)
Beech and Spruce forest	Sanger sequencing	10,000 each	Eukaryotic Diversity, novel enzymes for organic matter degradation	Damon *et al.* (2012)
Activated sludge	Illumina platform	2.4 Gbp	Enzymes involved in nitrogen metabolism	Yu and Zhang (2012)
Soil near metal smelter	Sanger sequencing	30,000 cDNA for each soil sample	Metal tolerant genes	Lehembre *et al.* (2013), Ziller *et al.* 2017

Paddy soil	Pyrosequencing on 454 GS FLX instrument (454 Life Sciences)	35,851 and 27,304 SSU rRNA tags	Structural and functional succession in oxic layer and anoxic soil	Kim and Liesack, 2015
Artificial pond water sample	Illumina platform.	14,792,542 to 16,756,988 reads	Differentially expressed transcripts during diurnal cycle	Trench-Fiol and Fink, 2020

2.6. Methodological challenges in metatranscriptomics

2.6.1. Efficient mRNA extraction protocol and mRNA instability

Key step in the metatranscriptomic approach is extraction of good quality of mRNA. Average half-life of mRNA molecules varies among different microbial species with a range between seconds to minutes and can be influenced by nutritional status of individual cells (Selinger *et al.* 2003; Hambraeus *et al.* 2003; Redon *et al.* 2005; Deutscher, 2006). It has been observed that genes with similar biological functions have similar rates of messenger RNA degradation while housekeeping genes possess more stable messenger RNAs (Bernstein *et al.* 2002; Hambraeus *et al.* 2003). For RNA studies, sampling should be fast and samples collected should be frozen instantly at -70°C or transfer into a RNA preservation solution to minimize the changes that occur in transcriptional profiles due to sampling (Yadav *et al.* 2014). RNA isolation from environmental samples such as soils is very daunting task due to many reasons: ineffective cell lysis, presence of RNases and adsorption of RNA to the soil particles. Various methods are employed to lyse cells such as bead beating, liquid nitrogen grinding (Volossiouk *et al.* 1995), microwave based rupture (Orsini and Romano-Spica, 2001) and enzymatic lysis (Zhou *et al.* 1996), out of which bead beating has been observed to be the most efficient method and is an initial step in various extraction protocols (Lakay *et al.* 2007). Further, it was observed that the conditions which are optimized for RNA extraction such as low pH for extraction buffers and high salt concentration to inactivate RNases also contribute towards hindrances in RNA extraction (Chomczynski and Sacchi, 1987). Complex organic molecules present in the soil samples such as humic and fulvic acid obstruct the nucleic acid precipitation and further limit the template availability hence inhibit the PCR amplification (Arbeli and Fuentes, 2007; Opel *et al.* 2010). Various methods such as use of powdered activated charcoal, precipitation with ammonium sulphate, treatment of soils with $CaCO_3$, addition of polyvinyl polypyrollidone (PVPP), purification of nucleic acid with $CaCl_2$, RNA extraction at low pH i.e., pH 5 followed by purification through cetyl trimethylammonium bromide (CTAB) and vitamin supplemented Q-Sepharose columns. Presence of genomic DNA in RNA sample can overestimate its concentration during quantification and can be considered as transcripts post sequencing, thus this problem can be resolved by treating the extracted nucleic acid with DNase I (Carvalhais *et al.* 2012). To overcome these challenges faced during RNA extraction various highly efficient commercial kits are also available in the market which are less time consuming and easy to use.

2.6.2. Enrichment of mRNA

Isolation and enrichment of messenger RNA molecules is a significant step in the metatranscriptomic studies as total RNA pool of an sample consist of all the three types of RNA: rRNA, tRNA and mRNA (Karpinets *et al.* 2006; He *et al.* 2010). Several methods have been implemented to enrich mRNA from various samples: (a) exonuclease treatment, (b) subtractive hybridization, (c) treatment of RNA with duplex specific nuclease (DSN) (Yi *et al.* 2011) and (d) size separation mediated by gel electrophoresis (McGrath *et al.* 2008). Comparison of these approaches shows that their efficiencies differ depending upon RNA integrity and the environment from which sample was collected. A recent approach for mRNA enrichment involves the use of a specific probe mix such as use of antisense ribosomal RNA probes engendered by *in vitro* transcription of amplified products using coupled DNA as template (Stewart *et al.* 2010).

Presence of prokaryotic or eukaryotic RNA as an impurity in a requisite RNA sample is also another issue in metatranscriptomic studies. Prokaryotic RNA can be enriched by using poly-dT probes which can bind to poly-dA site of eukaryotic mRNA. Similarly, eukaryotic mRNA can be enriched by constructing cDNAs using adaptors carrying oligo dT primers which will bind to 3' poly-dA tail of eukaryotic mRNA only (Yadav *et al.* 2014) while non-eukaryotic RNA with polyA tail are rare and if present, degrades rapidly (Belasco, 2010; Dreyfus and Regnier, 2002).

2.6.3. cDNA synthesis and amplification

Samples such as soils yield a less quantity of mRNA and needs an additional step of amplification. In case of eukaryotic RNA, this process requires oligo-dT primers for *in vitro* transcription and large quantity of anti-sense strands of RNA are synthesized. These strands are further used as template to synthesize double stranded cDNA by reverse transcription using random primers. Another method i.e., multiple displacement amplification (MDA) which was originally applied to amplify a little amount of DNA, could be used on complementary DNA as well (Blanco *et al.* 1989; Gilbert *et al.* 2008). Disadvantages of applying MDA to cDNAs is it compromises quantitative analysis of metagenomic data due to amplification biases thus limits the data interpretation to the presence or absence of transcripts (Yilmaz *et al.* 2010).

High-throughput sequencing techniques require complementary DNA as a template in which errors can be introduced during reverse transcription by reverse transcriptase and can undergo shearing, size selection and ligation of adaptors. It has been observed that rate of reverse transcription depends upon length of transcripts such as shorter the transcript, more efficiently it will be reverse transcribed (Stewart *et al.* 2010). In addition, template switching in high homology regions of cDNA by reverse transcriptase can be helpful in generating chimeric cDNA molecules (Cocquet *et al.* 2006).

2.7. Ecologically significant genes procured via metatranscriptomic approach

Metatranscriptomic analysis of various environments based upon expressed genes is the more appropriate approach to unravel the eukaryotic community functions in the environment. Numerous environments have been examined for mining novel genes and microorganisms playing significant role in industries, human health and in maintaining ecosystem through metatranscriptomic approach.

2.7.1. Organic matter decomposing enzymes

Degradation of recalcitrant organic matter from plant litter, human litter or animal waste is a central component of the carbon cycling and responsible for good soil health. Earlier laboratory processes have investigated mainly the cultivable microfora of the soil for organic matter decomposition but now with the advancement in microbial ecology, various studies have focused on the functional diversity of these organic matter rich soils such as forest soils and peat lands etc through metatranscriptomic approach (Marmeisse *et al.* 2017; Kuramae *et al.* 2013; Ivanova *et al.* 2016). Metatranscriptomics has identified numerous functional genes for organic matter degradation from various sample sites such as forests soils, agricultural soils etc and various classified metatranscripts had revealed a diverse array of oxidative as well as hydrolytic enzymes involved in degradation of organic matter such as cellulose, hemicelluloses, lignin etc (Kellner *et al.* 2011; He *et al.* 2013; Takasaki *et al.* 2013; Hori *et al.* 2018). Takasaki et al. (2013) identified numerous cDNAs carrying protein coding sequences for glycoside hydrolase enzymes such as glucanase, α-glucosidase, β-glucosidase and chitinase by using metatranscriptomic approach. Similar outcomes were reported when forests soils were explored through this technique for the presence of some oxidative and hydrolytic enzymes for lignocellulose degradation such as cellulases, hemicellulases, glycoside hydrolases and redox enzymes such as manganese peroxidases (Hori *et al.* 2018). In 2012, Damon et al. reported putative homologs to different enzymes from forest soils which participated in utilization of soil macro and micronutrients and majority of these protein coding sequences were identified as plant cell wall degrading enzymes active in degradation of cellulose, sugars, phosphates, sulfates and other compounds (Damon *et al.* 2012). Mining of other organic matter degrading enzymes such as phytases, proteases, lipases and P450 monooxygenases, cellobiohydrolase from various environments suggested metatranscriptomics as a potential significant approach in mining ecological significant genes from environmental sample (Damon *et al.* 2012; Ivanova *et al.* 2016; Kuramae *et al.* 2013).

2.7.2. Aromatic and xenobiotic degrading genes

Soils are dump for various xenobiotic compounds or synthetic chemicals such as phenyls, polyaromatic compounds, petroleum hydrocarbons, polychlorinated biphenyls, herbicides, insecticides and fungicides and were primary cause of biospheric pollution (Rabinovich *et al.* 2004; Sinha *et al.* 2009). Singh et al. (2018) highlighted a potential significant role of rhizospheric microbiota in degradation of aromatic as well as xenobiotic compounds by reporting numerous transcripts and pathways involved in degradation of these compounds (Singh *et al.* 2018). Microbial communities flourishing in freshwater sediments contaminated with aromatic and xenobiotic compounds were reported to have high gene expression of microbial transcripts related to unique catabolic and anabolic pathways (nitrate reduction, polyester synthesis, methanogenesis, hydrocarbon degradation) to exploit these contaminants as a source of energy (acetyl-CoA, succinyl-CoA) and transforming them into stable forms, concurrently recruiting stress-response and gene transfer mechanisms to adapt such stress (Reid *et al.* 2018; Falk *et al.* 2019). Similarly, de Menezes et al. (2012) reported the higher expression of transcripts related to dioxygenase genes, stress response genes and phenanthrene degradation genes on the exposure of soil microbial communities to phenanthrene. Various soil fungi especially mycorrhizal fungi were also reported to have a significant role in degrading natural as well as polyaromatic hydrocarbons, phenanthrene and pyrenes (Fernández-Luqueño *et al.* 2010; Yu *et al.* 2011; He *et al.* 2020). In a research involving metatranscriptomic analysis of mycorrhizal association of *Salix purpurea* cv in both contaminated as well as non-contaminated soils in Canada, suggested the active involvement of various fungal genera such as white rot fungi ascomycota, ectomycorrhizal (ECM) basidiomycota and ECM ascomycota and reported the higher expression of CAZy enzymes and ectomycorrhizal enzymes such as exo-β-1, 3-glucanase to combat the stressful conditions (Gonzalez *et al.* 2018).

2.7.3. Heavy metal tolerant genes

Contamination of soil, water and air with heavy metals due to urbanization and industrialization is a widespread concern. Heavy metal contamination resulted into deterioration of vast land areas which makes them unsuitable for cultivation and other activities. Various metatranscriptomics studies have been reported to explore the microbiota of such sites to find some novel mechanisms and genes with a potential to tolerate such huge concentrations of these metals (Marmeisse *et al.* 2017). Several cysteine rich proteins (CRPs) encoding genes that conferred Cd tolerance to Cd sensitive yeast mutants have been isolated from soil metatranscriptomes. These proteins of unknown taxonomic origin shared conserved cysteine motifs could be considered as environmental metallothioneins (EMTs) (Ziller *et al.* 2017). In a study involving screening of soil metatranscriptomic library reported that majority of metal tolerant cDNAs belong to unknown taxonomic groups which further advocate the utilization of metatranscriptomic approach for mining new ecologically significant genes (Lehembre *et al.* 2013). Metatranscriptomics revealed the active participation of cDNAs encoding putative aldehydrate dehydrogenase and serine protease inhibitor proteins in tolerating a range of metal ion concentrations (Cd^{2+}, Cu^{2+}, Zn^{2+} and Co^{2+}) isolated from polluted soil environment (Mukherjee *et al.* 2019a, 2019b).

2.8. Metatranscriptomics of the soil

Increase in the pollutants in soil due to anthropogenic activities increases stress level among the microbes which further results into “competitive exclusion” with removal of dominant species and increase in number of less abundant species in the soil environment. Competitive exclusion is further followed by “microbial resilience” as the affected soil microbiota try to regain their structure and function after an initial disturbance (Allison and Martiny, 2008). Studies pertaining to potentially toxic metals such as cadmium and copper have indicated the presence of ‘competitive exclusion’ in soil environment (Degens *et al.* 2001; Zhang *et al.* 2009). Eukaryotic life forms hold a repository of genes responsible for various activities in soil, while past studies generally investigated soil fungi for centuries (Rossman, 1998), but development of DNA-based analysis have paved their way to prove that phylogenetically, eukaryotes are as diverse as bacteria. Environmental RNA-based metatranscriptomics has been developed for different soil environments and is assumed to better reflect ongoing biological activities in these soils (Urich *et al.* 2008; Shrestha *et al.* 2009; Damon *et al.* 2012). 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). However, eukaryotic metatranscriptomics is still in its infancy. Researchers who engage in this field have therefore to face methodological and conceptual issues which were tried to resolve by the methodology used in this thesis. Eukaryotic 3’ polyadenylated messenger RNAs can be isolated from ribosomal, non-coding RNAs and bacterial mRNAs that dominate environmental metatranscriptomes to study the functional potential of the eukaryotes present in the soil (Yadav *et al.* 2014; Ziller *et al.* 2017). Furthermore, molecular diversity surveys based on the nuclear 18S rRNA gene showed that all major eukaryotic clades are represented in soil samples (Richards and Bass, 2005; Bailly *et al.* 2007; Lesaulnier *et al.* 2008; Damon *et al.* 2012). Proofs of concept for this protocol have now been established for various ecosystems such as soils (Grant *et al.*

2006; Bailly *et al.* 2007; Kellner *et al.* 2011; Damon *et al.* 2011; Lehembre *et al.* 2012) or bovine (Qi *et al.* 2011) and termite digesters (Todaka *et al.* 2007; Tartar *et al.* 2009).

Such advanced molecular biology approaches quantify potential molecular indices that may aid in assessment of environmental risks due to potentially toxic metals. Use of novel biomarkers procured from environment is an alternative approach to standard toxicity tests that can provide more information about the organism's response to individual potential toxicants present in soil environment (Kammenga *et al.* 2000, Scott-Fordsmand and Weeks, 2000). Such an approach, although not without practical challenges offers a number of advantages including (i) the potential for detection of low levels of exposure, (iii) the potential to link molecular or cellular responses to ecologically relevant life-cycle traits (such as reproduction and growth), (iv) the ability to tease out toxicant-specific effects and the evaluation of subtle non-lethal effects on laboratory and field populations of contaminants. Hence, the expanding functional genomic database provides a huge resource from which candidate molecular biomarkers can be identified and subsequently exploited to address the issues related to metal toxicity in soil environment.

2.9. Summary of literature review

As discussed in this literature review, urbanization and industrialization has led to the contamination of environment with potentially toxic metals (PTMs) and caused various health hazards. To cope up with such toxic concentrations of PTMs, living beings especially eukaryotes developed various mechanisms and phenomena. Cellular and molecular features of such metal toxicity, tolerance and homeostasis have been explored in various naturally resistant or tolerant eukaryotic organisms as well as metal sensitive model eukaryotic organisms. Various experimental strategies followed to uncover these mechanisms and their potential members include culture based analysis and sequence based analysis such as genomics and metagenomics. In traditional methods, research is limited by the fact that only cultivable members of an environment can be explored while genomic sequencing involved the screening of various genomes for sequences which showed homology to already known metal tolerant or resistance genes followed by experimental validation. The major drawbacks of both these approaches involved missing out of uncultivable members in traditional methods while in genomic approach it is unlikely to uncover novel genes and tolerance mechanisms as it is revolving around already identified genes only. With the advancement in sequencing technology, metagenomics, a recent omic approach with a potential to uncover the number of novel species and genes flourishing in an environment was established. Metagenomic involves total DNA content of an environment and provided knowledge about total diversity of an environment but unlikely to predict whether these genes or species were active at the time of sampling or not. Thus, to explore the functional diversity of an environment and its novel metal tolerant affiliates, another approach i.e., metatranscriptomics involving study of total mRNA of an environment have been established. Metatranscriptomic involves the use of high throughput sequencing techniques for community level transcriptional profiling of reverse transcribed RNA and functional screening of mRNA derived cDNA libraries in eukaryotic hosts such as *Saccharomyces cerevisiae* for novel species and genes. Unlike metagenomics which provides a snapshot of the genetic composition of microbial communities in an environment at any given time interval, metatranscriptomics helps in understanding the expression potential of numerous known and unknown genes present in an environment at the time of sampling. Various previous works cited in this literature review supported metatranscriptomics as a promising approach to explore diverse environments for their hidden treasures in the form of diverse microbial communities. Thus, all these recent findings and factors supported this approach based upon expressed genes as a suitable and powerful tool for exploring soil environment in order to characterize novel genes whose over-expression could restore tolerance or conferred resistance to the microbiota of soil or could be exploited for assessment of various toxicants present in the environments.

MATERIALS AND METHODS

Chapter 3 MATERIALS AND METHODS

3.1. Biological materials and culture conditions

For all the bacterial transformations, MegaX DH10B™ T1R Electrocomp™ cells, Invitrogen, USA, One Shot® TOP10 Electrocomp™ *E. coli* cells (Invitrogen, USA) were used. All the bacterial strains were maintained on Luria Broth (LB) media (Appendix I) at 37ºC and bacterial transformations were done on LB media amended with ampicillin. For yeast transformations, wild type strain of *S. cerevisiae* BY4741 and auxotrophic mutant strains derived from wild type BY4741 were used (Table 3.1).

Table 3.1: Bacterial and yeast strains used in the study

Strains used in the study		
Bacterial strain		
One Shot® TOP10 Electrocomp™ *E. coli* cells, MegaX DH10B™ T1R Electrocomp™ cells, Invitrogen, USA,		
Yeast Strain	**Phenotype**	**Genotype**
BY4741	Wild type	*MATα his3D1 leu2D0 met15D0 ura3D0*
ycf1$^{\Delta}$	Cd sensitive	*MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; YDR135c::kanMX4*
zrc1$^{\Delta}$	Zn sensitive	*MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; YMR243c::kanMX4*
cot1$^{\Delta}$	Co sensitive	*MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; YOR316c::kanMX4*
cup1^{s} (DTY4)	Cu sensitive	*MATα trp1-1 leu2-3 112his3Δ1 gal1 ura3-50, cup1Δ::URA3+*
cup1^{s} (DTY3)	One copy of isogenic copperthionein	*MATα leu2-3 112his3Δ1 trp1-1 ura3-50 gal1 CUP1s*

Cd-sensitive *ycf1*$^{\Delta}$, Cu-sensitive *cup1*$^{\Delta}$ (DTY4), Zn-sensitive *zrc1*$^{\Delta}$ and Co-sensitive *cot1*$^{\Delta}$ mutant strains of *S. cerevisiae* were used for heterologous complementation of cDNAs. The yeast mutant *ycf1*$^{\Delta}$ lacks *YCF1* gene which is an ABC transporter encoding gene responsible for conferring Cd tolerance through the vacuolar transport of Cd conjugates (Li *et al.* 1997). Similarly, copper sensitive *cup1*$^{\Delta}$ lacks both the copies of copperthionein gene while its isogenic wild type strain DTY3 (*cup1*^{s}) carries single copy of copperthionein gene *CUP1* (Longo *et al.* 1996). In case of *zrc1*$^{\Delta}$, *ZRC1* gene encoding a transporter protein that helps in sequestering of Zn into vacuoles is knocked out (Li and Kaplan, 1998). The hypersensitivity to Co in *cot1*$^{\Delta}$ strain is due to deletion of *COT1* gene that mediates the efflux of cobalt ions into the vacuole (Conklin *et al.* 1992). For heterologous complementation assays, *ycf1*$^{\Delta}$, *cot1*$^{\Delta}$ and *zrc1*$^{\Delta}$ strains transformed with empty vector (EV) were used as negative control whereas strain BY4741 transformed with EV were used as positive control. For maintaining yeast strains, YEPD media (Hi-media Laboratories, India) (Appendix I) was used while yeast transformations was done on synthetic defined without uracil (SD –Ura) media (Appendix I) at 30ºC.

3.2. Soil sampling

Soil samples were collected from agro-forestry land in Pierrelaye (PL) (49°1′45”N, 2°10′32″E) in northwest of Paris, France. This site was previously maize growing crop field which had been converted into an agro-forestry site where poplars are being cultivated at the time of sampling. These sites became polluted with potentially toxic metals (PTMs) due to prolonged irrigation with untreated wastewater for many years. Twenty soil cores were collected from three different sites i.e., PL1, PL2, PL3 and equal volumes of each core was mixed together to form a composite sample after sieving them through 2.0 mm mesh. Subsamples were taken from the composite samples and transported to the lab in frozen state at -80ºC for further experiments.

3.3 Physicochemical properties of metal polluted soil

Soil sample was analyzed for its physicochemical properties such as pH (1:2 soil:water ratio, v/v), total nitrogen, organic carbon, total phosphorus available phosphorous and metal content of the sample.

3.3.1 Total organic carbon (Walkley, 1947)

Organic carbon was determined by the following method:

Reagents (Appendix II)

Potassium Dichromate ($K_2Cr_2O_7$), Normal Solution, $FeSO_4$ solution, 0.5 N H_2SO_4, Concentrated H_2SO_4, Orthophosphoric acid (H_3PO_4) (85%) and Indicator.

Standardization of ferrous sulphate solution

Standardization of ferrous sulphate solution was done by adding 10 ml of $K_2Cr_2O_7$ solution to 20 ml of concentrated H_2SO_4 followed by addition of distilled water (200 ml), H_3PO_4 (10 ml) and mixed indicator (1 ml) to the mixture. The mixture was shaken vigorously and titrated against ferrous sulphate solution (present in burette) until the solution's color changed from blue to green. Further, $K_2Cr_2O_7$ solution (0.5 ml) was added to change the color back to blue. Drop wise addition of $FeSO_4$ solution with continued swirling again changed the color of the solution from blue to green and the total volume of $FeSO_4$ solution used (X) to the get the end point was recorded.

For soil analysis

The weighed oven dried soil sample (W1) was taken and sieved through a 10 mm sieve. The sieved sample weight was recorded as W_2 and powdered so that it passes through the 425-micron sieve. After sieving, 5 g of sieved soil was measured and placed in a flask to which $K_2Cr_2O_7$ solution (10 ml) and concentrated H_2SO_4 (20 ml) was added. The mixture was mixed by swirling and allowed to stand for 30 min for oxidation of the organic matter. Further added 200 ml of distilled water to the mixture along with orthophosphoric acid (H_3PO_4) (10 ml) and mixed indicator (1 ml) followed by vigorous shaking. The mixture was again titrated against $FeSO_4$ solution (with the increment of 0.5 ml) with constant mixing, until the color of the solution changed from blue to green. Further, addition of 0.5 ml of $K_2Cr_2O_7$ solution (0.5 ml) to titrate mixture again changed the color from green to blue again. Ferrous sulphate solution was again added to titrate mixture (drop-wise) and the change of color from blue to green was recorded immediately after the addition of a single drop.

Calculations

The total volume (V ml) of $K_2Cr_2O_7$ required to oxidize organic matter in the sample is given by the following formula:

$$V= 10.5 (1 - r/X)$$

where

r = Total volume of $FeSO_4$ used

X= Total volume of $FeSO_4$ used in standardized test

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

$$\text{Organic matter (\% by weight)} = \frac{0.67\ W2V}{W1W3}$$

where

W_2 = weight of soil sample passing 10 mm Seive

V = total volume of $K_2Cr_2O_7$ used to oxidize the organic matter

W_1 = weight of soil sample taken before sieving

W_3 = weight of soil sample used for the experiment

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

The basic principle involved in total nitrogen estimation states that digestion of soil sample with concentrated H_2SO_4 in the presence of a catalyst converts the organic nitrogen into ammonium sulphate which further releases ammonia by distillation with concentrated alkali solution. Ammonia formed is then absorbed by H_2SO_4 and the excess acid is titrated with standard alkaline solution.

Reagents used (Appendix II)

Potassium Sulphate or Anhydrous Sodium Sulphate, Mixed catalyst (Copper Sulphate), Concentrated H_2SO_4, Sodium hydroxide solution, 0.5 N standard H_2SO_4 solution, 0.25 N standard NaOH solution Alkaline Sodium Sulphide Solution. And Methyl Red Indicator solution

Procedure

Finely crushed soil sample was added to the Kjeldahl flask along with the catalyst (10 g) and concentrated H_2SO_4 (30 ml) followed by boil the contents till the solution becomes clear. Then the contents were further boiled gently for 2 hours. After cooling, the content was transferred to the separating funnel containing 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 was left. Titration of the excess H_2SO_4, added in the beaker along with methyl red indicator, was done against standard sodium hydroxide solution. Blank for the soil sample was conducted in presence of all chemicals except soil sample.

Calculations

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) and metal estimation

Total phosphorus content and metal content was determined spectrophotometrically with acid digested soil samples.

3.3.3.1. Acid digestion of soil samples

Acid digestion of soil sample was done by nitric acid and perchloric acid.

Reagents used (Appendix II)

Concentrated nitric acid (HNO_3), Perchloric acid ($HClO_4$) (60%), 0.5N HCl and 5N H_2SO_4.

Procedure

The soil samples were digested with a combination of nitric acid (HNO_3) and perchloric acid ($HClO_4$) (Hseu, 2004). For acid digestion, soil sample (1 gm) was placed on a digestion tube and 10 ml of concentrated HNO_3 was added followed by digestion of the sample contents for 1 hr at 145°C in acid proof digestion having fume exhaust. The reaction mixture was further allowed to cool and 10 ml of concentrated HNO_3 and 5 ml of perchloric acid ($HClO_4$) was added to the mixture. The mixture was then heated at about 100°C for 1 hr and then raised the temperature to 200°C. The digestion procedure was continued until the contents became colorless and white matter remained in the digestion tube. After removing from the heating plate the solution was cooled and 50 ml of distilled water was added to the tube followed by filtration of the contents through a Whatmann filter paper number 42.

3.3.3.2. Metal estimation

The above digested samples from above step were then used to estimate the metal contents by performing inductively coupled plasma atomic emission spectroscopy (ICP-AES) (ARCOS, Simultaneous ICP Spectrometer, SPECTRO Analytical Instruments GmbH, Germany) (Moor *et al.* 2001).

3.3.3.3. Total phosphorous

The reagent and procedure used to estimate total phosphorus was given below:

Reagents (Appendix II)

Solution A, Solution B (freshly prepared)

Procedure

Total phosphorus present in a soil sample was determined by colorimetry for acid digested samples. The digested samples were transferred to volumetric flasks (50 ml) and 35 ml of distilled water was added to it. To the sample, solution B (10 ml) was added and volume was made up to 50 ml with distilled water. The color was allowed to develop and the intensities were read at 690 or 880 nm. The standard curve was prepared with the standard (KH_2PO_4) concentration of 0-1 µg/ml.

Calculations

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

3.3.4. Available Phosphorous (Olsen, 1954)

Available phosphorus in soil was determined by Dickman and Bray’s Method

Reagents (Appendix II)

Ammonium Fluoride solution, 0.5 N HCl, Extracting Solution, Dickman and Bray’s Reagent, Stannous Chloride ($SnCl_2 . 2H_2O$) Solution, Standard Phosphorous Solution

Procedure

For extraction of phosphorous, extraction buffer was added to 5 g of the soil sample and contents were shaken for 5 mins and filtered. To estimate the amount of available P 5 ml of Dickman and Bray’s Reagent was added to 5 ml of soil extract followed by addition of 7.5 ml of Boric acid then 1 ml of $SnCl_2$ and mixed thoroughly. For preparing standard curve, 5 ml of each Standard Phosphorous Solution was added instead of soil extract followed by addition of other reagents such as Dickman and Bray’s Reagent (5 ml) was Boric acid (7.5 ml of) and $SnCl_2$ (1 ml) was added and mixed thoroughly. Blank was prepared by replacing 5 ml of Standard Phosphorous Solution with 5 ml H_2O. Further the intensity of the blue color developed was recorded at 690 nm.

Calculations

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. Total RNA extraction

Procedure

Soil samples were crushed with the help of chilled vibrating agate cup milland total RNA isolation was performed according to the protocol provided with the RNA PowerSoil® Total RNA Isolation Kit (Mo Bio laboratories, Carlsbad, CA). To extract total RNA from a soil sample, 2 g of crushed soil was added to the 15 ml bead tube containing 2.5 ml of bead solution followed by addition of solution SR1 (0.25 ml) and solution SR2 (0.8 ml) into the bead tubes. Further, a mixture of phenol: chloroform: isoamyl alcohol (3.5 ml) was added to the bead tube and vortex until the disappearance of biphasic layers. Bead tubes were then centrifuged for 10 minutes at 2500 x g at room temperature and the upper aqueous layer was carefully transferred to clean collection tube. Solution SR3 (1.5 ml) was mixed to the tube containing aqueous phase, mixed together and incubated for 10 min at 4°C. After incubation, tubes were centrifuged (2500 x g) for 10 min at room temperature and transferred the supernatant to new collection tube. Solution SR4 (5 ml) was poured to the collection tube containing the supernatant and tube was incubated at room temperature for 30 min. After incubation, tube was centrifuged at room temperature at 2500 x g for 30 minutes and the supernatant was decanted. The tubes were inverted to drain out all the liquid and the pellet containing RNA was re-suspended in solution SR5 (1 ml). This RNA suspension was then transferred to the saturated column and allowed to gravity flow. The RNA capture column was then washed with solution SR5 (2 ml) and the solution was again allowed to gravity flow. The RNA column was then eluted with solution SR6 and transferred the eluted RNA to new collection tube (2.2 ml) followed by addition of solution SR4 (1 ml). The tube was further inverted to mix and stored at tube at -20°C for at least 10 minutes. After incubation, again centrifuged the tube at 13000 x g for 15 minutes at room temperature and decanted. The pellet was then air dried and re-suspended the RNA in 100 μL solution SR7. The final RNA suspension was further treated with RNase free DNase I and store at -20°C after purification.

3.5. Metagenomic DNA isolation

Total genomic DNAs were extracted by manual method (Bailly *et al.* 2007). Crushed soil samples were (0.6 - 0.65 g) transferred into 2 ml tube containing 0.5 g glass beads (0.1 mm and 0.6 mm) and poured 950 μL of denaturing/ lysis buffer solution (denaturing solution (100 μL) + lysis solution (850 μL) + β - mercaptoethanol (50 μL) + yeast tRNA 4 μL (10 mg/mL). The mixture was then centrifuged at 15000 rpm for 5 min at 4°C and transferred the supernatant into a fresh tube. Added 500 µL of phenol and 500 µL of Chloroform:Isoamyl alcohol (24:1), vortex the mixture and centrifuged at 15000 rpm for 10 min at 4°C. The supernatant was then transferred in a fresh tube and Choroform:Isoamyl alcohol (24:1) (500 µL) was added into it followed by mixing and centrifugation at 15000 rpm for 10 min at 4°C. The upper aqueous phase was transferred into fresh microcentrifuge tubes, added 1/10 volume of 3M sodium acetate and 2.5 volume of 100% ethanol to the tube. The tubes were then incubated at -70 °C and centrifuged for 15000 rpm for 25 min at 4 °C. Supernatant from the tube was discarded and pellet was air dried. The pellet was then re-suspended in ultrapure water and incubated for 15 min on ice followed by addition of 65 µL of 4 M LiCl into the tubes. Inverted the tubes 5-6 times and incubated the tubes at 4°C overnight. The tubes were then centrifuged at 15000 rpm for 25 min at 4°C and supernatant containing metagenomic DNA was collected and stored at -20 °C.

3.6. Qualitative and quantitative analysis of nucleic acids

3.6.1. Non denaturing agarose gel electrophoresis

Agarose gel (0.7% - 2% w/v) was prepared in 0.5X TBE buffer (pH 8.0) (Appendix II) and ethidium bromide (EtBr) (0.5 μg/ml) was added before pouring the gel into tray. Nucleic acids (DNA or RNA) mixed with 6X loading dye (Appendix II) were loaded into the wells on agarose gel and voltage of 1-5 V/cm was provided. Nucleic acids were migrated on the basis of charge to volume ratio and visualized under UV transilluminator (wavelength 312 nm). Equipment used for RNA samples were washed with detergent and rinsed in autoclaved double distilled water prior to use to eradicate the ribonuclease contamination.

3.6.2. Formamide agarose gel electrophoresis

Nucleic acid especially RNA has propensity to form both secondary and tertiary structures that can slow down the separation by agarose gel electrophoresis. Hence the electrophoresis of RNA required to be performed under denaturing conditions such as formamide agarose gel electrophoresis (Byrant and Manning, 1998).

Procedure

RNA sample (20-40 μg) were added into the solution of MOPS (0.5X), formaldehyde (17.5%) (v/v) and formamide (50%) (v/v) in a final volume of 25 μl and denatured at

65°C for five minutes. Prior to loading, 5X loading dye and EtBr (1 mg/ml) was added directly to the denatured samples and loaded into the wells on 1% (w/v) agarose gel prepared in 1X MOPS buffer (Appendix II) and 9% formaldehyde (v/v). RNA samples were then fractionated in the presence of 1X MOPS buffer at 60V and RNAs were then visualized under UV light.

3.6.3. Quantification of nucleic acids

Quantification of DNA and RNA were determined by measuring their O.D. at wavelength 260 nm through spectrophotometry (SAFAS UVmc2, SAFAS Monaco). One absorbance unit corresponds to approximately 40 µg/ml of RNA or single stranded DNA and approximately 50 µg/ml of double stranded DNA (Sambrook *et al.* 1989). Presence of impurities such as protein or polysaccharides in the samples was determined by evaluating the ratio between OD 260/230 nm and OD 260/280 nm respectively. This ratio for pure samples lies closer or higher than 1.8 for DNA and 2.0 for RNA. Integrity of nucleic acids especially RNA was determined by Bioanalyzer 2100 (Agilent Technologies, USA).

3.7. cDNA synthesis and size fractionation

Total RNA extracted from soil sample was further used to synthesize cDNAs using Mint-2 cDNA synthesis kit (Evrogen, Moscow, Russia) as per the protocol provided. To synthesize cDNAs, 3 µg of total soil RNA was added into a PCR tube with 10 µM of 3’-end CDS adaptor and 5’-end plugOligo-3M adaptor and incubated at 70°C for 2 min. The 5’-end plugOligo-3M adaptor comprised of (a) a stretch of oligo (dG) sequence capable of binding to the complementary oligo (dC) sequence at 3' end of the first strand cDNA by reverse transcriptase (RT) mint MMLV RT (b) Restriction site (SfiIA) and (c) M1 primer sequence. The 3' -end CDS adaptor comprised of three components (a) a stretch of oligo dT sequence that bind to poly(A) tail eukaryotic mRNAs, (b) SfiIB restriction site and (c) M1 primer sequence. Furthermore, RT master mix was prepared by adding dNTPs (10 mM), DTT (20 mM), 5X first strand buffer and IP solution and added into each vial containing total RNA and adaptors to make a final volume of 15 μL. The first strand cDNA was synthesized at 42 °C by Mint RT and synthesis of second strand cDNA was further carried out with the help of thermostable Encyclo DNA polymerase (Evrogen) via M1 primers which recognizes PlugOligo-3M and CDS-4M adaptor sequences (Figure 3.1). Reaction mixture of volume 100 μL was prepared for each first strand cDNA sample with constituents: 10X Encyclo buffer, dNTP mix (10 mM), M1 PCR primers (10 μM), 4 μL of first strand cDNA (4 μL ≈ 800 ng of total soil RNA), 50X Encyclo polymerase mix and the final volume make up with sterile RNase free water. Second strand synthesis was further followed by PCR amplification up to 3 cycles at 95 °C for 15 sec, 66 °C for 20 sec and 72 °C for 3 min. The resulting PCR product was further purified by phenol chloroform extraction and precipitated into purified form by ethanol (2.5 volume) and 1/10th volume of 3.0 M sodium acetate.

3.8. Size fractionation and validations of cDNA fractions

To achieve this, double stranded cDNAs were first exposed to three PCR cycles with same parameters to have good amount of cDNAs without affecting their average length significantly. Further, double stranded cDNA was size fractionated with the help of two

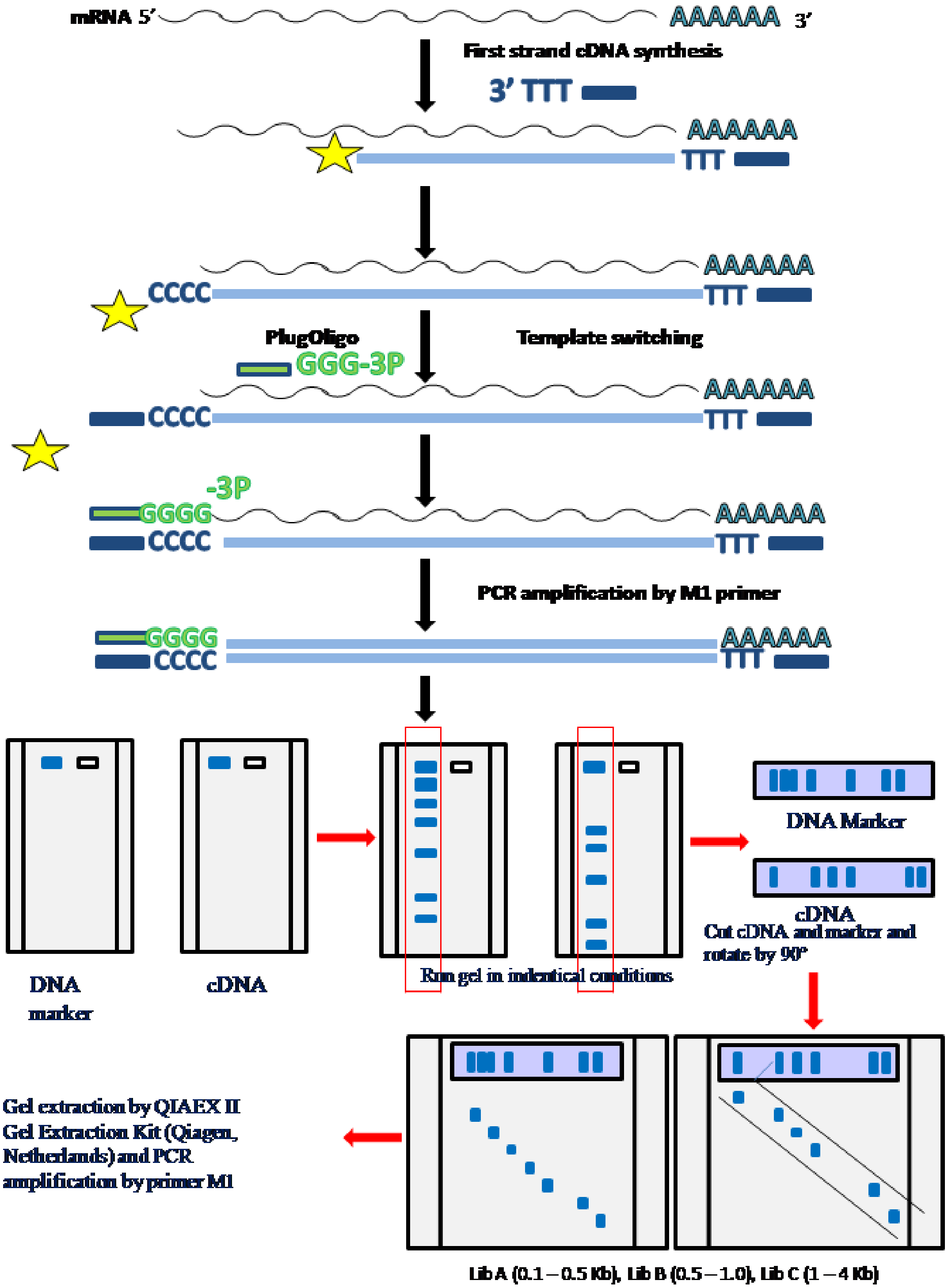


Figure 3.1: Schematic outline of Mint-2 cDNA synthesis and bi-directional size fractionation.

consecutive agarose gels of same consistency (Figure 3.1). cDNAs were loaded onto the 0.7% agarose gel along with DNA marker (1 Kb) in separate gel of same consistency under identical conditions. After electrophoresis, the size fraction ranging from 0.1 - 4 Kb was excised from both cDNA loaded lane and DNA marker lane. Both the gel slices were then rotated by 90°, placed in separate but identical gel trays and poured over by 1.4% low melting point agarose (Bioprobe Systems, Montreuil, France) around the gel slices. Electrophoresis was performed for 10 h at 2.6 V.cm-1. Agarose gel containing DNA marker was stained with EtBr and cDNA carrying unstained gel was superimposed on it to excise the size fractionated bands by visualizing in a dark reader transilluminator (Clare Chemical Research, Inc., USA). cDNAs from three gel slices of size range 0.1-0.5 Kb (fraction A); 0.5-1 Kb (fraction B) and 1-4 Kb (fraction C) were extracted and purified by QIAEX II Gel extraction kit (Qiagen, Netherlands). The purified product of three size ranges were then amplified by PCR using primer M1 under same PCR conditions with higher number of cycles (22- 30 cycles).

Different sized cDNA fractions were further validated by performing electrophoresis for each fragment in 1% agarose gel and by amplifying each fraction with different gene families representing different size groups. Each cDNA fraction (10 ng) was used as a template for amplification upto 35 cycles in a PCR machine. For the largest fraction C, i.e., >1 Kb size, primers pairs (BTf/BTr and EF1f/EF1r) corresponding to the representative gene families beta-tubulin (β-tub) and Elongation factor 1-α (EF1α) were selected for validation. Similarly, fraction B (0.5-1 Kb) was amplified by primer pair ribStF/ribStR for the gene families coding 40S Ribosomal protein S3 and primer pair MsrAf/MsrAr for gene Peptide methionine sulphoxide reductase. For the smallest fraction i.e., fraction A (< 0.5 Kb), none of the representative gene family member was found in this size range to generate primers that would work competently for metatranscriptomic cDNA fractions.

3.9. cDNA cloning

Each fraction of cDNAs (500 ng) synthesized was further digested with restriction enzyme SfiI at SfiIA and SfiIB sites located in the sequences of 5' and 3' adaptor used during first strand cDNA synthesis. After phenol chloroform extraction and purified cDNAs were ligated downstream of the phosphoglycerokinase-1 (PGK1) promoter derived from *S. cerevisiae* at SfiIA and SfiIB sites in yeast expression vector pFL61 (Figure 3.2) (Minet *et al.* 1992; Bailly *et al.* 2007; Yadav *et al.* 2014). Resulted recombinant plasmids (1 µl) were electro-transformed into electrocompetent bacterial cells (MegaX DH10B™ T1R Electrocomp™ cells, Invitrogen). To achieve this, electrocompetent *E. coli* cells were thawed on ice and added into chilled microcentrifuge tube containing 1 µl of recombinant cDNA plasmid. The recombinant plasmid and cell mix was then transferred into chilled 2 mm cuvette and tapped to ensure their complete mixing and to avoid formation of bubbles. Chilled cuvette was then exposed to electroporation in BioRad GenePulse® II electroporator under the parameters: voltage 2.5 kV; resistance 650 Ω and capacitance 25 µF.

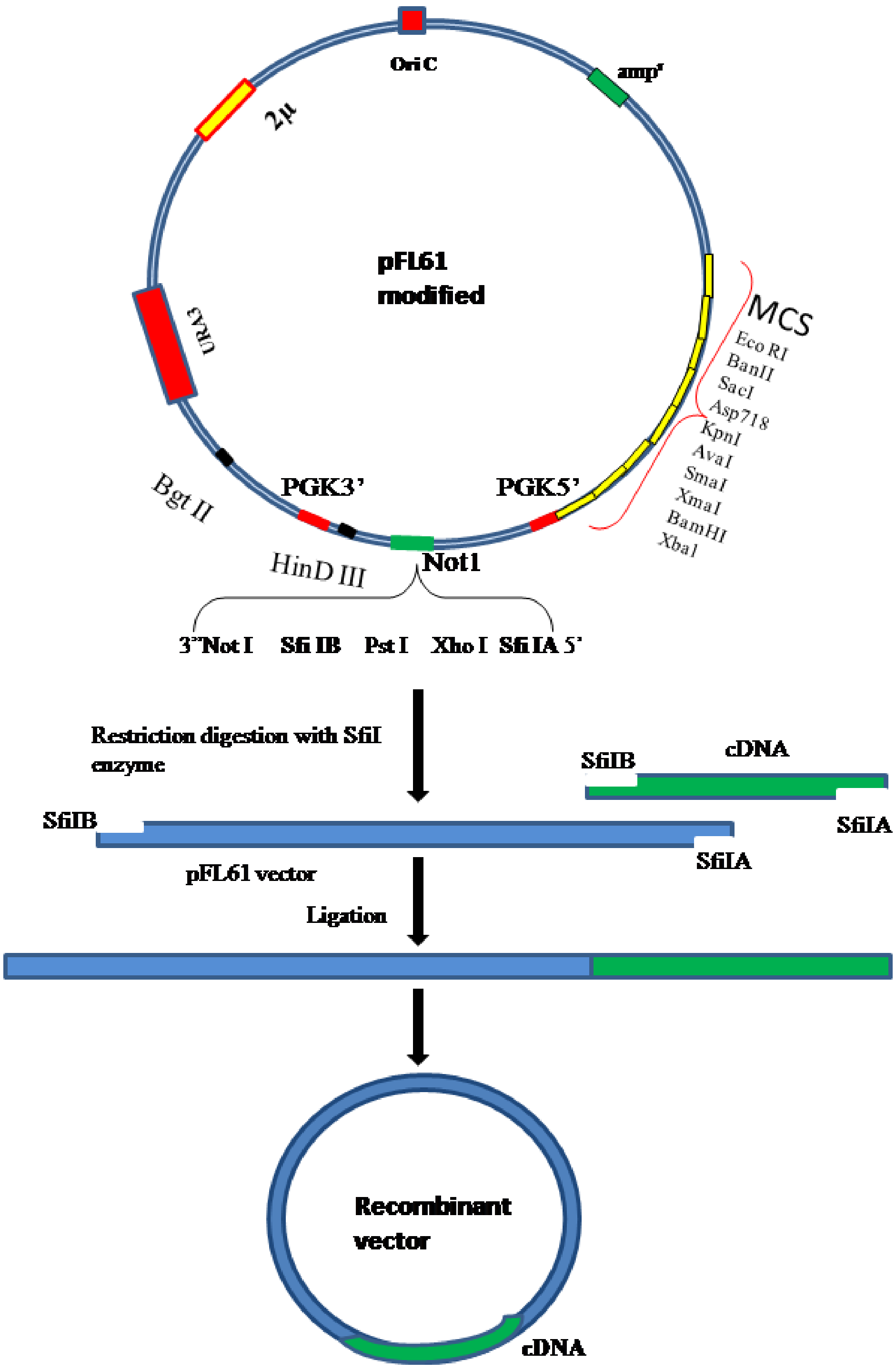


Figure 3.2: Illustration of construction of recombinant vector pFL61. Restriction digestion of cDNA and vector pFL61 with SfiI enzyme and ligation.

Simultaneously, 1 ml of S.O.C medium (Appendix I) was added to the cuvette and the cell suspension was transferred to a 15 ml falcon tube for incubation at 37°C (225 rpm) for one hour. After one hour, 100-200 μl of the culture was spread on LA plates supplemented with ampicillin (amp) (Appendix I) and incubated overnight at 37 °C. Approximately, 10^6 amp resistant transformed bacterial colonies grown on LA + amp plates were pooled together to constitute cDNA libraries corresponding to each cDNA fraction.

3.10. Plasmid isolation

Recombinant plasmids from each library A, B and C were isolated by using ZymoPURE^{TM} II Plasmid Maxiprep Kit. Pooled colonies were further transferred to flasks containing LB broth supplemented with ampicillin and incubated overnight at 37°C with vigorous shaking (300 rpm). Bacterial culture was harvested by centrifugation at 6000 x g for 15 min at 4°C. Pellet was re-suspended in 14 ml of ZymoPURE^{TM} P1 and 14 ml of ZymoPURE^{TM} P2 was added followed by immediate vigorous mixing by inverting the tube 4-6 times and placed at room temperature for 5 min. Further, ZymoPURE^{TM} P3 (14 ml) was added into the tubes, mixed thoroughly by inverting the 5-6 times followed by incubation on ice for 5 min. Centrifuged the tube for 10 minutes at maximum speed and supernatant was collected in another tube. The neutralized lysate was loaded into ZymoPURE^{TM} syringe filter and collected into fresh 50 ml tubes. ZymoPURE^{TM} binding buffer (14 ml) was added to the clarified lysate and mixed thoroughly by inverting the tube 10 times. The mixture is loaded into Zymo-Spin^{TM} V-P column and centrifuged at 500 x g for 2 mins. Added 5 ml of ZymoPURE^{TM} WASH 1 to the Zymo-Spin^{TM} V-P column and centrifuged at 500 x g for 2 mins. Added 5 ml of ZymoPURE^{TM} WASH 2 to the Zymo-Spin^{TM} V-P column and centrifuged at 500 x g for 2 min. Placed the column into a fresh collection tube and centrifuged it at ≥ 10,000 x g for one min. Transferred the column into clean microcentrifuge tube and added 400 µl of ZymoPURE^{TM} Elution buffer to column. Incubated for 2 min and centrifuged at ≥ 10,000 x g for one min. Stored the eluted final product in a microcentrifuge at -20ºC for further use.

3.11. Colony PCR

Colony PCR was performed to assure the success of transformation and presence of recombinant plasmids inside both bacterial as well as yeast cells. A pin point of bacterial colony was mixed into 5 µL of ultrapure water and incubated at 95ºC for 5 min to cell lysis while in case of yeast cells, cells were suspended in 20 mM of NaOH and incubated at 95ºC for 20 min. After cell lysis, the eppendorfs tubes were placed on ice immediately, vortexed for five sec and centrifuged. Supernatant (3 µL) was then transferred into fresh PCR tube and PCR mix was added. PCR mix was prepared as given below:

Content	**Final Concentration**	**Volume in µL**
10X buffer	1X	1.5 µl
$MgCl_2$.	0.5 or 1.0 mM	1 µl
2 mM dNTPs	200 µM	2 µl
Forward Primer NF	0.2 µM	1 µl
Reverse Primer NR	0.2 µM	1 µl
Template	<1,000 ng	variable
Taq Polymerase	1.25 U	µl
Nuclease free H_2O	NA	Final volume 20 µl

Amplification was performed as follows: Initial denaturation at 94ºC for 5 minutes, followed by 25 cycles of 94ºC for 30 seconds, annealing at 55ºC for 1 min, extension at 72ºC followed by final extension at 72ºC for 10 min. After PCR amplification, the amplified products were then visualized on 0.8% agarose gel via electrophoresis.

3.12. Yeast Transformation

Different sized cDNA libraries were further screened for heterologous complementation of cDNA insert in hypersensitive yeast strains of *S. cerevisiae* (Figure 3.3). The cDNA libraries were transformed into these metal hypersensitive strains with the help of Li-Ac method (Gietz and Schiestl, 2007).

Procedure

Primary Culture (Day 1):

For primary culture, single colony of yeast mutant was transferred from freshly streaked yeast extract peptone dextrose (YEPD) plate into YEPD broth and incubated at 30ºC overnight at 200 rpm.

Secondary Culture (Day 2):

Secondary culture was prepared by inoculated 50 µl of primary culture into 100 ml Erlenmeyer flask containing 20 mL of YPED broth and incubated at 30ºC at 200 rpm overnight along with two flasks containing 40 mL of the YEPD broth media for further use.

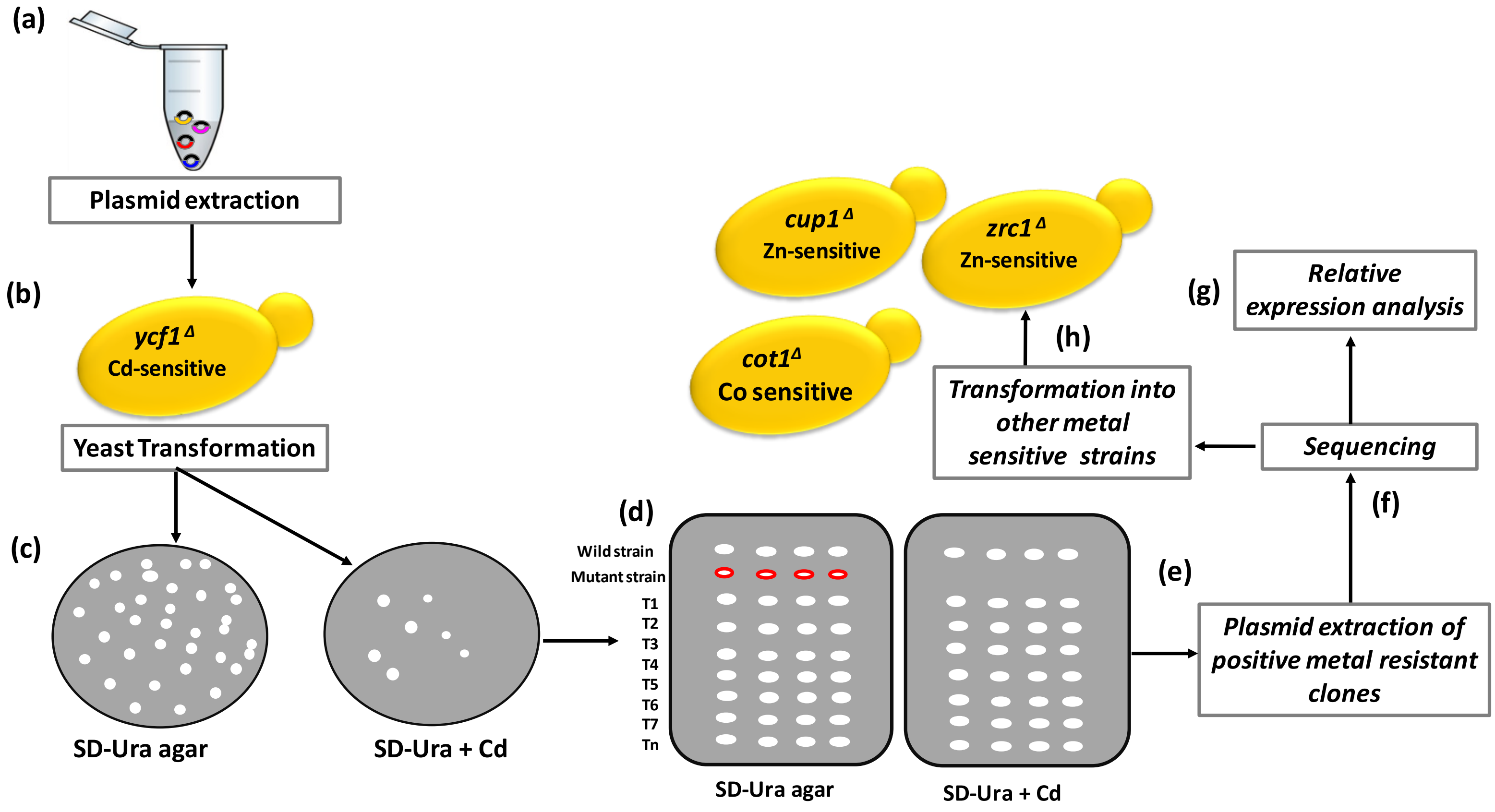


Figure 3.3: Illustration of functional screening of soil metatranscriptomic cDNA library: (a) Plasmid extraction from transformed bacteria; (b) Screening of libraries by yeast complementation ; (c) Selection of transformants on Cd amended media plates; (d) Drop out assay; (e & f) Recombinant plasmid extraction from resistant clones and their sequencing; (g) Relative expression of Cd tolerant cDNAs in the presence of Cd and (h) Transformation of cDNAs into other metal sensitive yeast mutants for further characterization.

Day 3:

The growth of secondary culture was then measured at 600 nm by making a dilution of 1/10 to avoid the saturation of spectrophotometer due to dense growth in the broth (If OD600 is 0.3 then the final OD will be considered as 0.3x10= 3). After measuring the growth, preheated 40 ml YEPD media (day 2) was inoculated with appropriate volume of secondary culture to achieve a final O.D.600 = 1 and final volume 50 ml with sterile YEPD media. The flask was incubated to achieve O.D.600 = 2 i.e., 2×10^7 (for around two hrs approximately) and the culture was then transferred in 50 ml falcon tubes. Falcons were then centrifuged at 3000 rpm for 5 min and pellet washing was done after subsequent centrifugation steps repeated at least thrice. The pellet received after centrifugation was then re-suspended with 1 ml of water and suspension was transferred to 1.5 mL microcentrifuge tube. The tube was centrifuged for 30 sec in bench centrifuge and supernatant was discarded. The pellet was again re-suspended in 1 ml water and aliquots of 100 µL each were prepared. Aliquots were then centrifuged for 30 sec with bench centrifuge, discarded the supernanat and added the following mixture into it:

Constituents	**1x**
PEG 3500 50% w/v	240 µL
LiAc 1M	36 µL
Salmon sperm DNA 10 mg/mL (boiled 5 min and quick chilled on ice)	10 µL
DNA (1 µL) + H_2O	74 µL
Total	**360 µL**

Cell suspension and transformation mixture carrying tube was then incubated for 1 hr at 42ºC in water bath. The tubes were then centrifuged for 30 sec with bench centrifuge and the supernatant was discarded. The pellet was then re-suspended with 1 ml water and the cell suspension (100 µl) was plated on the selective media (SD - Ura) plates and remaining 900 µl on SD - Ura + metal plates (300 µl on each plate). The plates were then incubated for 48-72 hrs at 30ºC.

3.13. 5-Flouroorotic acid (5-FOA) test

The 5-Flouroorotic acid (5-FOA) test is an extremely useful method of selection for various genetic and molecular biological manipulations requiring the detection of rare *ura3-* cells. 5-FOA is converted into a toxic product 5-fluoro-uracil in the presence of *URA3* gene product which is lethal to the yeast cells. Hence, cells have to eject or flush out the plasmid to survive on 5-FOA containing media which subsequently result into losing their ability to tolerate toxic concentrations of potentially toxic metals such as Cd (Figure 3.4). This test was performed by streaking the transformants on SD + Ura + FOA (1g/L) plates and incubated for 3-5 days at 30ºC. The transformants grown on SD + Ura + FOA plates was checked for their growth on following media:

Media	**Observations**
SD - Ura	Should not grow as no Ura3 plasmid is there due to flushing out of the Ura3 plasmid and if it is still growing then there is some mutation in the genome.
SD + Ura	Should grow
SD + Ura + Metal	Should not grow as Ura3 plasmid carrying metal tolerant insert is absent and if it is growing then there is some mutation in the genome.

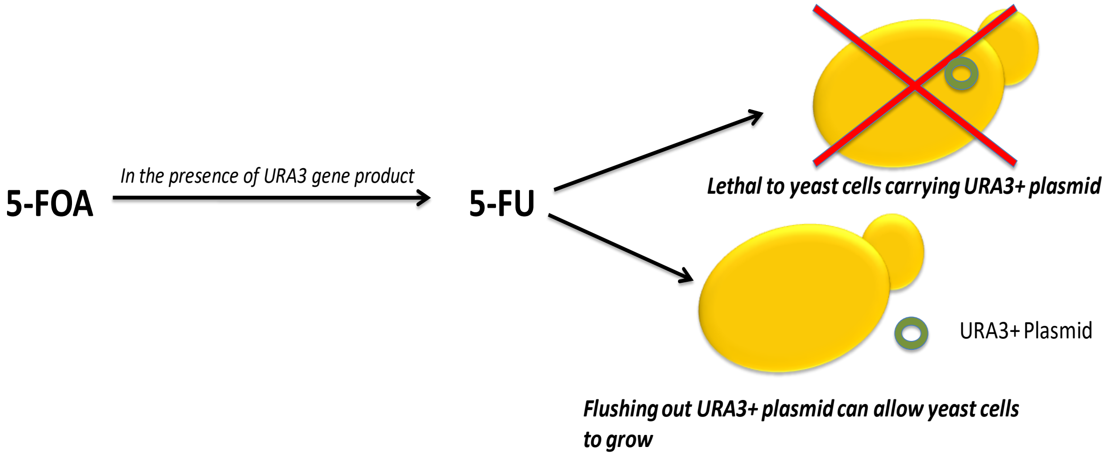


Figure 3.4: Illustration of impact of 5-Flouroorotic acid (5-FOA) on yeast cells carrying *URA3*$^+$ gene. 5-FOA is converted into a toxic product 5-fluoro-uracil in the presence of *URA3*$^+$ gene product which is lethal to the yeast cells and can only survive by ejecting or flushing out of plasmid.

3.14. Yeast complementation assay

Screened Cd tolerant uracil auxotropic yeast transformants were further tested for their ability to tolerate toxic concentrations of Cd, a potentially toxic metal, by two functional complementation assays: (i) Drop out assay (ii) Growth assay. Functional complementation assays on metal amended media provides us knowledge about capability of the screened Cd tolerant cDNA in terms of both different cell concentrations (culture dilutions) and different metal concentrations in both solidified media and liquid media.

3.14.1. Drop out assay

Screened Cd tolerant yeast transformants were inoculated from fresh plates into test tubes containing 10 mL SD-Ura broth and incubated at 30ºC and 220 rpm for 24 hrs. The O.D.

600 was measured for the cultures and set to the level of 1.0. Dilutions of pre-grown cultures with O.D. 600 of 1.0 were made up to 0.0001 with SD-Ura broth and 5.0 µl of each dilution was dropped on SD-Ura control plates and SD-Ura supplemented with 40 µM $CdSO_4$ (Figure 3.5). These plates were incubated at 30ºC for 48 hrs.

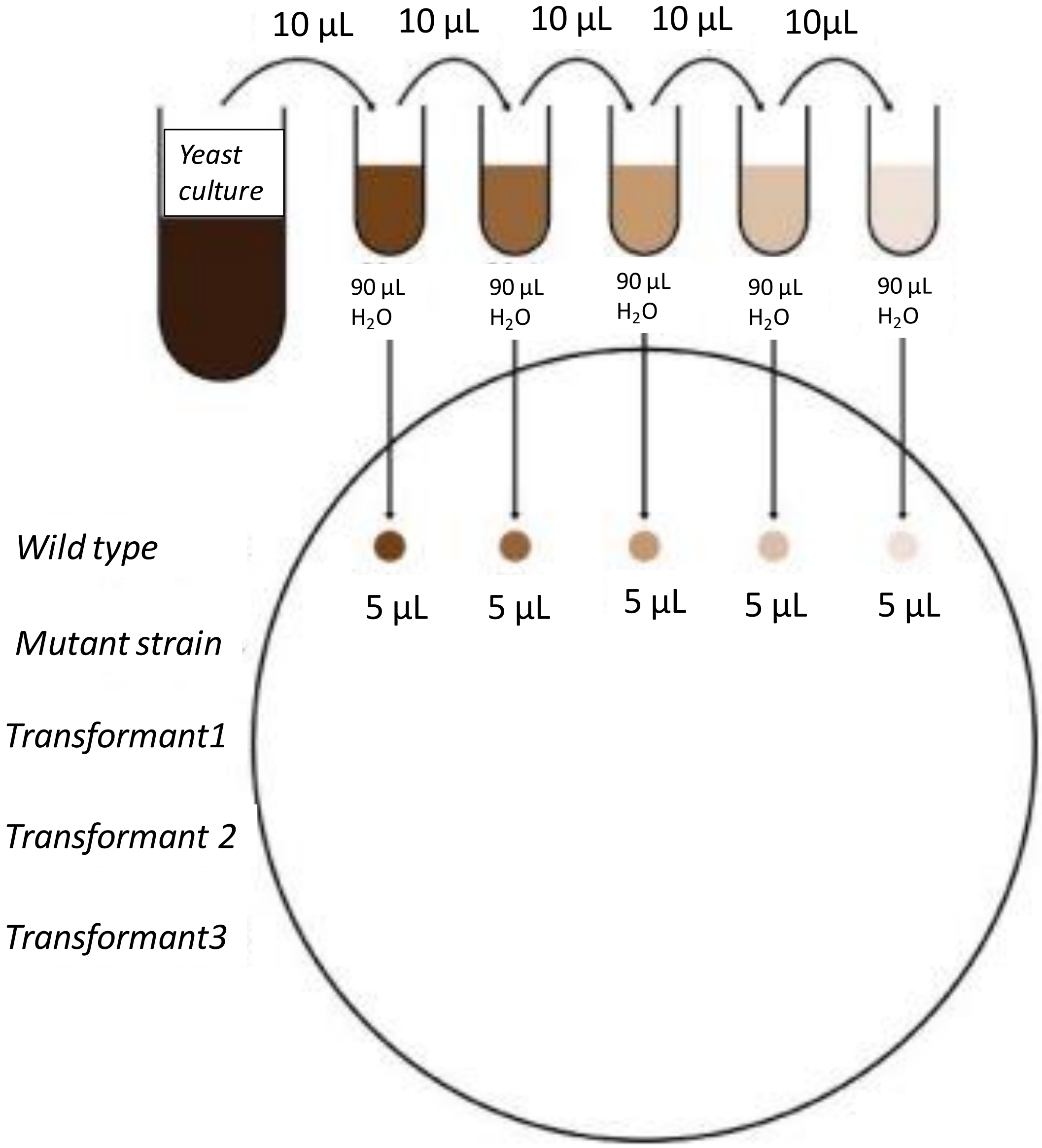


Figure 3.5: Illustration of drop out assay performed to screen metatranscriptomic cDNA library for Cd tolerant cDNAs.

3.15. Yeast Plasmid isolation

Yeast plasmid was isolated from the screened yeast transformants carrying Cd tolerant cDNAs by using Zymoprep™ Yeast Plasmid Miniprep II (Zymo Research, USA) as per the protocol provided.

Procedure

To isolate recombinant plasmid from the yeast cells, the yeast transformants were inoculated from fresh culture plates into 5 mL SD-Ura broth and incubated for 24 hrs. Fresh culture (0.1 ml – 0.5 ml) was transferred into microcentrifuge tube (1.5 ml) and centrifuged for 2 min at 600 x g. Supernatant was discarded and re-suspended the pellet by adding solution 1 (200 µl) and cell lysis was initiated by adding 3 µl of Zymolase^{TM} to each tube. Re-suspended the pellet by vortex and incubated the tubes at 37ºC for one hr. After incubation, 200 µl of the solution 2 was added to the tubes and addition of the solution 3 (400 µl) followed by mixing the cell suspension. The tubes were then centrifuged at maximum speed for around 3 min and transferred the supernatant to new Zymo-Spin-1 column tubes. Spin the column for 30 sec and washed the column with 550 µl of wash buffer and again spin for 60 sec. The Zymo-Spin-1 column was then placed on a fresh microcentrifuge tube and 10 µl of the TE buffer was added to it. Again spin the column for 30-60 sec to elude plasmid into fresh microcentrifuge tube.

Isolated plasmids (2-3 µl) were then electro-transformed into One Shot® TOP10 Electrocomp™ *E. coli* cells (Thermofisher, USA) through electroporator (BioRad

GenePulse® II electroporator). Bacterial colonies were selected on LA + amp plates and further sub-cultured on LA + amp plates.

3.16. Bacterial plasmid Isolation

Bacterial plasmid was isolated with the help of QIAprep Spin Miniprep Kit (Qiagen, Germany) as per the protocol provided.

Procedure

To isolate recombinant plasmid from transformed bacterial cells, overnight bacterial culture was harvested by centrifugation at > 8000 rpm (6800 x g) for 3 min. the pellet retrieved was then re-suspended in buffer P1 (250 µL) followed by addition of buffer P2 after few min and the tubes were inverted 5-6 times. Further, buffer N3 (350 µL) was added to the tube and mixed thoroughly by inverting the tube 5-6 times again. The tubes were centrifuged for 10 min at 13,000 rpm and 800 µL of the supernatant was then poured into the QIAprep 2.0 spin column. Centrifugation of the column for 30 - 60 sec was followed by washing the column with buffer PB (0.5 mL) and a centrifugation for 30-60 sec. The column was then washed with buffer PE (0.75 mL) and centrifuged for 30-60 sec. Centrifugation was then repeated to remove out all the residual volume from the column. The QIAprep 2.0 column was transferred to fresh microcentrifuge tube and eluted the DNA with 50 µl of EB buffer or water. Purity and quantity of extracted plasmid DNA was determined by agarose gel electrphoresis.

3.17. Bioinformatic analysis

Recombinant plasmid isolated from the electro-transformed bacterial cells was further sequenced to identify the taxonomic features of cDNA insert. The sequencing of cDNA insert was done by vector specific primers NF and NR (Table 3.2). The sequences were compared for homologous sequences by using BLASTX analysis of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The cDNA sequences were further analyzed for presence of full length ORFs using https://www.ncbi.nlm.nih.gov/orffinder/. Homologous sequences along with cDNA sequences were aligned by multiple sequence alignment using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Secondary structure of the deduced polypeptide was determined by http://bioinf.cs.ucl.ac.uk/psipred/ to know the localized organization among the amino acids due to various assumed different spatial arrangements. The potential O- and N- linked glycosylation sites if present in cDNAs, were predicted using NetOglyc 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc-3.1/) and NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/). Potential phosphorylation sites present in the polypeptide were determined by NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/). The sequences obtained in this study were deposited at NCBI under the respective accession numbers.

Phylogenetic analysis of full ORF carrying cDNAs was done using MEGA X (Kumar *et al.* 2018) with the sequences from various taxa corresponding to the similar genes and the phylogenetic tree was reconstructed using neighbor joining algorithm employing 1000 bootstrap replicates.

3.18. Growth assay

Growth assay was performed to analyze the tolerance profile of tolerant cDNA with respect to the different ranges of metal concentrations. To achieve this, flasks containing 20 mL of SD-Ura broth were inoculated with the pre-cultures of the screened transformants and incubated for 6 hr at 30ºC and 220 rpm. Cadmium metal in the final concentration of 40 μM, 60 μM and 80 μM was added into these flasks and allowed to grow for next 42 hr. Measured the 1/10 diluted cell density at 600 nm after 48 hr. Similarly, growth assay for other potentially toxic metals i.e., Cu, Zn and Co were performed by first transforming the isolated recombinant plasmid carrying Cd tolerant cDNA into respective yeast mutants i.e., *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$. Screened Cu, Zn and Co tolerant transformants were then sub-cultured on SD-Ura plates amended with respective metal. These transformants were then inoculated into 5 mL of SD-Ura broth amended with respective metal. For Cu metal, the final concentration of 150 μM, 300 μM and 500 μM was added into these flasks, for Zn metal, the concentration range is 10 mM, 11 mM and 12 mM while concentration for Co metal ranges from 2 mM, 3mM and 4 mM. Each flask containing metal amended media inoculated with transformant was allowed to grow for next 42 hr and 1/10 diluted cell density was measured at 600 nm after 48 hr.

Wild strain BY4741 transformed with pFL61 vector was used as positive control while the respective mutant strains transformed with pFL61 was also used as negative controls in both the experiments. Both positive and negative controls were executed in parallel to the growth assay study for Cd, Cu, Zn and Co tolerant transformants.

3.19. Quantitative reverse transcription PCR

Cadmium sensitive yeast mutant *ycf1*$^{\Delta}$ carrying tolerant cDNA was further tested for the level of expression of transformed cDNA in comparison to housekeeping genes of *S. cerevisiae* under similar parameters. To accomplish this, flasks containing 50 ml of SD-Ura broth media amended with different concentrations of Cd metal (0, 40, 60 and 80 μM) was inoculated with pre-culture of Cd tolerant transformant and incubated for 48 hrs at 30ºC. Cells were pellet out by centrifugation at 3000 r.p.m for 5 mins and grounded with liquid nitrogen. Total RNA was isolated by using Trizol reagent and cDNA was synthesized by using PrimeScript™ first strand cDNA synthesis kit (Takara, Japan) as per instruction manual. Gene expression analysis of Cd tolerant cDNA was performed by using SYBR® Premix Ex Taq™ (Tli RNase H Plus) (Takara, Japan) as directed in user's manual with a final volume of 20 μl (100 times diluted RT cDNA 6.8 μl, 0.2 μM forward primer, 0.2 μM reverse primer. Amplification programme used was: 95ºC for 2 mins (1 cycle), 95ºC for 15 sec, 54ºC for 15 sec and 68ºC for 20 sec (35 cycles). The qPCR primers for housekeeping genes (Table 3.2) are designed from *ACT* (Actin; SGD:S000001855), *TAF10* (TATA binding protein-Associated Factor; SGD:S000002574) and *TFC1* (Transcription Factor class C; SGD:S000000327) as controls (Teste *et al.* 2009) and their corresponding expression analysis via Normfinder suggested that most stable expression and minimum stability value under Cd stress was for *TFC1*. Hence in this study, *TFC1* was used as reference genes to calculate comparative expression level of genes by using the formula $2^{-\Delta\Delta CT}$ (Livak and Schmittgen, 2001). For all these measurements, RNA was extracted from three independent biological samples known as experimental replicates and three technical replicates were performed for each of the RNA replicates.

3.20. Inductively coupled plasma mass spectrophotometry

Hypersensitive yeast mutant *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying tolerant cDNAs were further tested for the level of accumulation of respective metal used in the study. To achieve this, each flask containing 50 ml of SD-Ura media supplemented with different range of metal concentrations Cd (40 µM - 80 µM), Cu (150 µM - 500 µM), Co (2 mM - 4 mM) and Zn (10 mM -12 mM), were inoculated with exponential phase cultures of transformed *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *cot1*$^{\Delta}$ and *zrc1*$^{\Delta}$ cells and incubated at 30 °C at 220 rpm. Cell density of grown cultures was measured after 48 hrs and cells were harvested. Metal accumulation by specific yeast transformant culture was further determined by acid digestion of cell pellet (1 gm) by nitric acid and perchloric acid (2:1, v/v) (Hseu, 2004). Digested product was leveled up to the final volume of 50 ml with distilled water and estimated for quantity of metal available by ICP-MS (Element XR, Thermo Fisher Scientific, Germany).

Table 3.2: PCR primers designed for various experiments conducted in this work

Primer	**Sequence**	**Specificity**
PCR primers for validation used for validation of cDNAs		
BTf	5’-GGTAACCAAATCGGTGCTGCTTTC-3’	β-Tubulin for library C
BTr	5’-ACCCTCAGTGTAGTGACCCTTGGC-3’	β-Tubulin for library C
EF1f	5’-GTCGTYGTYATYGGHCAYGT-3’	Elongation factor 1-alpha for library C
EF1r	5’-TGYTCNCGRGTYTGNCCRTCYTT-3’	Elongation factor 1-alpha for library C
ribStf	5’-CHSKHACYGABRTCATCATCCG-3’	40S Ribosomal protein S3 for library B
ribStR	5’-AADCCRTCRGTGAACTTCATG-3’	40S Ribosomal protein S3 for library B
MsrAf	5’-CGCCGCCGGCTGYTTYTGGGG-3’	Peptide methionine sulphoxide reductase for library B
MsrAf	5’-ATRGTRGTYNWCATGGACCTGTTCTTGGGGC-3’	Peptide methionine sulphoxide reductase for library B
pFL61 vector specific primers used for complementation studies		
pFL61 NF	5’-CAGATCATCAAGGAAGTAATTATCTAC-3’	Specific for vector pFL61
pFL61 NR	5’-CAGAAAAGCAGGCTGGG AAGC-3’	Specific for vector pFL61
qPCR primers for yeast (*S. cerevisae*) used for relative gene expression studies		

qACTF	5’-CGAATTGAGAGTTGCCCCAG-3’	Actin
qACTR	5’-CAAGGACAAAACGGCTTGGA-3’	Actin
qTAF10F	5’-GCTAACAACAGTCAGGCGAG-3’	TATA binding protein-Associated Factor
qTAF10R	5’-GAGCCCGTATTCAGCAACAG-3’	TATA binding protein-Associated Factor
qTFC1F	5’-ACGACGCTGCTTTGGAAAAT-3’	Transcription Factor class C
qTFC1R	5’-GCTTTTCATTGTTTCCGCGG-3’	Transcription Factor class C
qubiqF	5’-CGTATCAGCAGCGTCTCATC-3’	cDNA PLBe1 (UFP)
qubiqR	5’-CAATCGCCAGTGATGAGGAC-3’	cDNA PLBe1 (UFP)
qvitF	5’-ACCGTCCAAGTTCAATTGCC-3’	cDNA PLCe10 (vWD)
qvitR	5’-GAATGTCCTGTTCGTCGGTG-3’	cDNA PLCe10 (vWD)
qHspF	5’- GACCGTATGCCCCAAATGTC 3’	cDNA PLCc43 (Hsp40)
qHspR	5’-GGGAATTACACCAGGCTCCT-3’	cDNA PLCc43 (Hsp40)
qGSHF	5’-AGCAGCTCCCACCTCTATTG-3’	cDNA PLCd43 (PCS)
qGSHR	5’-TGAGAGGGAGATGCAAGCAA -3’	cDNA PLCd43 (PCS)

3.21. Functional diversity

To study the diversity of sampling site, PL soil samples from Agro-forestry land at Pierrelaye (PL) in North-West of Paris, France were processed for nucleic acid extraction.

3.21.1. Nucleic acid extraction and library preparation

Total genomic DNAs were extracted by manual method as described in Section 3.3 (Bailly *et al.* 2007) and total RNA were extracted from the frozen soil samples by RNA PowerSoil® Total RNA Isolation Kit (Mo Bio laboratories catalogue no 12866-25, Carlsbad, CA) as per the instructions provided. Extracted total genomic DNA was further treated with RNaseA and total RNA was treated with DNaseI to remove any DNA and RNA impurities. Quality of DNA and RNA were analyzed by agarose gel electrophoresis and quantity and purity were determined by spectrophotometry (SAFAS UVmc2, SAFAS Monaco). First strand cDNA from total RNA was synthesized using random hexamers and SuperScript reverse transcriptase (Invitrogen). Both types of nucleic acids, genomic DNA as well as cDNAs extracted from soil samples were amplified by different sets of degenerate primers specific to 380 bp V4 region of 18S rDNA gene and its transcribed rRNA (Table 3.3) (Hadziavdic *et al.* 2014). Each forward and reverse degenerate primer consists of an adaptor sequence for MiSeq sequencing (provided by IGA Technology Services S.r.l. Unipersonale Via Jacopo Linussio, 51, 33100, Udine Z.I.U., Italy), four random nucleotide sequences known as linker sequences, followed by an eight nucleotide sample specific indexed sequence and followed by 18S rRNA V4 specific forward primer sequence of twenty nucleotides or reverse primer sequence of eighteen nucleotides (Table 3.4) (Manoharan *et al.* 2017). The purpose of adaptor sequence is to enable the DNA sequence attachment to the MiSeq flow cell, linker sequence increases the overall melting temperature while sample specific indexed sequences was used to unveil the diversity present in soil sample.

Table 3.3: Soil samples with two different set of primers used for amplification of V4 region of 18S rDNA and 18S rRNA

Soil Samples with two different sets of primers					
S. No	**SET1**	**SET2**	**S. No**	**SET1**	**SET2**
1	F1R1	F10R9	**8**	F3R6	F4R1
2	F4R3	F2R2	**9**	F3R7	F4R2
3	F7R6	F5R4	**10**	F3R8	F4R4
4	F6R5	F8R7	**11**	F1R6	F10R5
5	F3R3	F9R9	**12**	F9R11	F9R8
6	F11R6	F10R6	**13**	F8R9	F10R4
7	F3R5	F5R3	**14**	F9R10	F10R3

The PCR was performed in two stages: first PCR step includes ten PCR cycles of the samples with primers containing Illumina sequencing primer, sample specific index sequences and gene specific primer sequences. Second PCR of 23 cycles was performed with primer carrying Illumina adapter sequence and gene specific primer sequence. The PCR was performed in two stages : step 1 94ºC - 3 min ten cycles of 94ºC - 10 sec, 53ºC - 30 sec, 72ºC - 40 sec followed by 23 cycles of 94ºC - 10 sec, 48ºC - 30 sec, 72ºC - 40 sec then final extension at 72ºC for 2 min. These amplified products (cDNA and genomic DNA amplicons) were then purified by Qiaquick PCR purification kit (Qiagen, Netherlands) and mixed together in equimolar concentrations. Pooled amplified products were then sent for NGS analysis by high throughput sequencing on Illumina MiSeq platform. For each soil sample, amplification was done with two different sets of degenerate primers (Table 3.3 & 3.4) and each amplification experiments was performed in two technical replicates.

3.21.2. NGS Analysis

Further, the NGS reads we received were assumed to be rRNA derived as the amplification was done using various sets of degenerate primers specific for V4 region of 18S rRNA. A paired end (PE) sequencing chemistry was used which further produced raw reads of equal length. The raw reads were quality checked using FastQC which includes GC bias, K-mer quality, duplication levels. After quality check of the reads, the datasets were analysed using Mothur (v.1.38.1) (Figure 3.6) and the pipeline used for analysis followed standard operating procedure from Schloss et al. (2009). During Mothur runs, forward and reverse reads R1 and R2 were combined using (make.contigs) command. Sequences that were not able to accomplish any one of the following norms were excluded: maximum read length of 600 bases, the presence of ambiguities and nucleotide mismatch (more than 1) to the primer using (screen.seqs) command. Duplicate sequences were removed using (unique.seqs) command. The sequences were customized to the targeted V3-V4 regions using pcr.seqs command and the unique sequences retrieved were further aligned from SILVA reference database. The columns containing only gap characters were taken out using (filter.seqs) command with no loss of any information followed by pre-clustering of sequences using (pre.cluster) command. Chimeras were detected using UCHIME algorithm and taxonomic classification of seque-

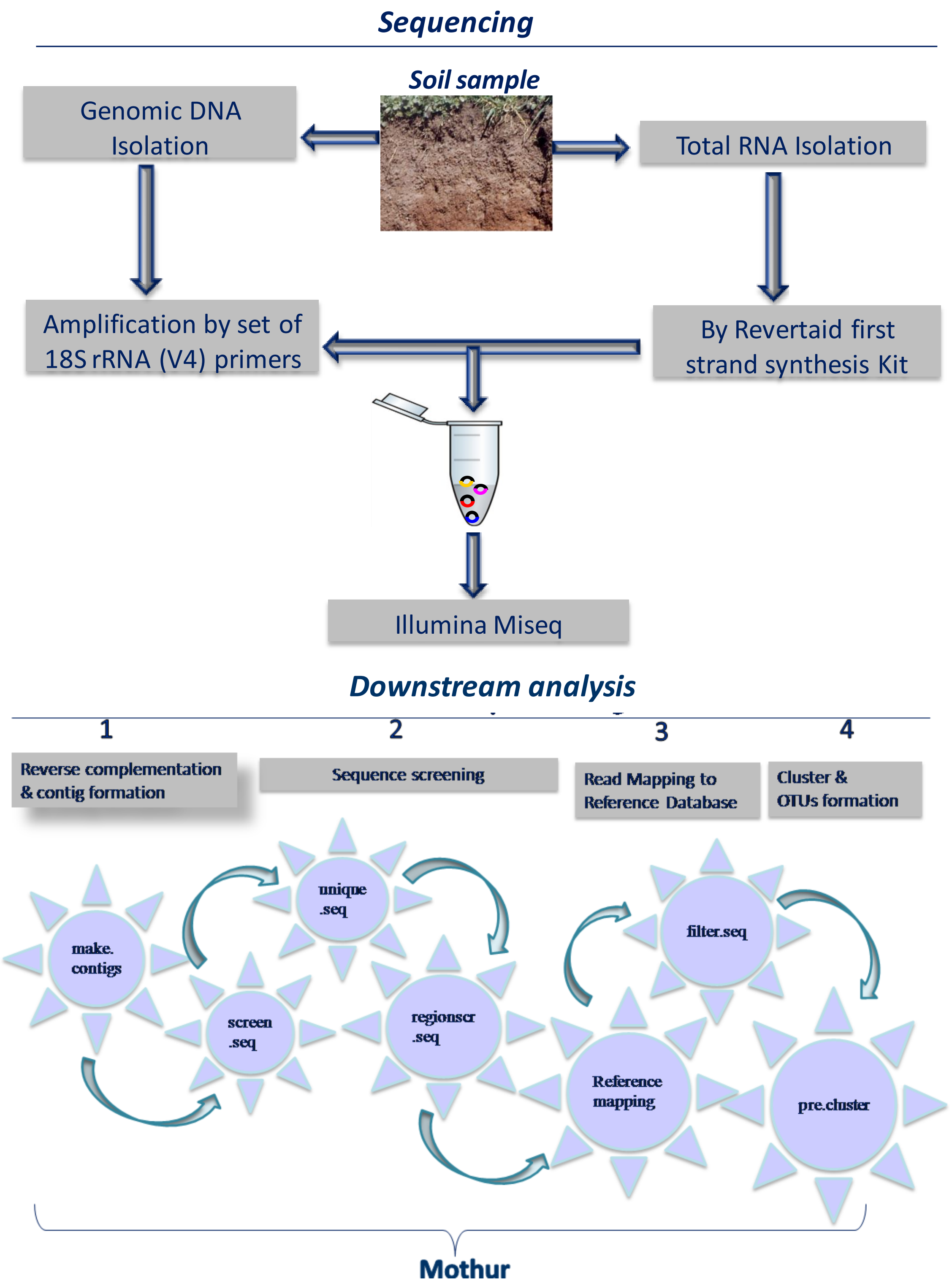


Figure 3.6: Workflow for downstream analysis of datasets: Sequenced data was analyzed using MOTHUR software (v.1.38.1)

nces was performed using the Naive Bayesian classification (80% confidence threshold) by (classify.seqs) command and those that were not classified to any one of the domains were eliminated using (remove.lineage) command. We considered each sample as a separate community.

3.21.3. Analysing diversity and phylotypes

Sequences were analyzed into phylotypes according to their taxonomic classification using (phylotype) command. A distance matrix was created and the sequences were clustered into operational taxonomic units (OTUs) at 3% dissimilarity cut-off (97% similarity) using (dist.seqs) and (cluster) command, respectively. The cut-off numbering of the phylotypes equals 4 which corresponded to order level used in (cluster.split) command. The number of sequences in each OTU was determined using (make.shared) command. The taxonomy for each OTU was specified using (classify.otu) command. Rarefaction curves describe the number of OTUs observed as a function of sampling effort which were performed using (rarefaction.single) command. The community richness and diversity indices were performed using summary.single command before heatmap as a graphical representation of the abundance of each OTU was generated.

3.22. Statistical analysis

All the data obtained were subjected to analysis of variance (ANOVA) and the significant differences among the means were compared with Tukey’s Test $p<0.05$ using GraphPad Prism version 8.0.

Table 3.4: Different sets of primers used for amplification of V4 region of 18S rDNA and 18S rRNA

Sample & Primer Set ID	Forward Primer Sequence	Reverse Primer Sequence
F1R1	NNNNacacacacCCAGCASCYGCGGTAATTCC	NNNNgatcgcgaACTTTCGTTCTTGATYRA
F10R9	NNNNgtacgactCCAGCASCYGCGGTAATTCC	NNNNcatcagtcACTTTCGTTCTTGATYRA
F4R3	NNNNtatgtcagCCAGCASCYGCGGTAATTCC	NNNNgtcgtagaACTTTCGTTCTTGATYRA
F2R2	NNNNacagcacaCCAGCASCYGCGGTAATTCC	NNNNcgctctcgACTTTCGTTCTTGATYRA
F7R6	NNNNactagatcCCAGCASCYGCGGTAATTCC	NNNNagactatgACTTTCGTTCTTGATYRA
F5R4	NNNNtagtcgcaCCAGCASCYGCGGTAATTCC	NNNNgtcacgtcACTTTCGTTCTTGATYRA
F6R5	NNNNtactatacCCAGCASCYGCGGTAATTCC	NNNNgactgatgACTTTCGTTCTTGATYRA
F8R7	NNNNtgacatcaCCAGCASCYGCGGTAATTCC	NNNNgcgtcagcACTTTCGTTCTTGATYRA
F3R3	NNNNgtgtacatCCAGCASCYGCGGTAATTCC	NNNNgtcgtagaACTTTCGTTCTTGATYRA
F9R9	NNNNacatgtgtCCAGCASCYGCGGTAATTCC	NNNNcatcagtcACTTTCGTTCTTGATYRA
F11R6	NNNNatgatcgcCCAGCASCYGCGGTAATTCC	NNNNagactatgACTTTCGTTCTTGATYRA
F10R6	NNNNgtacgactCCAGCASCYGCGGTAATTCC	NNNNagactatgACTTTCGTTCTTGATYRA
F3R5	NNNNgtgtacatCCAGCASCYGCGGTAATTCC	NNNNgactgatgACTTTCGTTCTTGATYRA
F5R3	NNNNtagtcgcaCCAGCASCYGCGGTAATTCC	NNNNgtcgtagaACTTTCGTTCTTGATYRA
F3R6	NNNNgtgtacatCCAGCASCYGCGGTAATTCC	NNNNagactatgACTTTCGTTCTTGATYRA
F4R1	NNNNtatgtcagCCAGCASCYGCGGTAATTCC	NNNNgatcgcgaACTTTCGTTCTTGATYRA

F3R7	NNNNgtgtacatCCAGCASCYGCGGTAATTCC	NNNNgcgtcagcACTTTCGTTCTTGATYRA
F4R2	NNNNtatgtcagCCAGCASCYGCGGTAATTCC	NNNNcgctctcgACTTTCGTTCTTGATYRA
F3R8	NNNNgtgtacatCCAGCASCYGCGGTAATTCC	NNNNacgacgagACTTTCGTTCTTGATYRA
F4R4	NNNNtatgtcagCCAGCASCYGCGGTAATTCC	NNNNgtcacgtcACTTTCGTTCTTGATYRA
F1R6	NNNNacacacacCCAGCASCYGCGGTAATTCC	NNNNagactatgACTTTCGTTCTTGATYRA
F10R5	NNNNgtacgactCCAGCASCYGCGGTAATTCC	NNNNgactgatgACTTTCGTTCTTGATYRA
F9R11	NNNNacatgtgtCCAGCASCYGCGGTAATTCC	NNNNtctactgaACTTTCGTTCTTGATYRA
F9R8	NNNNacatgtgtCCAGCASCYGCGGTAATTCC	NNNNacgacgagACTTTCGTTCTTGATYRA
F8R9	NNNNtgacatcaCCAGCASCYGCGGTAATTCC	NNNNcatcagtcACTTTCGTTCTTGATYRA
F10R4	NNNNgtacgactCCAGCASCYGCGGTAATTCC	NNNNgtcacgtcACTTTCGTTCTTGATYRA
F9R10	NNNNacatgtgtCCAGCASCYGCGGTAATTCC	NNNNatcagtcaACTTTCGTTCTTGATYRA
F10R3	NNNNgtacgactCCAGCASCYGCGGTAATTCC	NNNNgtcgtagaACTTTCGTTCTTGATYRA

RESULTS AND DISCUSSION

4.1. Construction of cDNA libraries from RNA isolated from metal polluted soil

4.1.1. Soil sampling and analysis

Soil samples collected from Agro-forestry land in Pierrelaye (PL) (49°1′45”N, 2°10′32″E) in northwest of Paris, France were tested for various physico-chemical properties (Figure 1). The soil possessed a sandy luvisol texture with pH of 7.1, organic carbon 1.6%; total phosphorus 291 mg/Kg; available phosphorus 14.2 mg/Kg and total nitrogen 0.12%. The metal content estimated was Zn 385 mg/Kg, Cu 64 mg/Kg and Cd 2.5 mg/Kg (Table 4.1).

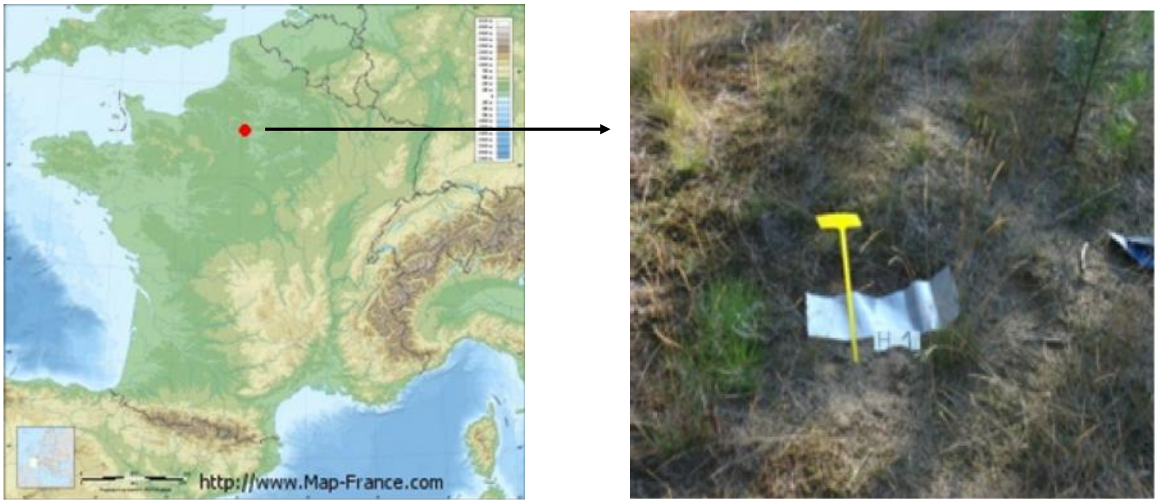


Figure 1: Soil sampling site: Agro-forestry land in Pierrelaye (PL) in northwest of Paris, France.

Table 4.1: Physico-chemical characteristics of the soil sampling site

SOIL SAMPLES	
PARAMETERS	**PL**
Coordinates	49°1’45’’N, 2°10’32’’E
Climate type	Semi-oceanic temperate
Vegetation	Poplar plantation (*Populus. sp.*)
Mean annual Temperature (°C)	10.9
Mean annual rainfall (mm)	630
Soil Type	Sandy luvisol soil
Soil pH	7.1
Parental material	Limestone
Soil temperature (°C)*	24
Soil water content (%)*	5.8
pH	7.1
Organic Carbon (%)	1.6
Available P (mg Kg^{-1})	14.2
Total P (mg Kg^{-1})	291
Total Nitrogen (%)	0.12
Available Metals (mg Kg^{-1})	
Zn	51.17
Pb	3
Cu	2
Cd	0.68
Total Metals (mg Kg^{-1})	
Zn	385
Pb	189
Cu	64
Cd	2.5

4.1.2 RNA isolation and cDNA synthesis

Total RNA was isolated from the soil sample with a yield between 330 to 980 ng per gram of soil and the quality of total RNA was determined by agarose gel electrophoresis. The two integral bands of small and large ribosomal subunits were more prominent after DNase I treatment (Figure 4.2). Extracted RNA was further analyzed by using Bioanalyzer, where presence of sharp bands corresponds to smaller (18S) and large sub units (28S) of rRNA and a broad smear (0.2 kb to 5 kb) corresponds to mRNA as observed in electrophoregram (Figure 4.3).

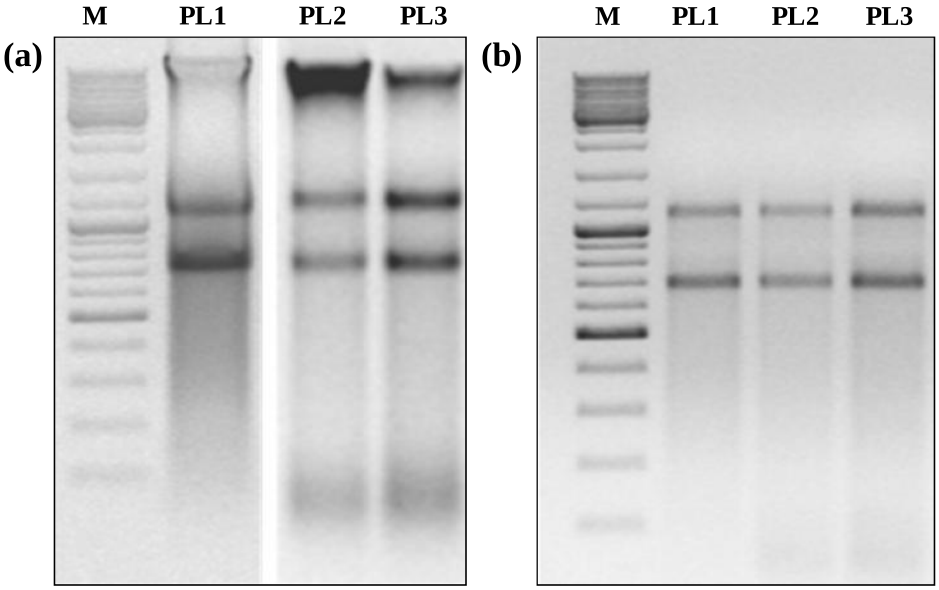
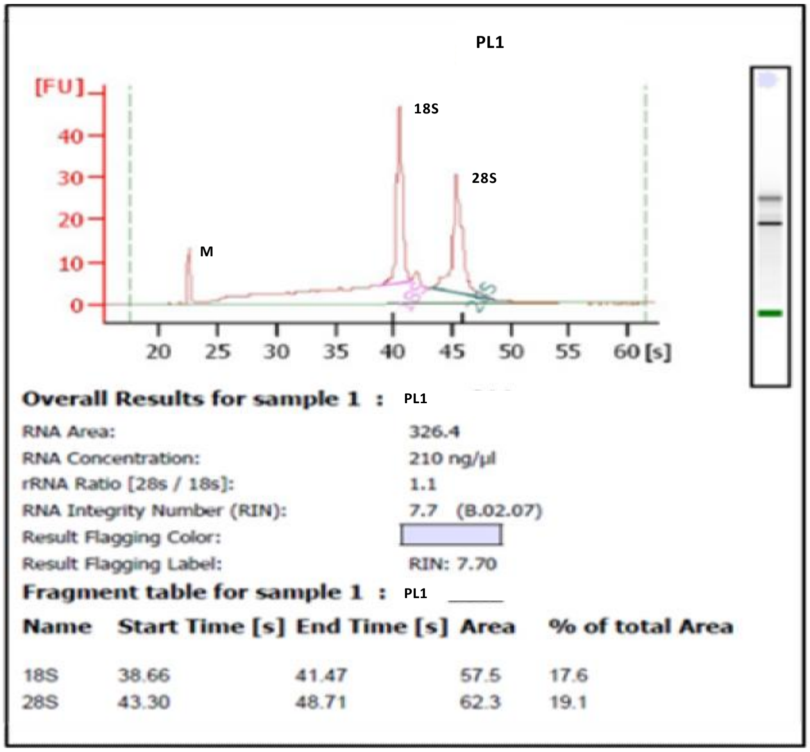


Figure 4.2: Total RNA isolation from soil sample PL. (a) Total RNA bands on agarose gel before DNase I treatment (b) Total RNA bands on agarose gel after DNase I treatment. M denotes 100 bp marker where as PL1, PL2 and PL3 are RNA sample collected from site PL.

Soil RNA is known to be comprised of approximately ninety percent of non-coding sequences especially rRNAs and seven percent of coding sequences belongs to eukaryotic realm (Ulrich *et al.* 2008). Though these figures can fluctuate with different soil samples, still bacterial biomass is always believed to be dominating in the soil. Hence, exclusion of prokaryotic nucleic acid and purification of micrograms of poly-A mRNAs is itself a daunting task due to least mRNA stability. Earlier studies have reported several methods for RNA extraction from various environmental samples, which had been extensively used to study the compositions and dynamics of microbial communities (Griffiths *et al.* 2013; Hurt *et al.* 2001; Wang *et al.* 2008). The rRNA pool of a soil sample provides only indirect view of functional status of the microbial communities flourishing in the soil environment. Hence, to monitor the functional genes or active mechanisms, phenomena or pathways in a community, we also have to study the mRNA content of the soil, which will give insight into active community of the soil sample

(a)

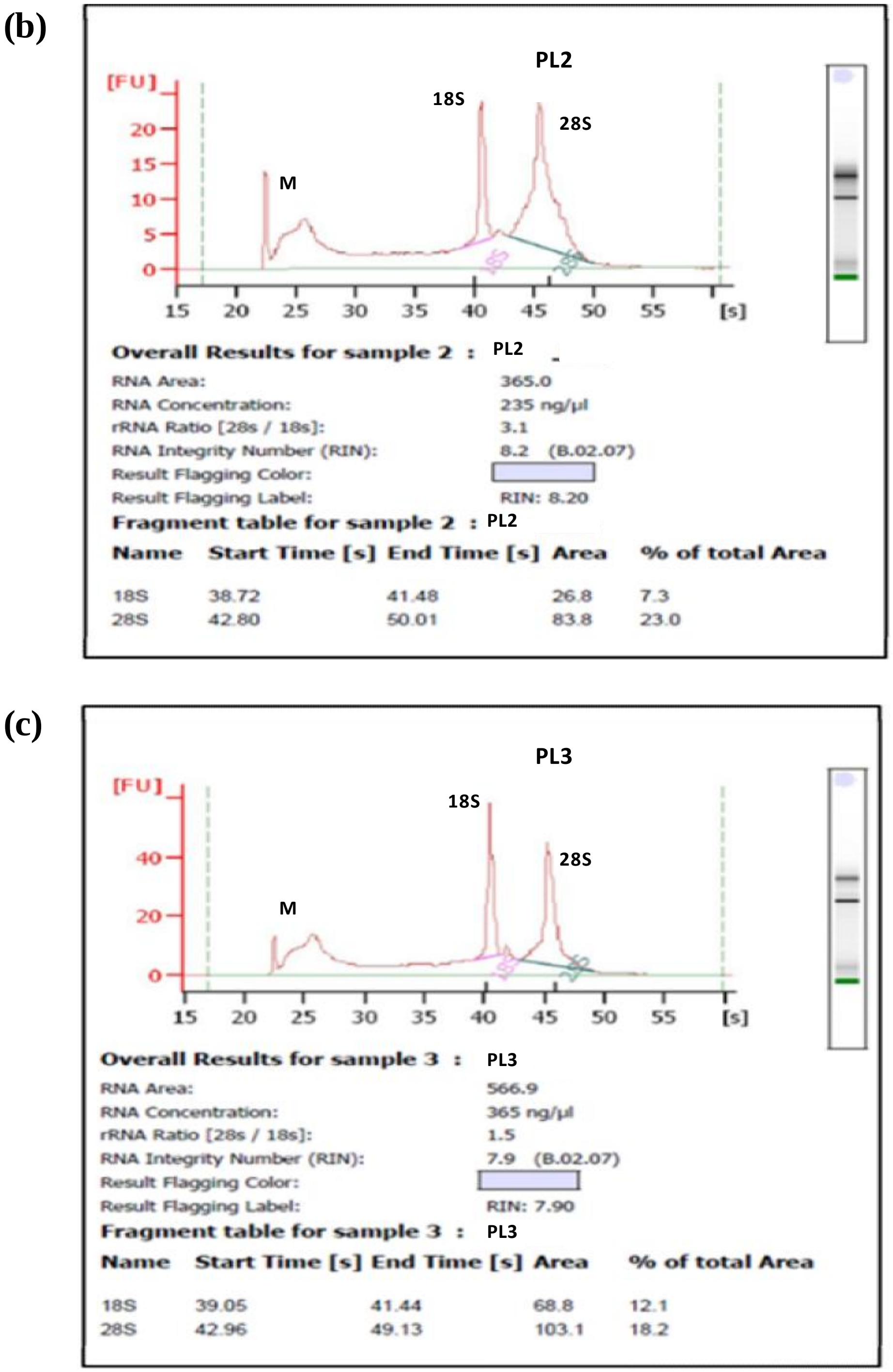


Overall Results for sample 1 : PL1

RNA Area: 326.4
RNA Concentration: 210 ng/µl
rRNA Ratio [28s / 18s]: 1.1
RNA Integrity Number (RIN): 7.7 (B.02.07)
Result Flagging Color: 🟪
Result Flagging Label: RIN: 7.70

Fragment table for sample 1 : PL1

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	38.66	41.47	57.5	17.6
28S	43.30	48.71	62.3	19.1



Overall Results for sample 2 : PL2

RNA Area: 365.0
RNA Concentration: 235 ng/µl
rRNA Ratio [28s / 18s]: 3.1
RNA Integrity Number (RIN): 8.2 (B.02.07)
Result Flagging Color: ☐
Result Flagging Label: RIN: 8.20

Fragment table for sample 2 : PL2

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	38.72	41.48	26.8	7.3
28S	42.80	50.01	83.8	23.0

Overall Results for sample 3 : PL3

RNA Area: 566.9
RNA Concentration: 365 ng/µl
rRNA Ratio [28s / 18s]: 1.5
RNA Integrity Number (RIN): 7.9 (B.02.07)
Result Flagging Color: ☐
Result Flagging Label: RIN: 7.90

Fragment table for sample 3 : PL3

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.05	41.44	68.8	12.1
28S	42.96	49.13	103.1	18.2

Figure 4.3: Electrophoregram generated for capillary electrophoresis of the total RNA extracted from the soils (a) PL1, (b) PL2, (c) PL3 through Bioanalyzer 2100 (Agilent Technologies, USA). Sharp peaks corresponding to marker (M), 18S and 28S subunits of rRNA.

Extraction of high quality mRNA is a challenging task due to short half-life of mRNA and presence of impurities such as humic acid and other organic compounds (Deutscher, 2006). Stability of mRNA molecules varies with different microbial species and depends upon nutritional status of each cell (Bernstein *et al.* 2002; Hambraeus *et al.* 2003; Redon *et al.* 2005). Genes with similar biological roles reported to display similar rates of mRNA degradation and housekeeping genes reported to have more stable messenger RNAs (Selinger *et al.* 2003; Hambraeus *et al.* 2003). In order to minimize the degradation of mRNA, hence the transcriptional profiles of sample sites during sampling, it is recommended to quick-freeze the samples in liquid nitrogen within few seconds after sampling (Carvalhais *et al.* 2003). Presence of impurities such as humic acids and other organic compounds inhibit the downstream processes such as amplification or reverse transcription of RNA for further experiments. Removal of these impurities will lessen the quantity of mRNA for further experiments. The low content of mRNA i.e., 1-5% of total RNA extract and more susceptibility of mRNA to degradation by RNases in comparison to rRNA also hinder the efficient extraction of intact mRNA from soil (Albert *et al.* 1994; Anderson *et al.* 2006). Extraction of intact mRNA from soils is also challenging due to low content of mRNA, ineffective cell lysis especially for eukaryotes and adsorption of RNA molecules on soil particles. Adsorption of mRNA molecule on soil particles is also enhanced by the conditions which are typically required for RNA isolation such as low pH or high salt conditions. Earlier, various methods has been employed as initial cell rupturing step for isolation of RNA and among all these methods, bead beating have been shown to be highly efficient method than those involving microwave-based rupturing, liquid nitrogen grinding and enzymatic lysis (Carvalhais *et al.* 2013). Further, the presence of genomic DNA as an impurity in RNA extracts results into overestimation of RNA quantity which can be overcome by treating RNA with DNaseI (Marchetti *et al.* 2012). Low content of mRNA can be enriched by various methods such as subtractive hybridization, exonuclease treatment for rRNA degradation, size separation via gel electrophoresis and duplex specific nuclease treatment (Pang *et al.* 2004; McGrath *et al.* 2008; Yi *et al.* 2011). While applying metatranscriptomic analysis for eukaryotic microbiota, eukaryotic mRNA can be isolated by using poly'dT oligonucleotides carrying magnetic beads with binding affinity towards 3' polyA tails of eukaryotic mRNA (Bailly *et al.* 2007). This enrichment technique is known to be most efficient method designed on the basis of the fact that non eukaryotic 3'-polyA RNAs are rare and degraded quickly if present (Belasco, 2010; Dreyfus and Regnier, 2002).

4.1.3. Synthesis of metatranscriptomic cDNA libraries

4.1.3.1. Synthesis of cDNA and size fractionation

Double stranded eukaryotic cDNA was synthesized by Mint-2 cDNA synthesis kit (Evrogen) using as little as 3 µg of total soil RNA. cDNA synthesis was initiated with a modified poly-dT primer and involves the "switching mechanism at 5' end of RNA template (SMART) technology" which favors the synthesis of full-length cDNAs bordered by short adaptor sequences on 3' and 5'-ends that helped in PCR amplification of metatranscriptomic cDNA. Therefore, the protocol (MINT-2 cDNA synthesis kit, Evrogen) used in this study has following characteristics: (i) Total RNA (µg) was starting material (ii) Long cDNAs were synthesized as it involves long range of PCR amplification cycles. (iii) As a consequence, pure double stranded cDNAs of eukaryotic origin was obtained from as low as 3 µg of total RNA which contains only few ng of eukaryotic mRNA.

Size fractionation of the cDNAs synthesized was performed (prior to amplification) in order to facilitate the synthesis of long strands of cDNA and diminish the preferential amplification of short ones from complex mixtures of long and short cDNAs (Wellenreuther *et al.* 2004). To achieve this, double stranded cDNAs were first exposed to three PCR cycles with primer M1 to have good amount of cDNAs without affecting their average length significantly. Amplified cDNA and marker DNA was then size fractionated by loading into two separate agarose gels of same consistency in identical trays. Segregated cDNA gel and marker DNA gel were superimposed on each other and cDNA gel bands of requisite size was excised. Bi-dimensional agarose gel electrophoresis facilitated the complete segregation of these pre-amplified cDNAs and minimized the probability of cross contamination of different cDNA fractions (Figure 4.4a). Size fractionation resulted into cDNA fractions of three different sizes i.e., fraction A: < 0.5 kb; fraction B: 0.5–1 kb and fraction C: >1 kb (Figure 4.4b). These fractions were further amplified up to 22 to 30 PCR cycles with primer M1 which gave three different sized discrete and distinct DNA smears while non-size fractionated amplified cDNAs gave a broad smear throughout the gel lane. Further, these fractions were tested for cross contamination by amplification with the primers derived from conserved genes of corresponding sizes: for fraction C: β-Tub and EF1-α having coding sequences of approximately 1300 and 1400 bp, for fraction B: 40S ribosomal protein S3 and peptide-methionine sulphoxide reductase (MsrA) with coding sequence of about 800 and 600 bp where as fraction A was too small to have a house keeping gene for validation. In case of

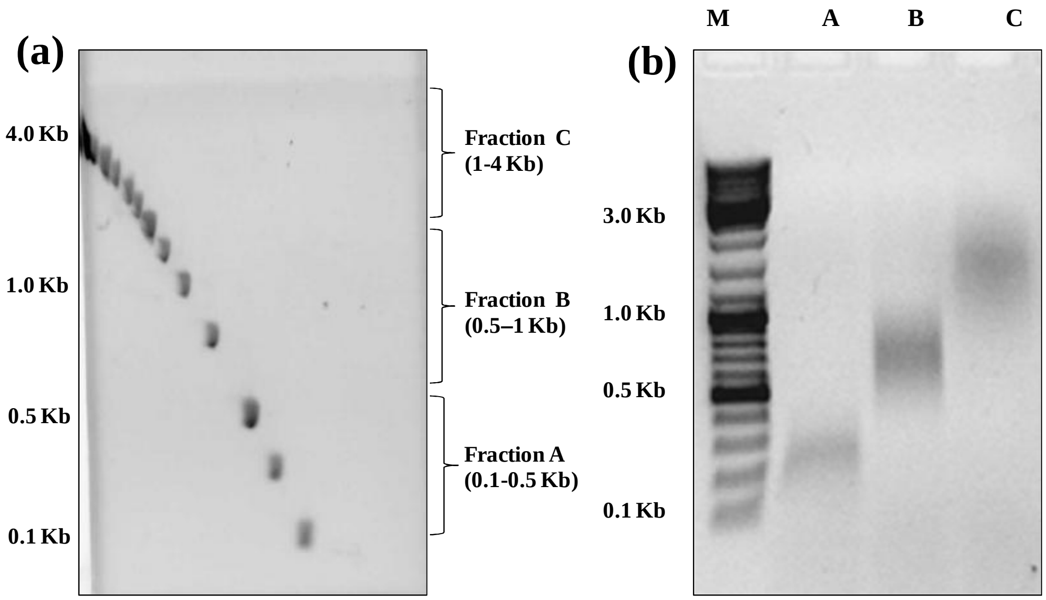


Figure 4.4: Size fractionation of cDNA synthesized from total RNA. (a). Unstained gel carrying cDNA was superimposed over the EtBr-stained DNA marker gel to visualize the fractionation (b) Agarose gel showing three fractions of cDNA i.e., fraction A, fraction B and fraction C extracted from excised cDNA gel slices.

largest fraction C (1-4 kb), positive amplification was observed for β-Tub and EF1-α fragments only while for fraction B (0.5-1 kb), positive amplification was observed for RibS3 and MsrA fragments only and no amplified product was visible on agarose gel for fraction A (Figure 4.5). Gel electrophoresis had shown complete correlation between size-fractionated cDNAs and detection of PCR product from housekeeping genes whose expected size was included in the subsequent fraction size range. These results demonstrated the efficiency of this method in procuring long and high quality eukaryotic cDNAs from few µg of total RNAs.

Synthesis of cDNA using SMART technique is a convenient method and does not require any mRNA enrichment or manipulative procedure before proceeding the reaction. This protocol is different from conventional first strand synthesis in terms of use of SMART

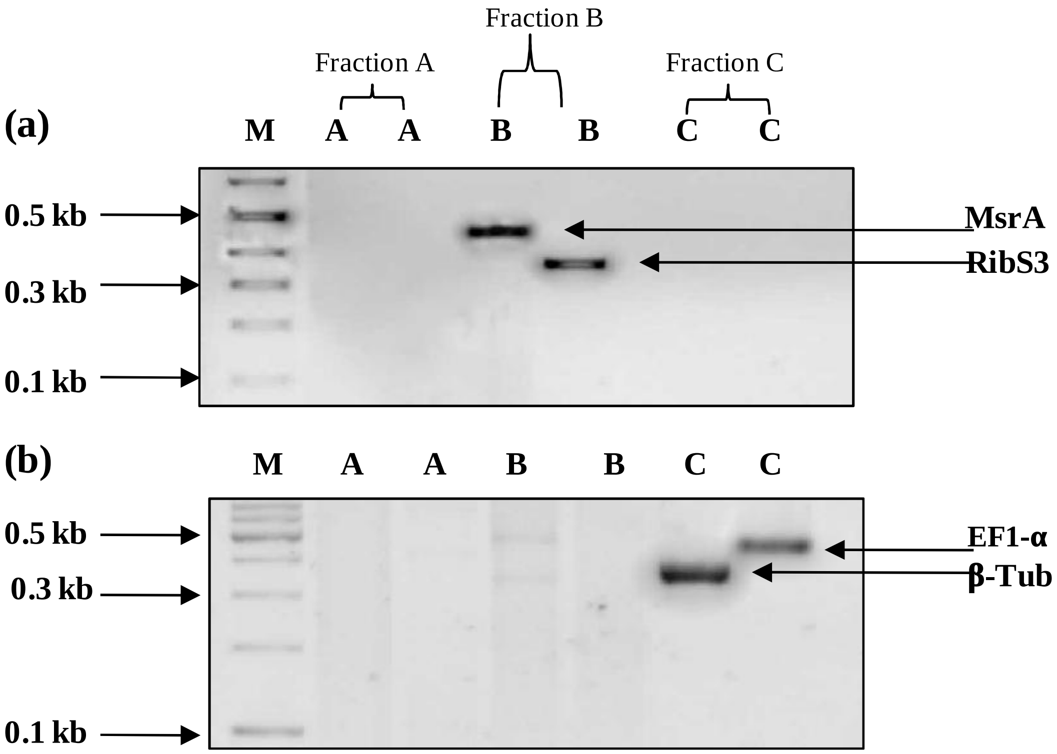


Figure 4.5: Validation of cDNA fractions with conserved gene family primers. PCR products were loaded on the agarose gel and amplified PCR product was observed only for (a) fraction B (MsrA and RibS3) and (b) fraction C (β-Tub and EF1-α).

RNA oligo (3'-end CDS adaptor and 5'-end plugOligo-3M adaptor) and MMLV RTase enzyme in the reaction. MMLV RTase is an RNaseH negative enzyme which ensures the addition of C residues and prevent the degradation of SMART RNA oligo while base pairing with these residues (Wellenreuther *et al.* 2004). In this study, size fractionation of cDNA synthesized was performed prior to PCR amplification step unlike conventional methods, where cDNA synthesis was directly followed by PCR amplification step. Size fractionation resulted into different sized cDNA fractions, which were further amplified in separate reactions. As large cDNAs are less frequent in a cDNA consortium, hence higher number of PCR cycles would be required in comparison to smaller ones, in order to obtain equivalent amount of amplified products for all the fractions. This would further increase the redundancy and PCR based errors. Hence, to circumvent these obstacles, as fewer PCR cycles as possible were suggested to be performed in previous studies (Levesque *et al.* 2003; Wellenreuther *et al.* 2004). These advancements in cDNA synthesis protocol removes the PCR biases generally observed in conventional methods, where smaller cDNAs were preferentially amplified over long cDNAs (Yadav *et al.* 2014).

4.1.3.2. cDNA library synthesis

To synthesize cDNA libraries, all the three cDNA fractions (500 ng each) and modified pFL61 yeast plasmids were subjected to restriction digestion by SfiI enzyme which recognizes SfiIA and SfiIB sites located in the PlugOligo-3M and CDS-4M adapter sequences respectively. Digested products were further purified by phenol-chloroform extraction and after precipitation, exposed to directional cloning downstream the *S. cerevisiae* PGK promoter in a modified pFL61 yeast expression vector. Ligated products were then electro-transformed into MegaX DH10B™ T1R Electrocomp™ cells, (Invitrogen, USA). Each of the three libraries contained at least 10^6 clones growing on LB agar media supplemented with ampicillin and these clones were pooled to constitute each library. Recombinant plasmid pools from each library were extracted by using ZymoPure^{TM} Plasmid Maxiprep Kit (Zymo Research, USA) and subjected to restriction digestion with SfiI enzyme. After gel electrophoresis, cDNA insert pool of each library was detected as a smear corresponding to their original size range (Figure 4.6).

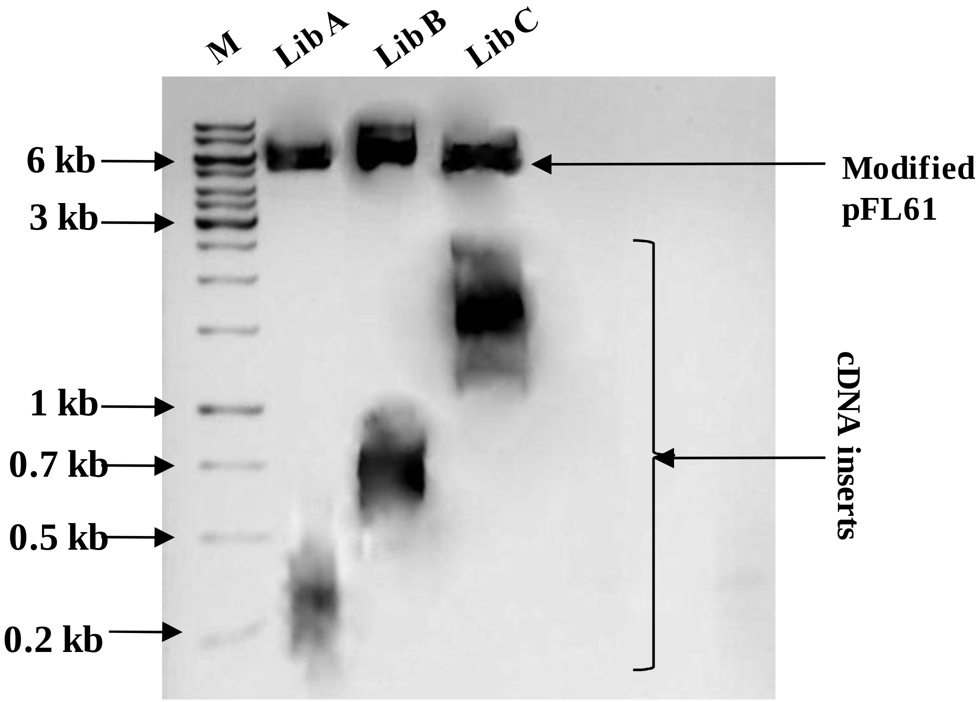
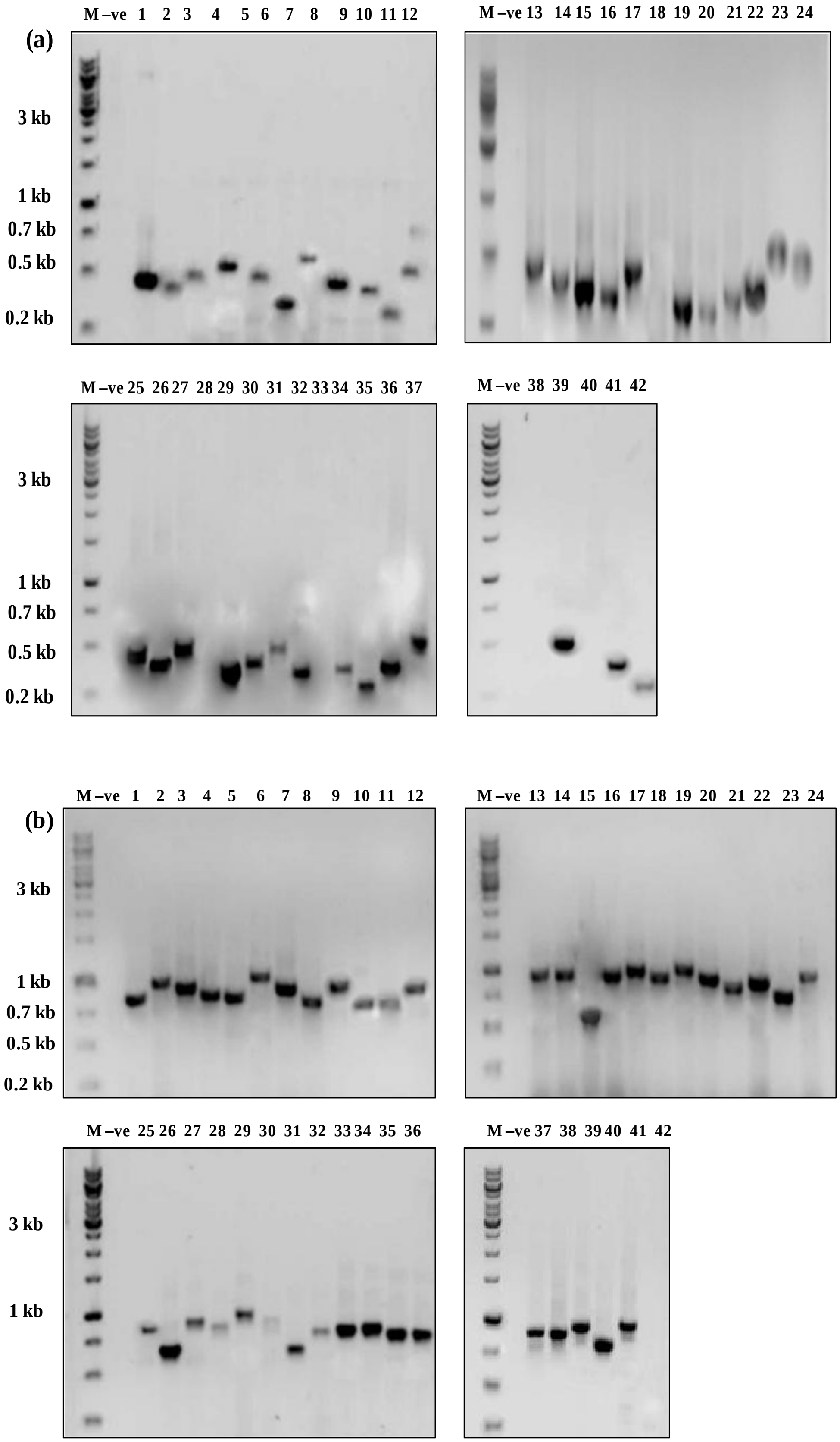


Figure 4.6: Agarose gel electrophoresis for SfiI digested cDNA libraries. First well is loaded with DNA marker 1 Kb, subsequent well with Sfi I digested Lib A, Lib B and Lib C. –4 kb) and (C) SfiI digested libraries PL-A, PL-B and PL-C. For each sub library, the size range of the cDNA inserts is similar as after size fractionation.

The cDNA inserts of 42 random colonies from each library was further amplified to crosscheck the presence of inserts. Absence of inserts was found merely 1% and as expected, the size of amplified products from each clone confined within the expected size cut off of corresponding library i.e., between 0.1 - 0.5 kb for Lib A; 0.5-1 kb for Lib B and 1-3 kb for Lib C (Figure 4.7).

Full length cDNAs derived from eukaryotic mRNAs are crucial tool to identify full length novel genes and splice variants and as a resource for full length cDNA carrying clones for studying the metagenomics and metatranscriptomics (Wiemann *et al.* 2003).


(a)
M –ve 1 2 3 4 5 6 7 8 9 10 11 12
M –ve 13 14 15 16 17 18 19 20 21 22 23 24
3 kb
1 kb
0.7 kb
0.5 kb
0.2 kb
M –ve 25 26 27 28 29 30 31 32 33 34 35 36 37
M –ve 38 39 40 41 42
3 kb
1 kb
0.7 kb
0.5 kb
0.2 kb
(b)
M –ve 1 2 3 4 5 6 7 8 9 10 11 12
M –ve 13 14 15 16 17 18 19 20 21 22 23 24
3 kb
1 kb
0.7 kb
0.5 kb
0.2 kb
M –ve 25 26 27 28 29 30 31 32 33 34 35 36
M –ve 37 38 39 40 41 42
3 kb
1 kb

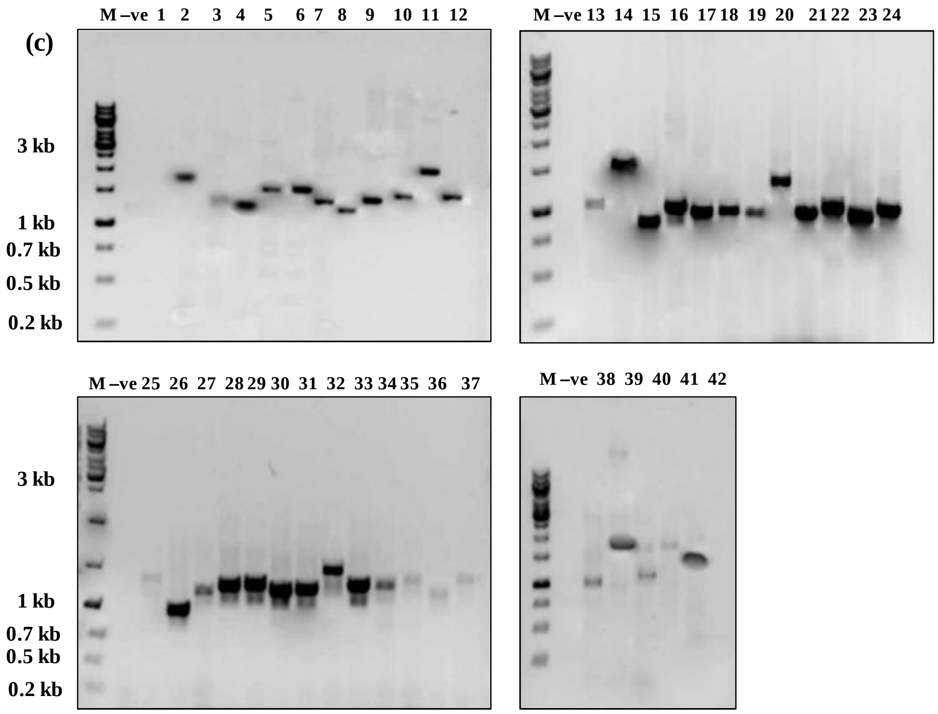


Figure 4.7: Validation of insert size of cDNA libraries by colony PCR. Panels a, b and c are the gel images of separated PCR products of 42 clones from each of the three libraries i.e., Lib A (0.1 - 0.5 kb); Lib B (0.5-1 kb) and Lib C (1-3 kb) respectively.

Unfortunately, traditional methods used to synthesize cDNA libraries contains large proportion of 5’ truncated clones due to incomplete reverse transcription of the template mRNA especially when large mRNAs or mRNAs with tendency to form secondary structure are involved (Gubler and Hoffman, 1983). For such reasons, full length cDNAs are always underrepresented in cDNA libraries synthesized through conventional methods. Several techniques have been developed to construct a library with full length cDNAs such as RNA oligo ligation to 5’ end of mRNA (Kato *et al.* 1994; Suzuki *et al.* 1997), 5’ affinity selection via eukaryotic initiation factor 4E (Edery *et al.* 1995) or 5’ cap biotinylation followed by biotin affinity selection (Carninci and Hayashizaki, 1999; Carninci *et al.* 1996). All of these techniques are quite laborious and increases the chances of mRNA degradation as they involves several enzymatic steps performed on mRNAs, hence these methods require quite high amount of starting mRNA i.e., 5-100 µg.

In contrasts to all these demerits, “template switching protocol” i.e., SMART technique used in this work for synthesis of cDNA libraries is very uncomplicated as well as vigorous and need only 0.025 to 1 µg of starting mRNA (Zhu *et al.* 2001). Another merit of this protocol is that there is no negative selection against the truncated mRNA species as the basic principle involves selection of full reverse transcribed mRNA molecules instead of selection via mRNA 7’ cap. This method exploits the specific characteristic of MMLV reverse transcriptase which, adds C residues specifically at 3’ end of the first strand cDNA on reaching the end of mRNA template, not on the prematurely terminated reverse transcripts. Further, these C residues form base pairing with three G present in an RNA oligo in the reaction and served as extended template for reverse transcription. This lead to synthesis of 5’ tagged full length cDNAs as G residues would be added only on full length reverse transcripts (not to the premature transcripts) and further amplified by RNA oligo-specific primer provided. cDNA library synthesized by SMART technique gives a good percentage of full length clones in comparison to those conventional methods or transcripts synthesized with other full length enriching techniques (Sugahara *et al.* 2001, Zhu *et al.* 2001). In conventional library construction, large insert carrying clones are rarely observed often due to truncation of longer transcripts during cDNA synthesis and strong size based bias against long cDNAs during cloning. While, in this study, cDNA size fractions after PCR amplification were subjected to restriction digestion and ligated into modified yeast shuttle vector pFL61 in separate reactions to obtain size fractionated sub-libraries.

Hence, in this work, we have modified the SMART technique to produce cDNA libraries with large inserts in convenient plasmid vector pFL61 and reported the synthesis of three different sized libraries (Lib A 0.1-0.5 kb; Lib B 0.5-1.0 kb; Lib C 1- 3 kb) enriched for full length clones. This technique is easy to use with no additional enzymatic steps are involved and size fractionation protocol we followed was the very efficient and can be performed with basic laboratory tools.

4.2. Heterologous expression of metal tolerant genes by yeast functional complementation

All three fractions of cDNA libraries (Lib A, Lib B and Lib C) were further screened for presence of Cd tolerant functional genes by transforming the recombinant plasmid libraries into hypersensitive uracil auxotrophic yeast mutant $ycf1^{\Delta}$. Transformation was done by lithium acetate method (Gietz and Schiestl, 2007) and the transformed cells were selected by plating onto a minimal medium without uracil and supplemented with 40 μM of $CdSO_4$ (Figure 4.8). It is a primary mode of screening where a complete library screening was performed for Lib C while only a small volume was screened for Lib B and Lib A to get Cd tolerant clones. After screening 5.2×10^6 clones of Lib C (5 times of the original size of the library), 498 clones showed tolerance to Cd. A total of 1,11,949 clones were screened for Lib B and 78 clones showed tolerance towards Cd and for Lib A, 80 clones were found tolerance to Cd among 1,74,020 clones screened.

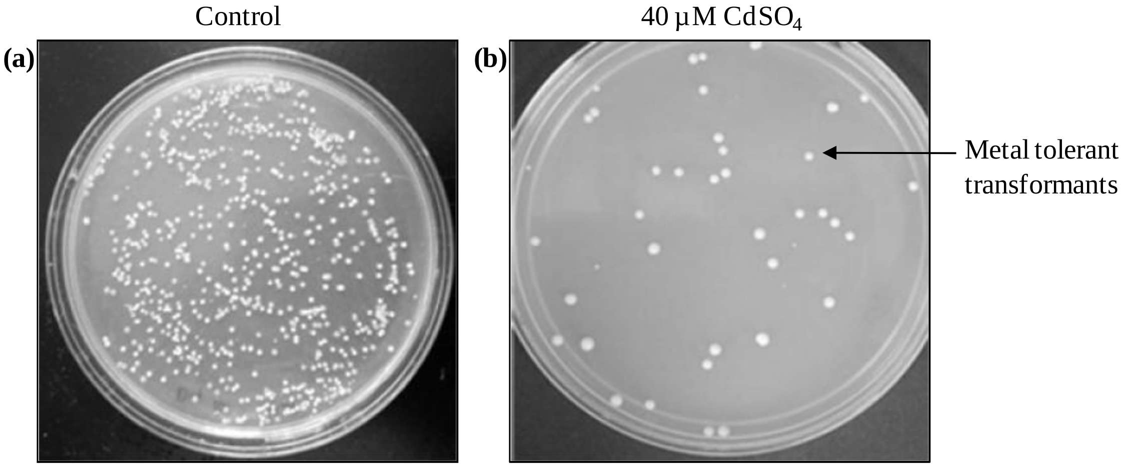


Figure 4.8: Primary screening of Cd tolerant cDNAs by transforming cDNA library into Cd sensitive yeast mutant *ycf1*$^{\Delta}$. Yeast transformants *ycf1*$^{\Delta}$ carrying cDNAs were spread on (a) SD-Ura (b) SD-Ura + $CdSO_4$. SD-Ura plates without metal considered as control.

As this work focus on screening of functional genes of eukaryotic origin from a soil sample thus relies on transcriptionally expressed genes present in soil metatranscriptomic library. Here, functional approach of metatranscriptomics was used to detect novel metal tolerant genes, which can be screened on a selective media supplemented with metal. Various studies have also been reported where function based approach of metatranscriptomics was exploited to screen the functional genes with ecologically significance such as organic matter degrading enzymes like lignocellulolytic, chitinolytic enzymes (Courty *et al.* 2008; Kellner and Vandenbol, 2010) and metal tolerant proteins (Ziller *et al.* 2017). However, in our work, it is known that many of the transformants obtained after this primary mode of screening are likely to be false positive ones in the sense that they might not carry a transforming gene, which confers tolerance to the studied metal. Therefore, these false positive clones must be identified by performing "laborious" additional secondary, tertiary and eventually quaternary screens. Hence, after primary screening for each of the library on Cd amended media, numerous clones were further found to be Cd tolerant which is further exposed to secondary screening i.e., drop assay.

4.2.1. Drop Assay

Drop assay is secondary mode of screening of clones carrying cDNA for confirming their Cd tolerant phenotype and to screen out the false positives. Drop test illustrate the capacity of yeast transformant to tolerate a particular metal varies with increase or decrease in cell density. This secondary mode of screening consists of spotting of the different dilutions of transformant yeast culture onto a fresh selective medium without uracil and supplemented with a lethal dose of Cd metal (Figure 4.9, 4.10 & 4.11). Drop tests reduced the total number of Cd tolerant yeast transformants from all the three libraries to 580 and used to classify them as "highly tolerant” (growth up to 0.0001 dilutions), “moderately tolerant” (growth up to 0.001 dilutions) and "low tolerant" (growth up to 0.01 dilutions) transformants on the basis of metal tolerant phenotype possessed by transformants. A total of 498 clones were screened as Cd tolerant clones from the largest sized library i.e., C during primary screening out of which 83 were observed as highly tolerant clones, 126 as moderately tolerant while 174 clones as low tolerant as observed in secondary screening i.e., drop assay. Similarly, for Lib B, a total of 78 clones conferred tolerance to Cd metal in which 6 were highly tolerant, 5 was moderately tolerant while 28 were low tolerant. For Lib A, the smallest sized library in this study, 20 transformants were observed to be highly tolerant, 7 as moderately tolerant and 33 as low tolerant.

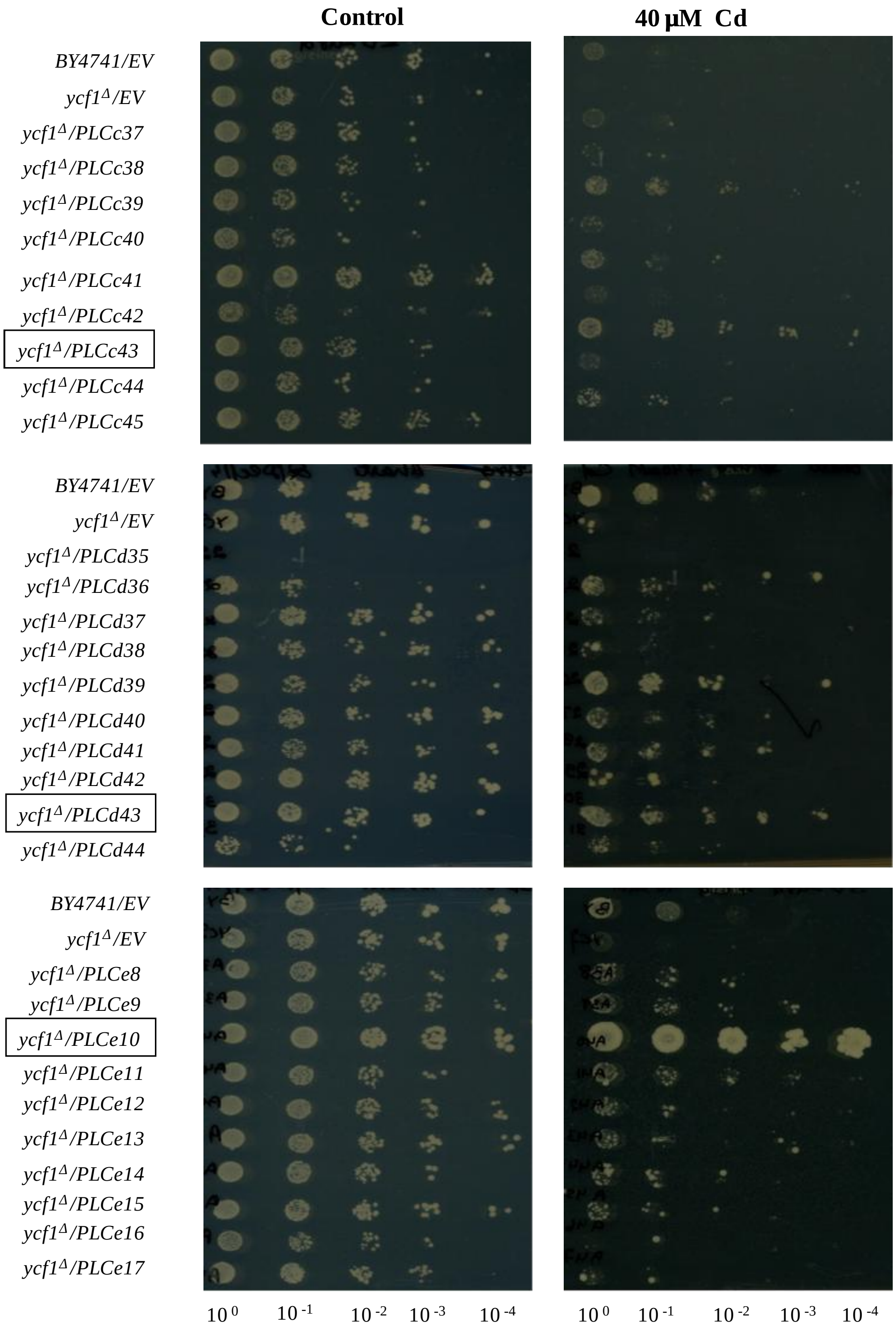


Figure 4.9: Secondary mode of screening of transformants carrying cDNA corresponding to Lib C for their metal tolerant phenotype conferred to mutant *ycf1*$^{\Delta}$ at Cd 40 µM. Wild strain BY4741 and *ycf1*$^{\Delta}$ were used as control and EV denotes to empty vector pFL61.

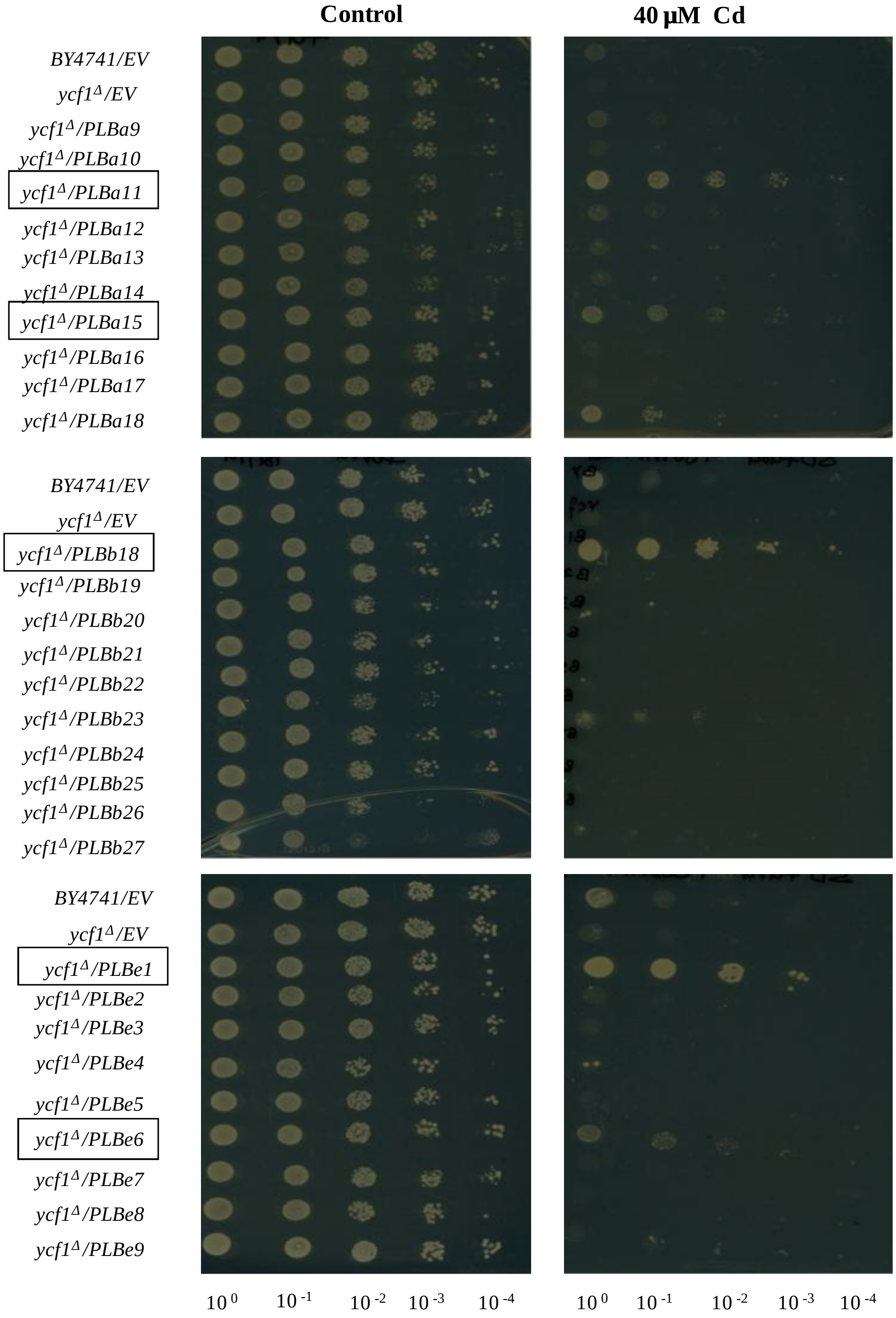


Figure 4.10: Secondary mode of screening of transformants carrying cDNA corresponding to Lib B for their metal tolerant phenotype conferred to mutant $ycf1^{\Delta}$ at Cd 40 µM. Wild strain BY4741 and $ycf1^{\Delta}$ were used as control and EV denotes to empty vector pFL61.

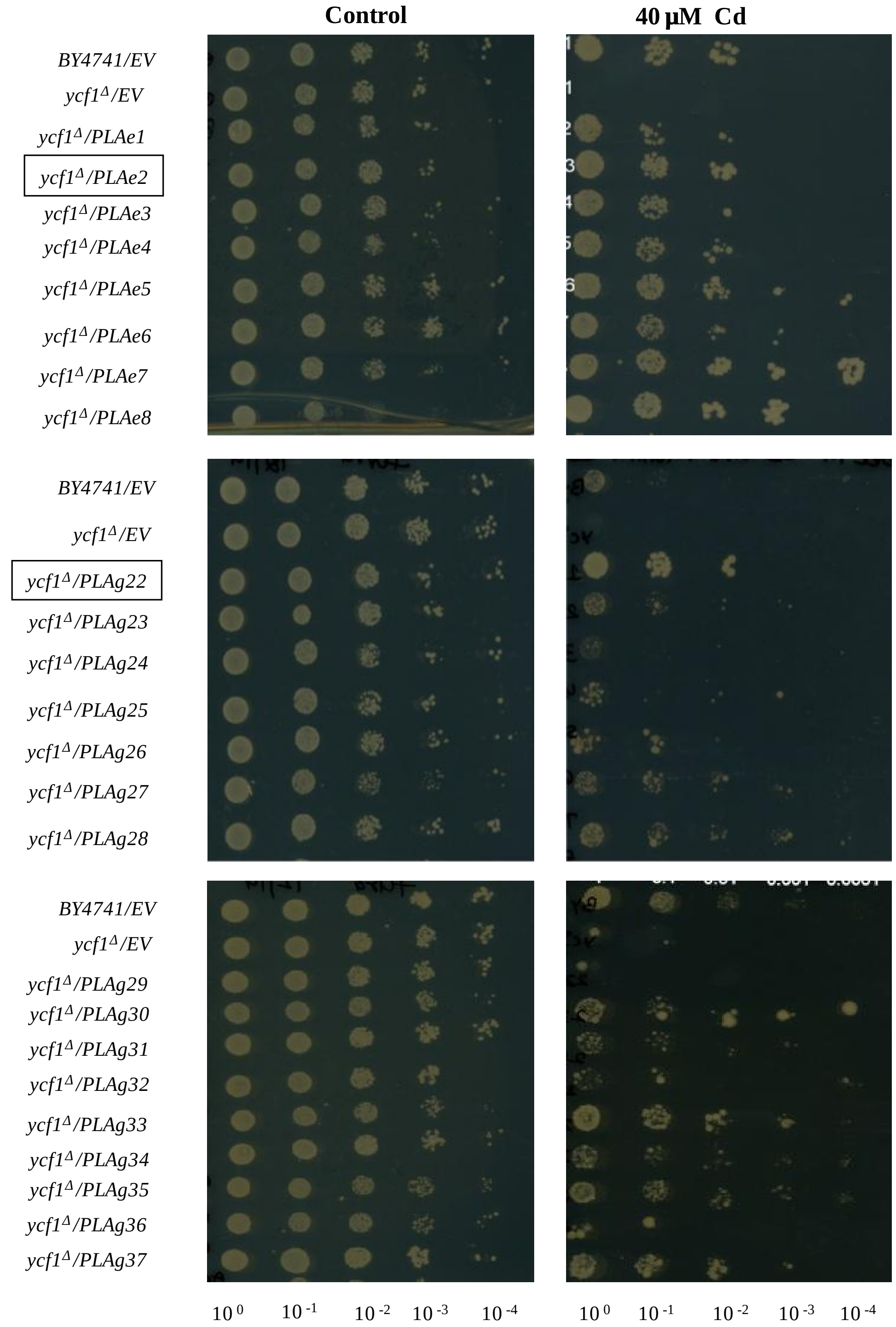


Figure 4.11: Secondary mode of screening of transformants carrying cDNA corresponding to Lib A for their metal tolerant phenotype conferred to mutant *ycf1*$^{\Delta}$ at Cd 40 µM. Wild strain BY4741 and *ycf1*$^{\Delta}$ were used as control and EV denotes to empty vector pFL61.

4.2.2. 5-Flouroorotic acid Test

5-Flouroorotic acid (5-FOA) test is a tertiary mode of screening, performed on the transformants capable of growing on Cd amended media plates during secondary screening. The transformants screened with the capability to rescue mutants from Cd toxicity were further crosschecked by streaking them on media amended with lethal dose of 5-FOA. In *S. cerevisiae*, 5-fluoroorotic acid (5-FOA) has been used in various genetic screening experiments to select and distinguish *URA3*$^+$ cells from *URA3*$^-$ cells (Boeke *et al.* 1984; Ko *et al.* 2008). 5-fluoroorotic acid is a uracil analogue and presence of which can be toxic to the cells. *URA3* gene encodes an important enzyme, orotidine-5-phosphate decarboxylase, which can convert 5-FOA into fluorouracil (5-FU) in uracil biosynthetic pathway (Jones and Fink, 1982). 5-FU is a pyrimidine analogue that interferes with function and processing of RNA and DNA molecules when incorporated in place of uracil or thymine into RNA and DNA, hence leading to cell death. Thus, *ura3* mutants required uracil from external source for growth and viability in 5-FOA medium. Presence of *URA3*$^+$ plasmid inside *ura3* mutants makes them sensitive towards 5-FOA yet they can survive in a 5-FOA medium by losing their *URA3*$^+$ plasmid and by supplementing the medium with uracil. Thus, in our study, the clones growing on 5-FOA+ Ura amended media, were assumed to have lost the plasmid carrying metatranscriptomic cDNA. True transformants, after losing their plasmid, have also lost their metal tolerant gene in parallel to their uracil prototrophy. Henceforth, to ensure that the metal tolerant phenotype of yeast cells was due to recombinant plasmid, not due to any mutation in yeast genome, drop assay on SD + Ura plate with 60 µM $CdSO_4$ was performed (Figure 4.12, 4.13 and 4.14) for those which enable to grow in presence of 5-FOA by flushing out

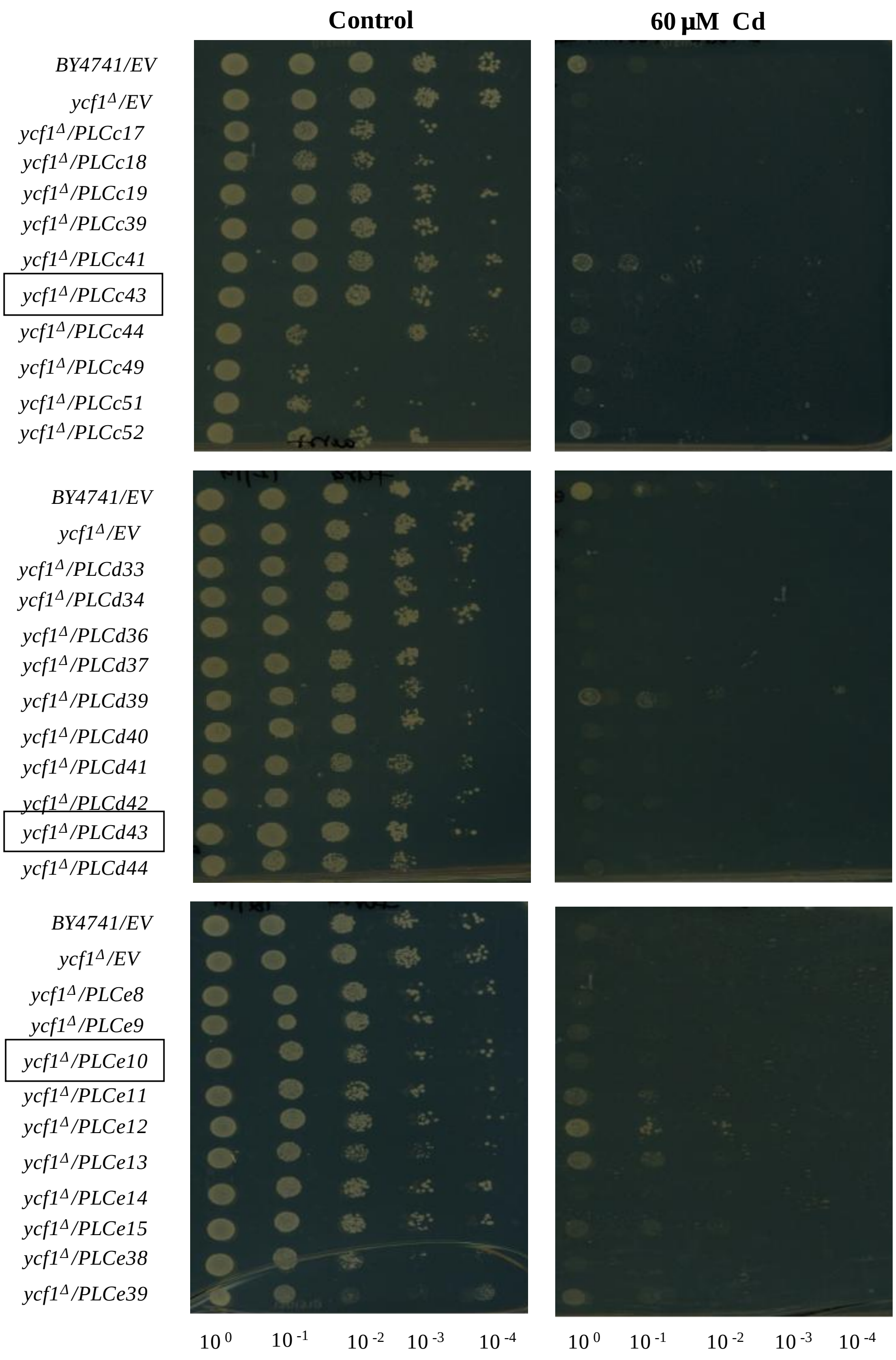


Figure 4.12: Drop test of yeast clones of Lib C recovered after 5-FOA screening at higher Cd concentration i.e., 60 µM $CdSO_4$ to confirm the role of metatranscriptomic cDNAs in conferring tolerance to the hypersensitive yeast mutant *ycf1*$^{\Delta}$. Wild strain BY4741 and *ycf1*$^{\Delta}$ were used as control and EV denotes to empty vector pFL61.

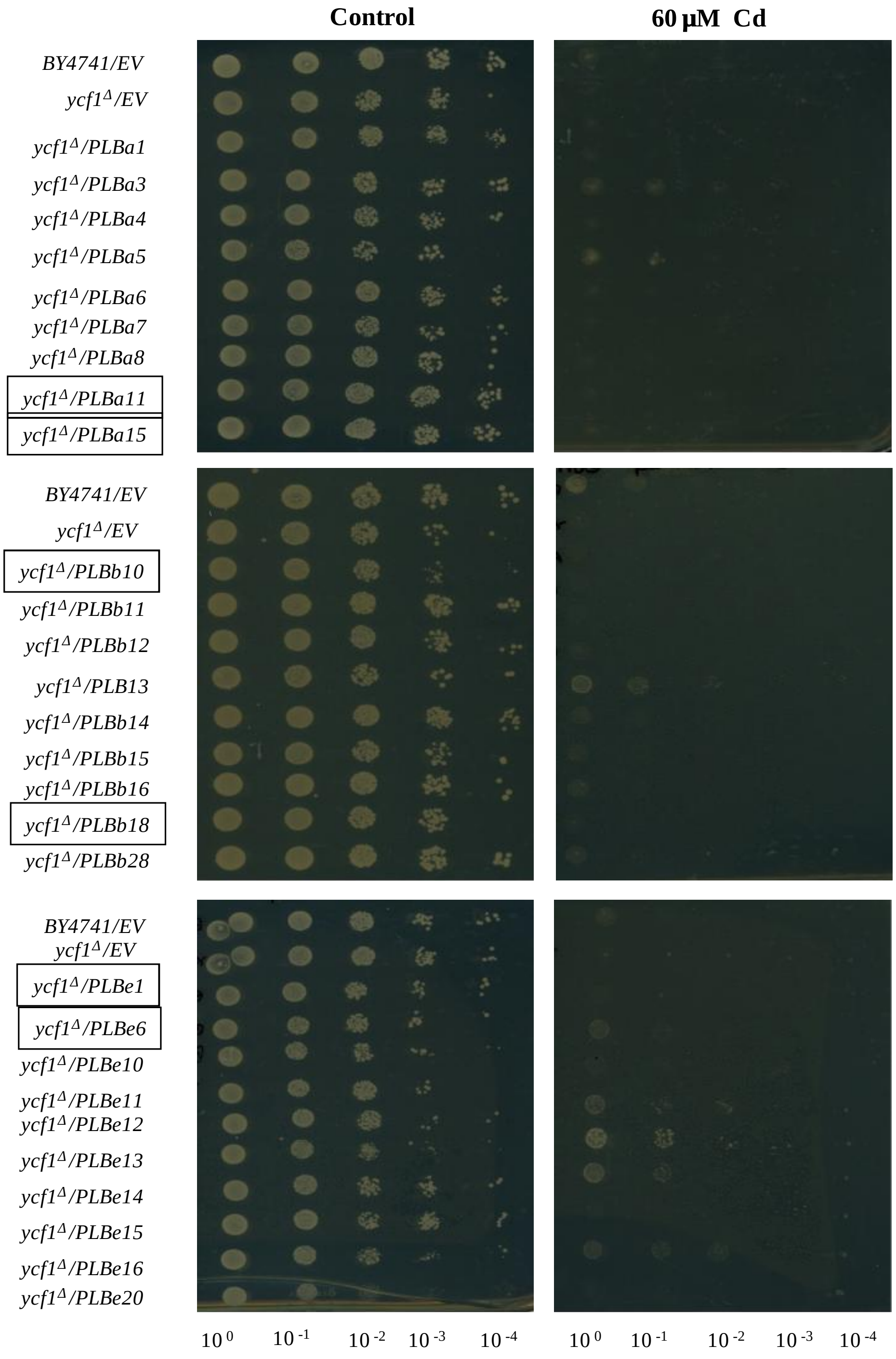


Figure 4.13: Drop test of yeast clones of Lib B recovered after 5-FOA screening at higher Cd concentration i.e., 60 µM $CdSO_4$ to confirm the role of metatranscriptomic cDNAs in conferring tolerance to the hypersensitive yeast mutant *ycf1*$^{\Delta}$. Wild strain BY4741 and *ycf1*$^{\Delta}$ were used as control and EV denotes to empty vector pFL61.

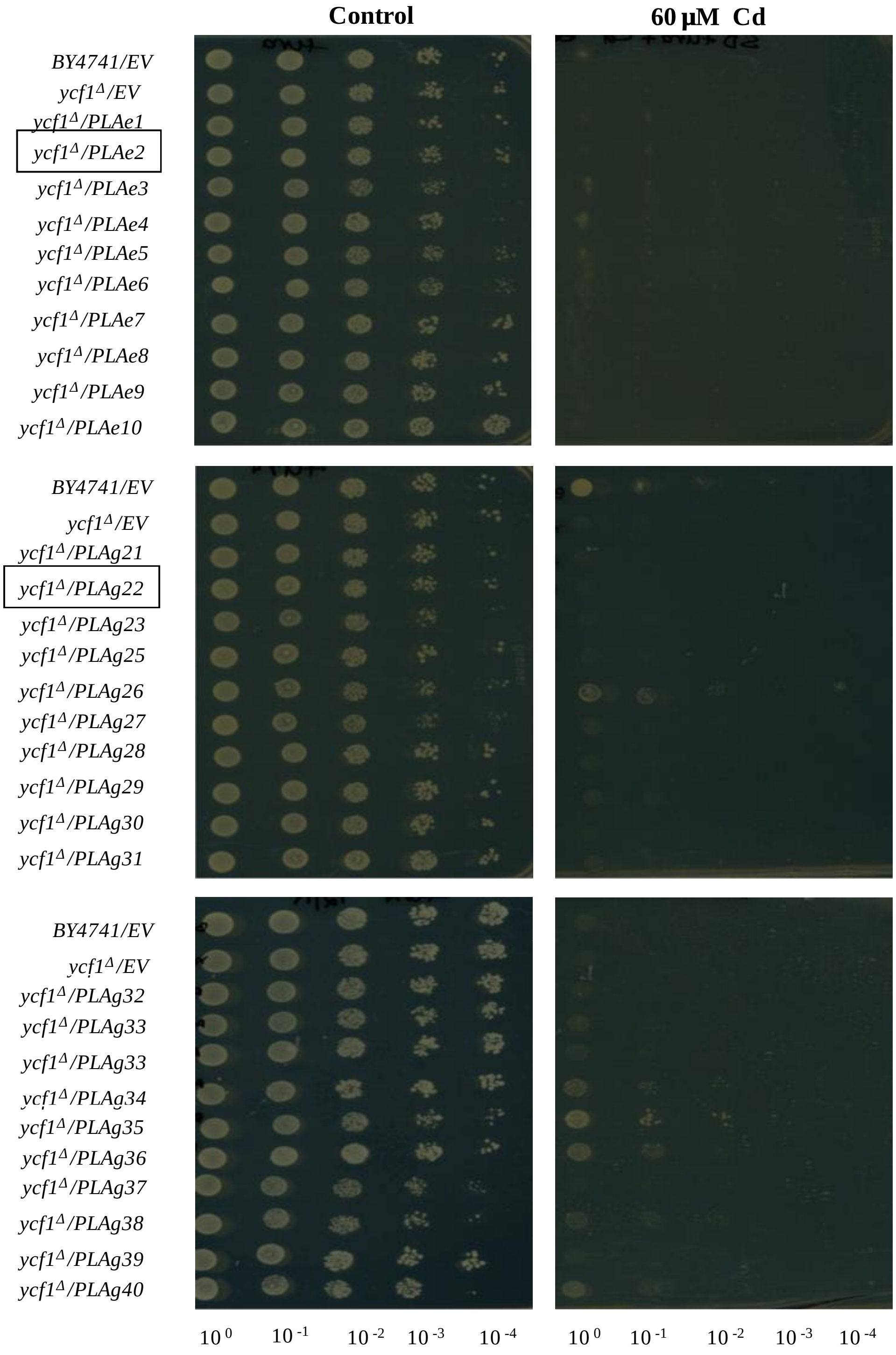


Figure 4.14: Drop test of yeast clones of Lib A recovered after 5-FOA screening at higher Cd concentration i.e., 60 µM $CdSO_4$ to confirm the role of metatranscriptomic cDNAs in conferring tolerance to the hypersensitive yeast mutant $ycf1^{\Delta}$. Wild strain BY4741 and $ycf1^{\Delta}$ were used as control and EV denotes to empty vector pFL61.

the *URA3*$^+$ plasmid. A total of 134 yeast transformants without plasmid were observed to be sensitive for 60 µM of Cd toxicity thus the number of true yeast transformants possessing Cd tolerant cDNAs were constricted to 98 for Lib C, 12 for Lib B and 24 for Lib A. Hence, these findings advocate the FOA test and drop test after FOA as a reliable protocol to confirm the role of cDNA in conferring tolerance towards Cd toxicity and helped in filtering out all the false positives (Table 4.2, 4.3 & 4.4).

4.2.3. Yeast Colony PCR

Colony PCR is a quaternary mode of screening required to distinguish the yeast transformants carrying the correct cDNA construct responsible for their metal tolerant phenotype from the false positives. This technique take benefit of the high sensitivity of PCR where amplification can be done from crude preparation of cells and small amount of template DNA is required to get a visualized band easily on an agarose gel. To ensure the presence of metatranscriptomic cDNA inserts inside the screened yeast transformants, yeast cells were exposed to 20 mM NaOH at high temperature i.e., above 95ºC for cell lysis for 5 min and the supernatant was then amplified with vector specific primers NF and NR. Presence of cDNA was marked by the sharp band of amplified product of requisite size in the agarose gel when observed in UV illuminator. For Lib C, bands between 1-3 Kb were observed for all the true transformants while for Lib B, the bands appear at 0.5 -1 Kb and for Lib C, it is between 0.1- 0. 5 Kb (Figure 4.15). Only bands from 14 transformants out of total 126 transformants screened from the PL library were not observed on agarose gel while remaining true transformants were moved to further level of study.

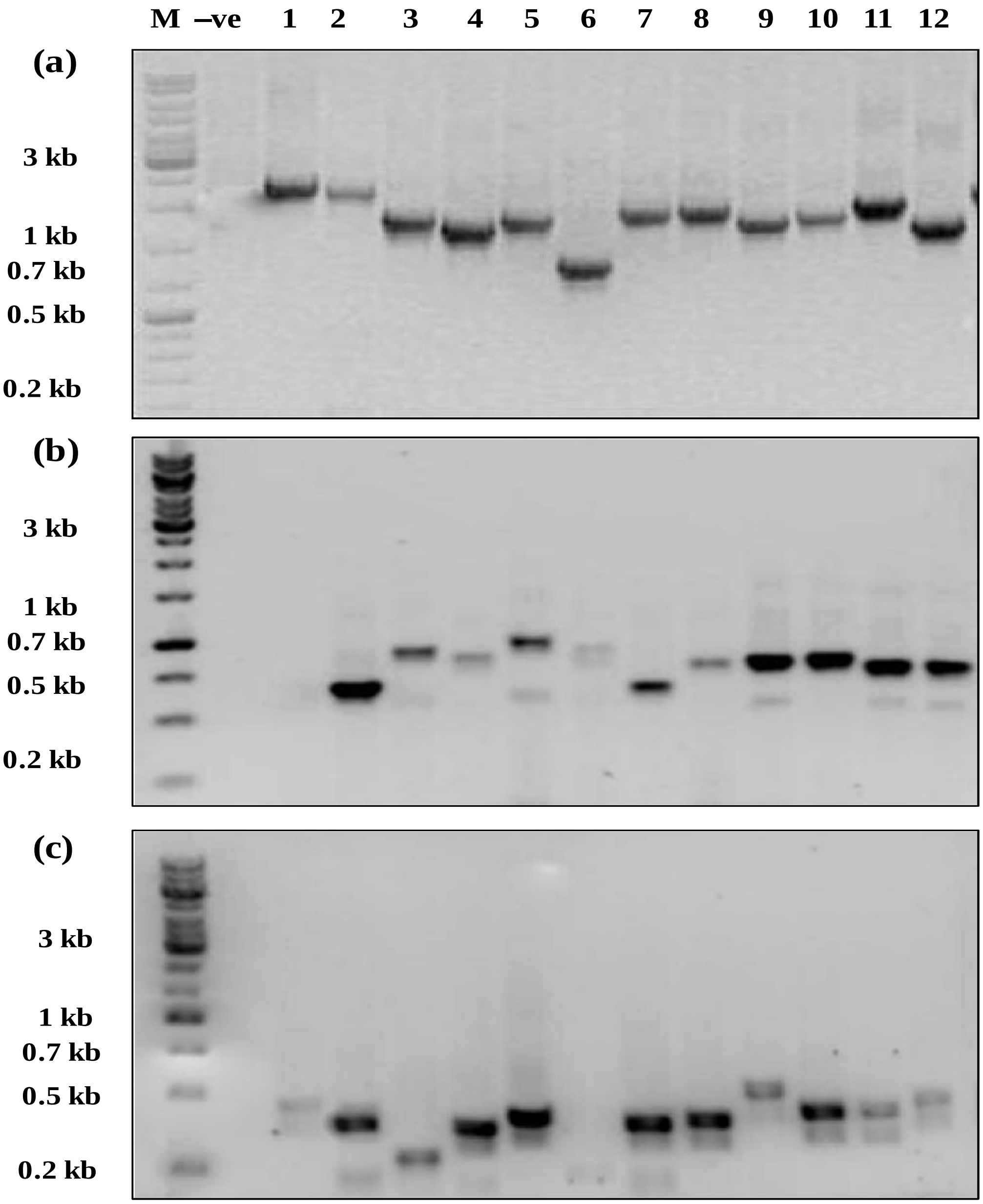
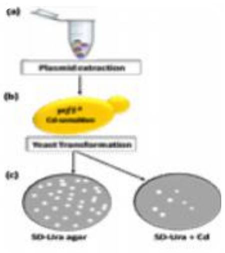


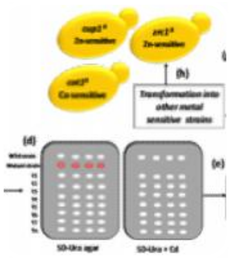
Figure 4.15: Quaternary screening of Cd tolerant yeast transformants. Agarose gel electrophoresis marked the presence of metatranscriptomic cDNA of requisite size inside yeast transformants *ycf1*$^{\Delta}$ after colony PCR. Panel A, B and C showing colony PCR products of transformants from three different sized cDNA libraries i.e., C (1-3 Kb), B (0.5-1 Kb) and A (0.1-0.5 Kb). M denotes 1 Kb marker DNA where as various yeast colony PCR products were denoted by numbers

Thus, colony PCR is a powerful tool which plays a significant role in screening process where hundreds of transformants are involved. In other terms, it is an advanced alternative to the traditional strategy of growing cultures from several colonies to get few micrograms of DNA from each culture for verifying the presence of desired construct by restriction digestion. Therefore, all these modes of screening shown in a flowchart below i.e., drop assay, 5-FOA test and colony PCR for yeast transformants had filtered out all the false positives and gave a significant number of true transformants (Table 4.5) which were explored further to know the taxonomic origin of cDNAs through bioinformatic analysis.



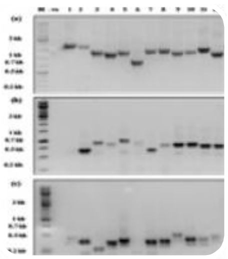
Primary screening of cDNA libraries A, B and C for Cd tolerance genes

- Transformation of Lib A, B and C into Cd sensitive uracil auxotrophic yeast mutant *ycf1*$^{\Delta}$ by Li-Ac method (Gietz and Schiestl, 2007).
- Transformants were selected by plating on Cd amended SD-Ura media plates



Secondary and tertiary modes of screening for Cd tolerant transformants

- <u>Secondary Screening</u>**:** Drop assay on Cd amended media
- <u>Tertiary Screening</u>**:** 5-FOA test and after FOA drop test



Quaternary modes of screening for Cd tolerant transformants

- <u>Quaternary screening</u>: Yeast Colony PCR
- *After all these screening procedures, plasmids were extracted from these positive yeast transformants and sent for sequencing*

Table 4.2: Tolerance level of yeast transformants corresponding to cDNA library C screened against Cd toxicity. Total 498 clones screened from Lib C, were tested by drop test assay in presence of 40 µM of $CdSO_4$.

Clone no.	Clone name
1	ycfPLCa1
2	ycfPLCa2
3	ycfPLCa3
4	ycfPLCa4
5	ycfPLCa5
6	ycfPLCa6
7	ycfPLCa7
8	ycfPLCa8
9	ycfPLCa9
10	ycfPLCa10
11	**ycfPLCa11**
12	ycfPLCa12
13	ycfPLCa13
14	ycfPLCa14
15	ycfPLCa15
16	ycfPLCa16
17	ycfPLCa17
18	ycfPLCa18
19	ycfPLCa19
20	ycfPLCa20
21	ycfPLCa21
22	**ycfPLCa22**
23	ycfPLCa23
24	ycfPLCa24
25	ycfPLCa25
26	ycfPLCa26
27	ycfPLCa27
28	ycfPLCa28
29	ycfPLCa29
30	ycfPLCa30
31	ycfPLCa31
32	**ycfPLCa32**
33	ycfPLCa33
34	ycfPLCa34
35	ycfPLCa35

Clone no.	Clone name
36	ycfPLCa36
37	ycfPLCa37
38	ycfPLCb1
39	ycfPLCb2
40	ycfPLCb3
41	**ycfPLCb4**
42	ycfPLCb5
43	ycfPLCb6
44	ycfPLCb7
45	ycfPLCb8
46	ycfPLCb9
47	ycfPLCb10
48	ycfPLCb11
49	ycfPLCb12
50	ycfPLCb13
51	ycfPLCb14
52	ycfPLCb15
53	ycfPLCb16
54	ycfPLCb17
55	ycfPLCb18
56	ycfPLCb19
57	ycfPLCb20
58	ycfPLCb21
59	ycfPLCb22
60	ycfPLCb23
61	ycfPLCb24
62	ycfPLCb25
63	ycfPLCb26
64	ycfPLCb27
65	ycfPLCb28
66	ycfPLCb29
67	ycfPLCb30
68	ycfPLCb31
69	ycfPLCb32
70	ycfPLCb33

Clone no.	Clone name
71	ycfPLCb34
72	ycfPLCb35
73	ycfPLCb36
74	ycfPLCb37
75	ycfPLCb38
76	ycfPLCb39
77	ycfPLCb40
78	ycfPLCb41
79	ycfPLCb42
80	ycfPLCb43
81	ycfPLCb44
82	ycfPLCb45
83	ycfPLCb46
84	ycfPLCb47
85	ycfPLCb48
86	ycfPLCb49
87	ycfPLCb50
88	ycfPLCb51
89	ycfPLCb52
90	ycfPLCb53
91	ycfPLCb54
92	ycfPLCc1
93	ycfPLCc2
94	ycfPLCc3
95	ycfPLCc4
96	ycfPLCc5
97	ycfPLCc6
98	ycfPLCc7
99	ycfPLCc8
100	ycfPLCc9
101	ycfPLCc10
102	ycfPLCc11
103	ycfPLCc12
104	ycfPLCc13
105	ycfPLCc14

Clone no.	Clone name
106	**ycfPLCc15**
107	**ycfPLCc16**
108	**ycfPLCc17**
109	**ycfPLCc18**
110	**ycfPLCc19**
111	ycfPLCc20
112	ycfPLCc21
113	ycfPLCc22
114	ycfPLCc23
115	ycfPLCc24
116	ycfPLCc25
117	ycfPLCc26
118	ycfPLCc27
119	ycfPLCc28
120	ycfPLCc29
121	ycfPLCc30
122	ycfPLCc31
123	ycfPLCc32
124	ycfPLCc33
125	ycfPLCc34
126	ycfPLCc35
127	ycfPLCc36
128	**ycfPLCc37**
129	ycfPLCc38
130	ycfPLCc39
131	ycfPLCc40
132	**ycfPLCc41**
133	ycfPLCc42
134	**ycfPLCc43**
135	ycfPLCc44
136	ycfPLCc45
137	ycfPLCc46
138	ycfPLCc47
139	ycfPLCc48
140	**ycfPLCc49**

▓	Highly tolerant	▒	Moderately Tolerant	░	Low Tolerant	Clones with bold white font=Selected after FOA test

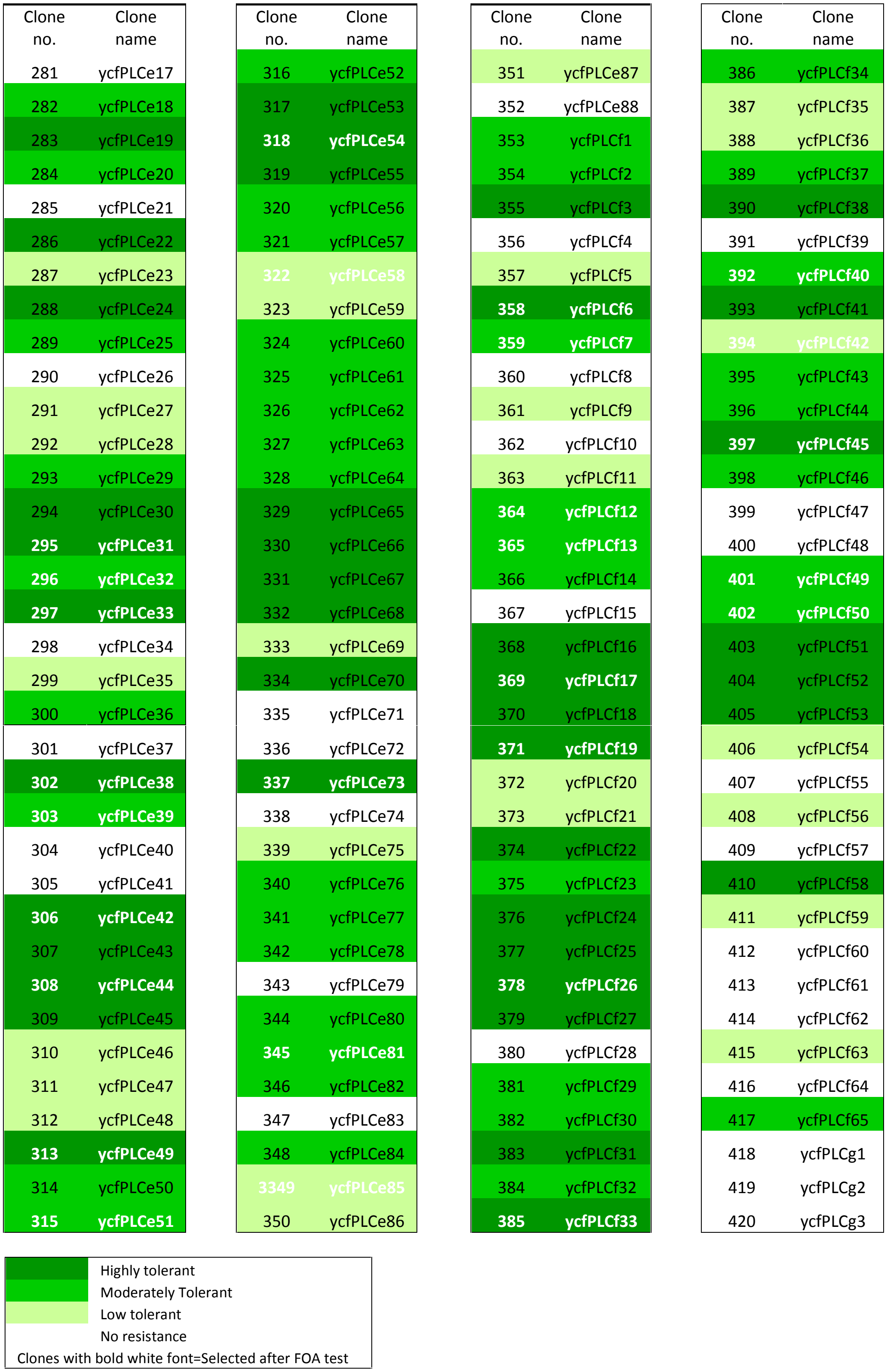
Clone no.	Clone name
■ 141	ycfPLCc50
🟩 **142**	**ycfPLCc51**
■ **143**	**ycfPLCc52**
■ **144**	**ycfPLCc53**
■ 145	ycfPLCc54
■ 146	ycfPLCc55
🟩 147	ycfPLCc56
■ 148	ycfPLCc57
■ 149	ycfPLCc58
🟩 150	ycfPLCc59
■ 151	ycfPLCc60
■ 152	ycfPLCc61
■ **153**	**ycfPLCc62**
■ 154	ycfPLCc63
■ 155	ycfPLCc64
🟩 156	ycfPLCc65
🟩 157	ycfPLCc66
■ 158	ycfPLCc67
■ 159	ycfPLCc68
■ 160	ycfPLCc69
■ 161	ycfPLCc70
□ 162	ycfPLCc71
■ **163**	**ycfPLCc72**
▨ 164	ycfPLCc73
■ 165	ycfPLCc74
■ 166	ycfPLCc75
■ 167	ycfPLCc76
■ 168	ycfPLCc77
■ **169**	**ycfPLCc78**
■ 170	ycfPLCd1
■ 171	ycfPLCd2
■ **172**	**ycfPLCd3**
■ 173	ycfPLCd4
□ 174	ycfPLCd5
■ 175	ycfPLCd6

Clone no.	Clone name
▨ 176	ycfPLCd7
🟩 177	ycfPLCd8
▨ 178	ycfPLCd9
▨ 179	ycfPLCd10
🟩 180	ycfPLCd11
🟩 181	ycfPLCd12
▨ **182**	**ycfPLCd13**
▨ **183**	**ycfPLCd14**
🟩 184	ycfPLCd15
■ 185	ycfPLCd16
■ 186	ycfPLCd17
■ 187	ycfPLCd18
■ 188	ycfPLCd19
■ **189**	**ycfPLCd20**
🟩 **190**	**ycfPLCd21**
🟩 191	ycfPLCd22
🟩 192	ycfPLCd23
🟩 193	ycfPLCd24
🟩 **194**	**ycfPLCd25**
□ 195	ycfPLCd26
■ **196**	**ycfPLCd27**
🟩 **197**	**ycfPLCd28**
■ 198	ycfPLCd29
■ 199	ycfPLCd30
■ **200**	**ycfPLCd31**
■ 201	ycfPLCd32
■ **202**	**ycfPLCd33**
▨ **203**	**ycfPLCd34**
□ 204	ycfPLCd35
■ **205**	**ycfPLCd36**
▨ **206**	**ycfPLCd37**
□ 207	ycfPLCd38
■ **208**	**ycfPLCd39**
🟩 **209**	**ycfPLCd40**
🟩 **210**	**ycfPLCd41**

Clone no.	Clone name
🟩 211	ycfPLCd42
■ **212**	**ycfPLCd43**
▨ 213	ycfPLCd44
■ 214	ycfPLCd45
■ 215	ycfPLCd46
■ 216	ycfPLCd47
■ 217	ycfPLCd48
■ **218**	**ycfPLCd49**
■ 219	ycfPLCd50
■ **220**	**ycfPLCd51**
■ **221**	**ycfPLCd52**
■ 222	ycfPLCd53
■ 223	ycfPLCd54
■ **224**	**ycfPLCd55**
■ **225**	**ycfPLCd56**
■ **226**	**ycfPLCd57**
■ 227	ycfPLCd58
■ 228	ycfPLCd59
▨ 229	ycfPLCd60
🟩 230	ycfPLCd61
▨ 231	ycfPLCd62
□ 232	ycfPLCd63
🟩 233	ycfPLCd64
🟩 **234**	**ycfPLCd65**
🟩 **235**	**ycfPLCd66**
🟩 236	ycfPLCd67
■ 237	ycfPLCd68
🟩 238	ycfPLCd69
🟩 239	ycfPLCd70
🟩 240	ycfPLCd71
■ 241	ycfPLCd72
■ 242	ycfPLCd73
🟩 243	ycfPLCd74
🟩 244	ycfPLCd75
■ 245	ycfPLCd76

Clone no.	Clone name
🟩 246	ycfPLCd77
🟩 247	ycfPLCd78
□ 248	ycfPLCd79
🟩 **249**	**ycfPLCd80**
■ 250	ycfPLCd81
■ 251	ycfPLCd82
🟩 252	ycfPLCd83
□ 253	ycfPLCd84
🟩 254	ycfPLCd85
■ 255	ycfPLCd86
■ 256	ycfPLCd87
■ **257**	**ycfPLCd88**
▨ 258	ycfPLCd89
■ **259**	**ycfPLCd90**
■ 260	ycfPLCd91
🟩 261	ycfPLCd92
🟩 262	ycfPLCd93
🟩 263	ycfPLCd94
■ 264	ycfPLCd95
■ 265	ycfPLCe1
🟩 266	ycfPLCe2
■ 267	ycfPLCe3
🟩 **268**	**ycfPLCe4**
■ **269**	**ycfPLCe5**
▨ 270	ycfPLCe6
■ **271**	**ycfPLCe7**
🟩 **272**	**ycfPLCe8**
🟩 **273**	**ycfPLCe9**
■ **274**	**ycfPLCe10**
🟩 **275**	**ycfPLCe11**
▨ **275**	**ycfPLCe12**
■ **277**	**ycfPLCe13**
■ **278**	**ycfPLCe14**
🟩 **279**	**ycfPLCe15**
□ 280	ycfPLCe16

■ Highly tolerant
🟩 Moderately Tolerant
▨ Low tolerant
□ No resistance
Clones with bold white font=Selected after FOA test



Clone no.	Clone name
281	ycfPLCe17
282	ycfPLCe18
283	ycfPLCe19
284	ycfPLCe20
285	ycfPLCe21
286	ycfPLCe22
287	ycfPLCe23
288	ycfPLCe24
289	ycfPLCe25
290	ycfPLCe26
291	ycfPLCe27
292	ycfPLCe28
293	ycfPLCe29
294	ycfPLCe30
295	**ycfPLCe31**
296	**ycfPLCe32**
297	**ycfPLCe33**
298	ycfPLCe34
299	ycfPLCe35
300	ycfPLCe36
301	ycfPLCe37
302	**ycfPLCe38**
303	**ycfPLCe39**
304	ycfPLCe40
305	ycfPLCe41
306	**ycfPLCe42**
307	ycfPLCe43
308	**ycfPLCe44**
309	ycfPLCe45
310	ycfPLCe46
311	ycfPLCe47
312	ycfPLCe48
313	**ycfPLCe49**
314	ycfPLCe50
315	**ycfPLCe51**

Clone no.	Clone name
316	ycfPLCe52
317	ycfPLCe53
318	**ycfPLCe54**
319	ycfPLCe55
320	ycfPLCe56
321	ycfPLCe57
322	**ycfPLCe58**
323	ycfPLCe59
324	ycfPLCe60
325	ycfPLCe61
326	ycfPLCe62
327	ycfPLCe63
328	ycfPLCe64
329	ycfPLCe65
330	ycfPLCe66
331	ycfPLCe67
332	ycfPLCe68
333	ycfPLCe69
334	ycfPLCe70
335	ycfPLCe71
336	ycfPLCe72
337	**ycfPLCe73**
338	ycfPLCe74
339	ycfPLCe75
340	ycfPLCe76
341	ycfPLCe77
342	ycfPLCe78
343	ycfPLCe79
344	ycfPLCe80
345	**ycfPLCe81**
346	ycfPLCe82
347	ycfPLCe83
348	ycfPLCe84
3349	**ycfPLCe85**
350	ycfPLCe86

Clone no.	Clone name
351	ycfPLCe87
352	ycfPLCe88
353	ycfPLCf1
354	ycfPLCf2
355	ycfPLCf3
356	ycfPLCf4
357	ycfPLCf5
358	**ycfPLCf6**
359	**ycfPLCf7**
360	ycfPLCf8
361	ycfPLCf9
362	ycfPLCf10
363	ycfPLCf11
364	**ycfPLCf12**
365	**ycfPLCf13**
366	ycfPLCf14
367	ycfPLCf15
368	ycfPLCf16
369	**ycfPLCf17**
370	ycfPLCf18
371	**ycfPLCf19**
372	ycfPLCf20
373	ycfPLCf21
374	ycfPLCf22
375	ycfPLCf23
376	ycfPLCf24
377	ycfPLCf25
378	**ycfPLCf26**
379	ycfPLCf27
380	ycfPLCf28
381	ycfPLCf29
382	ycfPLCf30
383	ycfPLCf31
384	ycfPLCf32
385	**ycfPLCf33**

Clone no.	Clone name
386	ycfPLCf34
387	ycfPLCf35
388	ycfPLCf36
389	ycfPLCf37
390	ycfPLCf38
391	ycfPLCf39
392	**ycfPLCf40**
393	ycfPLCf41
394	**ycfPLCf42**
395	ycfPLCf43
396	ycfPLCf44
397	**ycfPLCf45**
398	ycfPLCf46
399	ycfPLCf47
400	ycfPLCf48
401	**ycfPLCf49**
402	**ycfPLCf50**
403	ycfPLCf51
404	ycfPLCf52
405	ycfPLCf53
406	ycfPLCf54
407	ycfPLCf55
408	ycfPLCf56
409	ycfPLCf57
410	ycfPLCf58
411	ycfPLCf59
412	ycfPLCf60
413	ycfPLCf61
414	ycfPLCf62
415	ycfPLCf63
416	ycfPLCf64
417	ycfPLCf65
418	ycfPLCg1
419	ycfPLCg2
420	ycfPLCg3

🟩 Highly tolerant
🟩 Moderately Tolerant
🟩 Low tolerant
No resistance
Clones with bold white font=Selected after FOA test

Clone no.	Clone name
421	ycfPLCg4
422	ycfPLCg5
423	ycfPLCg6
424	ycfPLCg7
425	ycfPLCg8
426	**ycfPLCg9**
427	ycfPLCg10
428	ycfPLCg11
429	ycfPLCg12
430	ycfPLCg13
431	ycfPLCg14
432	ycfPLCg15
433	ycfPLCg16
434	ycfPLCg17
435	ycfPLCg18
436	ycfPLCg19
437	ycfPLCg20
438	ycfPLCg21
439	ycfPLCg22
440	ycfPLCg23

Clone no.	Clone name
441	ycfPLCg24
442	**ycfPLCg25**
443	ycfPLCg26
444	ycfPLCg27
445	ycfPLCg28
446	ycfPLCg29
447	ycfPLCg29
448	ycfPLCg30
449	ycfPLCg31
450	**ycfPLCg32**
451	ycfPLCg33
452	ycfPLCg34
453	ycfPLCg35
454	ycfPLCg36
455	**ycfPLCg37**
456	ycfPLCg38
457	**ycfPLCg39**
458	ycfPLCg40
459	ycfPLCg41
460	ycfPLCg42

Clone no.	Clone name
461	**ycfPLCg43**
462	ycfPLCg44
463	ycfPLCg45
464	ycfPLCg46
465	ycfPLCg47
466	**ycfPLCg48**
467	ycfPLCg49
468	ycfPLCg50
469	ycfPLCg51
470	**ycfPLCg52**
471	ycfPLCg53
472	ycfPLCg54
473	**ycfPLCg55**
474	**ycfPLCg56**
475	ycfPLCg57
476	ycfPLCg58
477	ycfPLCg59
478	ycfPLCg60
479	ycfPLCg61
480	**ycfPLCg62**

Clone no.	Clone name
481	ycfPLCg63
481	ycfPLCg64
482	ycfPLCg65
483	ycfPLCg66
484	ycfPLCg67
485	ycfPLCg68
486	ycfPLCg69
487	ycfPLCg70
488	**ycfPLCg71**
489	**ycfPLCg72**
490	ycfPLCg73
491	ycfPLCg74
492	ycfPLCg75
493	ycfPLCg76
494	**ycfPLCg77**
495	ycfPLCg78
496	ycfPLCg79
497	ycfPLCg80
498	ycfPLCg81

Table 4.3: Tolerance level of yeast transformants corresponding to cDNA library B screened against Cd toxicity. Total 78 clones screened from Lib B were tested by drop test assay in presence of 40 µM of $CdSO_4$.

Clones corresponding to Library B

Clone no.	Clone name
1	**ycfPLBa1**
2	ycfPLBa2
3	ycfPLBa3
4	ycfPLBa4
5	ycfPLBa5
6	**ycfPLBa6**
7	ycfPLBa7
8	ycfPLBa8
9	**ycfPLBa9**
10	ycfPLBa10
11	**ycfPLBa11**

Clone no.	Clone name
21	ycfPLBb3
22	ycfPLBb4
23	ycfPLBb5
24	ycfPLBb6
25	ycfPLBb7
26	ycfPLBb8
27	ycfPLBb9
28	**ycfPLBb10**
29	ycfPLBb11
30	ycfPLBb12
31	ycfPLBb13

Clone no.	Clone name
41	ycfPLBb23
42	ycfPLBb24
43	ycfPLBb25
44	ycfPLBb26
45	ycfPLBb27
46	ycfPLBb28
47	**ycfPLBe1**
48	ycfPLBe2
49	ycfPLBe3
50	ycfPLBe4
51	ycfPLBe5

Clone no.	Clone name
61	ycfPLBe15
62	ycfPLBe16
63	ycfPLBe17
64	ycfPLBe18
65	ycfPLBe19
66	ycfPLBe20
67	ycfPLBf1
68	ycfPLBf2
69	ycfPLBf3
70	ycfPLBf4
71	ycfPLBf5

12	ycfPLBa12	**32**	**ycfPLBb14**	52	ycfPLBe6	**72**	**ycfPLBf6**
13	ycfPLBa13	**33**	**ycfPLBb15**	53	ycfPLBe7	**73**	**ycfPLBf7**
14	ycfPLBa14	34	ycfPLBb16	54	ycfPLBe8	**74**	**ycfPLBf8**
15	**ycfPLBa15**	35	ycfPLBb17	55	ycfPLBe9	**75**	**ycfPLBf9**
16	ycfPLBa16	**37**	**ycfPLBb18**	56	ycfPLBe10	76	ycfPLBf10
17	ycfPLBa17	37	ycfPLBb19	57	ycfPLBe11	**77**	**ycfPLBf11**
18	ycfPLBa18	38	ycfPLBb20	58	ycfPLBe12	78	ycfPLBf12
19	**ycfPLBb1**	39	ycfPLBb21	59	ycfPLBe13		
20	ycfPLBb2	40	ycfPLBb22	60	ycfPLBe14		

Table 4.4: Tolerance level of yeast transformants corresponding to cDNA library A screened against Cd toxicity. Total 80 clones screened from Lib A were tested by drop test assay in presence of 40 µM of $CdSO_4$.

Clones corresponding to Library A

8	Clone name	Clone no.	Clone name	Clone no.	Clone name	Clone no.	Clone name
1	ycfPLAd1	21	ycfPLAf6	41	ycfPLAg6	61	ycfPLAg26
2	**ycfPLAd2**	**22**	**ycfPLAf7**	42	ycfPLAg7	62	ycfPLAg27
3	ycfPLAd3	23	ycfPLAf8	**43**	**ycfPLAg8**	63	ycfPLAg28
4	ycfPLAd4	24	ycfPLAf9	44	ycfPLAg9	**64**	**ycfPLAg29**
5	ycfPLAd5	25	ycfPLAf10	45	ycfPLAg10	65	ycfPLAg30
6	ycfPLAe1	26	ycfPLAf11	46	ycfPLAg11	66	ycfPLAg31
7	**ycfPLAe2**	27	ycfPLAf12	**47**	**ycfPLAg12**	**67**	**ycfPLAg32**
8	ycfPLAe3	**28**	**ycfPLAf13**	48	ycfPLAg13	**68**	**ycfPLAg33**
9	ycfPLAe4	29	ycfPLAf14	**49**	**ycfPLAg14**	**69**	**ycfPLAg34**
10	ycfPLAe5	30	ycfPLAf15	50	ycfPLAg15	**70**	**ycfPLAg35**
11	ycfPLAe6	31	ycfPLAf16	**51**	**ycfPLAg16**	**71**	**ycfPLAg36**
12	**ycfPLAe7**	32	ycfPLAf17	52	ycfPLAg17	72	ycfPLAg37
13	**ycfPLAe8**	33	ycfPLAf18	**52**	**ycfPLAg18**	73	ycfPLAg38
14	**ycfPLAe9**	34	ycfPLAf19	**53**	**ycfPLAg19**	74	ycfPLAg39
15	**ycfPLAe10**	35	ycfPLAf20	55	ycfPLAg20	75	ycfPLAg40
16	ycfPLAf1	36	ycfPLAg1	56	ycfPLAg21	76	ycfPLAg41
17	ycfPLAf2	37	ycfPLAg2	**57**	**ycfPLAg22**	77	ycfPLAg42
18	**ycfPLAf3**	38	ycfPLAg3	**58**	**ycfPLAg23**	78	ycfPLAg43
19	ycfPLAf4	39	ycfPLAg4	**59**	**ycfPLAg24**	79	ycfPLAg44
20	**ycfPLAf5**	40	ycfPLAg5	**60**	**ycfPLAg25**	80	ycfPLAg45

■	Highly tolerant
🟩	Moderately Tolerant
▨	Low tolerant
□	No resistance

Clones with bold white font=Selected after FOA test

Table 4.5: Transformants procured after primary, secondary, tertiary and quaternary screening

Library C	
Total number of clones screened	$5x10^6$
Total number of Cd tolerant transformants selected by drop test	498
Low Resistant	174
Moderate Resistant	126
Highly Resistant	83
Non tolerant	115
Number of true transformants screened after 5-FOA drop test at 60 µM of $CdSO_4$	98
Number of transformants screened after colony PCR	94
Library B	
Total number of clones screened	1,11,949
Total number of Cd tolerant transformants selected by drop test	78
Low Tolerant	28
Moderate Tolerant	5
Highly Tolerant	6
Non Tolerant	44
Number of transformants screened after 5-FOA and drop test at 60 µM $CdSO_4$	12
Number of transformants screened after colony PCR	12
Library A	
Total number of clones screened	1,74,020
Total number of Cd tolerant transformants selected by drop test	80
Low Tolerant	33
Moderate Tolerant	7
Highly Tolerant	20
Non Tolerant	20

Number of transformants screened after 5-FOA and drop test at 60 µM $CdSO_4$	24
Number of transformants screened after colony PCR	24

4.2.4. Plasmid extraction and sequence analysis

True yeast transformants selected after four reliable screening procedures were further studied to characterize and to know the taxonomic origin of cDNA sequences by using bioinformatics analysis. To achieve this, recombinant plasmid DNA carrying the desired construct was extracted from the true transformed *ycf1*$^{\Delta}$ cells by using Zymoprep™ Yeast Plasmid Miniprep II (Zymo Research, USA) and electro-transformed (3-4 µl) into One Shot® TOP10 Electrocomp™ *E. coli* cells (Thermofisher, USA). Recombinant plasmids were further extracted from the prokaryotic host i.e., electrocompetent *E.coli* cells and sequenced (Figure 4.16). Sequences retrieved were analyzed by BLASTX to find the homologous sequences between the screened cDNAs and the GenBank databases (Table 4.6). Out of 112 cDNA inserts, only 35 cDNAs were identified as unique good quality sequences, while other sequences were either identified more than once or could not be read. Preliminary BLAST analysis against databases to identify their putative biological functions reveals that several of these genes encode proteins similar but not identical to the proteins reported to confer metal tolerance among living beings e.g. germin like protein. Similarly, other genes encoded proteins of known functions with no reported role in heavy metal resistance, while other ones have no reported homologs in databases. The full-length ORF carrying cDNA sequences retrieved after BLASTX analysis were deposited at NCBI database under the respective accession numbers.

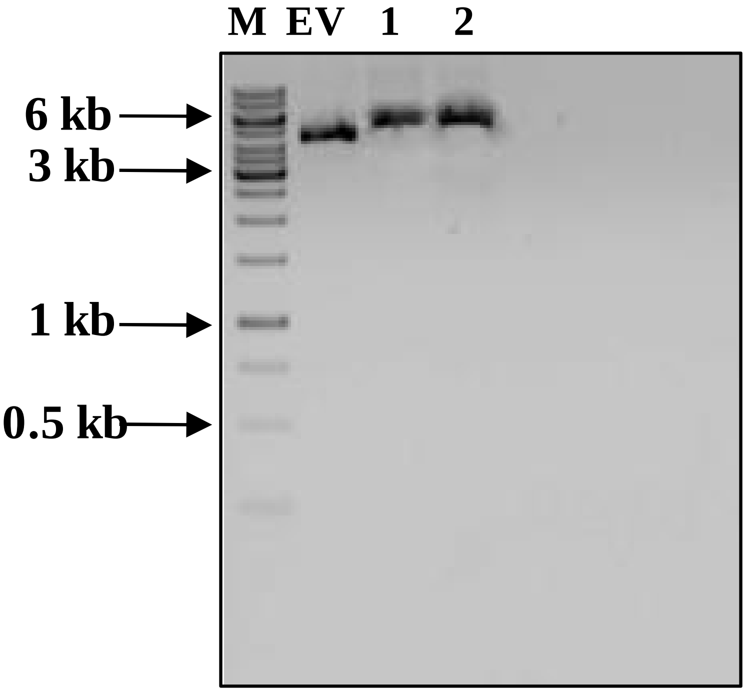


Figure 4.16: Agarose gel showing bands corresponding to recombinant plasmid extracted from Lib B transformants. M denotes 1 kb marker DNA, EV stands for empty pFL61 without cDNA and sample number 1 and 2 corresponds to Lib B transformants.

Recent developed strategies such as metagenomics and metatranscriptomics are shedding light on the microbial composition of a diverse environmental habitat and provides reliable information about their involvement in various ecosystem processes (Tringe and Rubin, 2005; Dinsdale *et al.* 2008). Furthermore, in terms of biotechnology, to understand the content of environmental gene pool, knowledge of metabolic potential of various microbial species flourishing in that environment is of utmost importance. Though metagenomics has been successfully exploited to find novel biocatalysts however, this strategy largely favours the unveiling of the archea and bacteria but not eukaryotes (Daniel, 2005; Steele *et al.* 2009). Absence of eukaryotes in previous studies was not intentional but the outcome of internal characteristics of the eukaryotic

Table. 4.6. Bioinformatic analysis of metatranscriptomic cDNAs screened for cadmium tolerance

S. No.	cDNA	Clone Name	Sequence length	Annotation BLASTX/Swiss Prot	e value	Organism NR
1	PLCc37	$ycf1^{\Delta}$PLCc37	1342	No significant similarity found	2e-34 (Min)	*Oxytricha trifallax*
2	PLCc43	$ycf1^{\Delta}$PLCc43	1039	DNJA2_MOUSE RecName: Full=DnaJ homolog subfamily A member 2	2.00E-103	*Acanthamoeba castellanii* str. Neff
3	PLCc52	$ycf1^{\Delta}$PLCc52	1513	Hypothetical protein	0.77	ND (not determined)
4	PLCc53	$ycf1^{\Delta}$PLCc53	928	Hypothetical protein	0.006	ND
5	PLCc62	$ycf1^{\Delta}$PLCc62	1041	MIOX4_ARATH RecName: Full=Inositol oxygenase 4	0	*Populus*
6	PLCd20	$ycf1^{\Delta}$PLCd20	1180	No significant similarity found	NA	NA
7	PLCd27	$ycf1^{\Delta}$PLCd27	907	Hypothetical protein	8.7	ND
8	PLCd31	$ycf1^{\Delta}$PLCd31	1312	No hit found	3.00E-64	NA
9	PLCd37	$ycf1^{\Delta}$PLCd37	1323	Hypothetical protein	3.00E-85	*Neofusicoccum* parvum UCRNP2
10	PLCd39	$ycf1^{\Delta}$PLCd39	1184	Hypothetical protein	ND	*Drosophila ananassae*
11	PLCd43	$ycf1^{\Delta}$PLCd43	1392	Gamma glutamylcysteinyl transferase [*Arabidopsis thaliana*]		*Arabidopsis thaliana*
12	PLCd49	$ycf1^{\Delta}$PLCd49	1737	LAMB1_HUMAN RecName: Full=Laminin subunit beta-1	1.00E-09	*Paramecium tetraurelia strain* d4-2
13	PLCd56	$ycf1^{\Delta}$PLCd56	1479	Hypothetical protein	4.4	ND
14	PLCe10	$ycf1^{\Delta}$PLCe10	1234	VIT1_CAEEL RecName: Full=Vitellogenin-1	3.00E-35	*Ascaris suum*
15	PLCe11	$ycf1^{\Delta}$PLCe11	1157	Hypothetical protein	NA	NA
16	PLCe13	$ycf1^{\Delta}$PLCe13	1002	Hypothetical protein	4.1	ND
17	PLCe38	$ycf1^{\Delta}$PLCe38	1126	sp|Q6DE96.1|IWS1A_XENLA RecName: Full=Protein IWS1 homolog A; AltName: Full=IWS1-like	4.00E-23	*Citrus sinensis*
18	PLCe42	$ycf1^{\Delta}$PLCe42	1345	EFTU_RECAM RecName: Full=Elongation factor Tu, mitochondrial (mitochondrion)	0	*Andalucia godoyi*

19	PLCf17	*ycf1*$^{\Delta}$PLCf17	1554	PP2C_LEICH RecName: Full=Protein phosphatase 2C	7.00E-21	*Monosiga brevicollis MX1*
20	PLCf19	*ycf1*$^{\Delta}$PLCf19	1203	No significant similarity found	NA	NA
21	PLCf26	*ycf1*$^{\Delta}$PLCf26	1442	No hit found	NA	NA
22	PLCg9	*ycf1*$^{\Delta}$PLCg9	1191	PSD3_SCHPO RecName: Full=Phosphatidylserine decarboxylase proenzyme 3	2.00E-159	*Reticulomyxa filosa*
23	PLCg52	*ycf1*$^{\Delta}$PLCg52	1212	GL12_ARATH RecName: Full=Putative germin-like protein subfamily 1 member	8.00E-18	*Populus trichocarpa*
24	PLCg56	*ycf1*$^{\Delta}$PLCg56	1272	sp|Q941D6.1|HDA14_ARATH RecName: Full=Histone deacetylase 14 [Arabidopsis thaliana] homologue à HDA1 de *S. cerevisiae* avec une e value de 1e-47	1e-114 (Min)	*Nannochloropsis gaditana*
25	PLCg62	*ycf1*$^{\Delta}$PLCg62	1134	No significant similarity found	NA	NA
26	PLCg71	*ycf1*$^{\Delta}$PLCg71	1669	Hypothetical protein	3e-13 (Min)	*Trichoderma harzianum*
27	PLBa11	*ycf1*$^{\Delta}$PLBa11	512	No significant similarity found	NA	NA
28	PLBa15	*ycf1*$^{\Delta}$PLBa15	573	Transcription factor (Conus magus)	9.00E-29	*Conus magus*
29	PLBb10	*ycf1*$^{\Delta}$PLBb10	806	Senescence-associated protein [Gossypium australe]	8.00E-36	*Gossypium australe*
30	PLBb18	*ycf1*$^{\Delta}$PLBc18	807	No significant similarity found	NA	NA
31	PLBe1	*ycf1*$^{\Delta}$PLBe1	694	Ubiquitin domain containing protein [Acanthamoeba castellanii str. Neff]	1.00E-134	*Acanthamoeba castellanii str. Neff*
31	PLBe6	*ycf1*$^{\Delta}$PLBe6	681	Reticulocyte binding protein homologue 1 (RH1), partial [*Plasmodium ovale curtisi*]	1.00E-51	*Plasmodium* ovale curtisi
32	PLBf11	*ycf1*$^{\Delta}$PLBf11	811	Putative reverse transcriptase [*Zingiber officinale*]	5.00E-20	*Zingiber officinale*
34	PLAe2	*ycf1*$^{\Delta}$PLAe2	324	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase [*Algoriphagus* sp. F21]	7.6	*Algoriphagus* sp. F21
35	PLAg22	*ycf1*$^{\Delta}$PLAg22	489	Hypothetical protein BVRB_022340 [*Beta vulgaris* subsp. *vulgaris*]	9.00E-10	*Beta vulgaris* subsp. *vulgaris*

communities (large sized genomes with non coding introns) as well as less sophisticated techniques to study their complex environments (Marmeisse *et al.* 2017). Therefore, to screen the eukaryotic gene pool with specific functions in an environmental sample, exploitation of RNA pool of the sample site through metatranscriptomic libraries is the most efficient way as it relies upon transcriptionally expressed genes (Warnecke and Hess, 2009; Bailly *et al.* 2007). Functional metatranscriptomics exploits the intrinsic characteristic of eukaryotic mRNA to possess 3’ poly-dA tail that can be reverse transcribed into complementary cDNA with the help of poly-dT primers (Wellenreuther *et al.* 2004). These cDNAs were further ligated into a suitable vector and cloned into a suitable prokaryotic or eukaryotic host for transcription and expression of recombinant protein (Kellner *et al.* 2011).

Further, in terms of selection of suitable host for expression of cDNAs, functional approach of metatranscriptomics involves the use of single cell eukaryotes such as *S. cerevisiae*. As a single cell eukaryotic host, *S. cerevisiae* is most commonly used host as it is easy to transform with a wide range of plasmids (integrative or replicating plasmids) carrying regulated or constitutive promoter for expression of eukaryotic cDNAs (Kellner *et al.* 2011; Da Silva and Srikrishnan, 2012). Another important characteristic of using *S. cerevisiae* as host is its ability to express various heterologous eukaryotic proteins with catalytic or structural features that required to be transported to specific intracellular, extracellular or membrane compartments to be biologically active (Lehembre *et al.* 2013; Damon *et al.* 2012). Various studies have reported the use of *S. cerevisiae* for expression of wide range of intracellular, secreted, transport or membrane bound active proteins of fungal (Ramesh *et al.* 2009), plants (Zhen *et al.* 2007) and animal origin (Mukherjee *et* *al.* 2019). The yeast genome project, which focus on developing new mutant strains of *S. cerevisiae* for its biotechnological applications results into numerous yeast mutants. These are attributed with specific characteristics such as *ABC* transporter gene lacking Cd sensitive *ycf1*$^{\Delta}$ strain used to screen Cd tolerant genes (Ziller *et al.* 2017) while strain W303 of *S. cerevisiae* was used to screen oligopeptide transporters from organic matter rich forest soil as it cannot utilise dipeptides as N source (Damon *et al.* 2011).

Availability of genome sequences of numerous microbial species from an environment is an essential resource for precise taxonomic and functional annotations of the environmental sequences and can be helpful in implementing novel gene mining protocols (Marmeisse *et al.* 2017). Various studies have elucidated the metabolic and ecological roles of eukaryotic micro-biota present in diverse and adverse environments such as organic matter rich sites and metal contaminated site through metatranscriptomics approach (Damon *et al.* 2011; Mukherjee *et al.* 2019). Eukaryotic microorganisms, especially fungi are known as major producers of primary metabolites such as organic acids and biologically active secondary metabolites such as antibiotics, immune-suppressors at industrial scale (Hofrichter and Esser, 2010). Lehembre *et al.* (2013) reported different categories of genes such as metallothioneins (MTs), cysteine rich protein (CRPs) responsible for Cd tolerance and BOLA genes responsible for Zn, Mn and Co tolerance by using soil metatranscriptomic approach. While exploring metatranscriptomics for novel genes, Ziller *et al.* (2017) reported metal binding properties of five new CRPs (environmental MTs) conferring tolerance to heavy metals like Zn and Cd.

4.3. Characterization of cadmium tolerant cDNAs

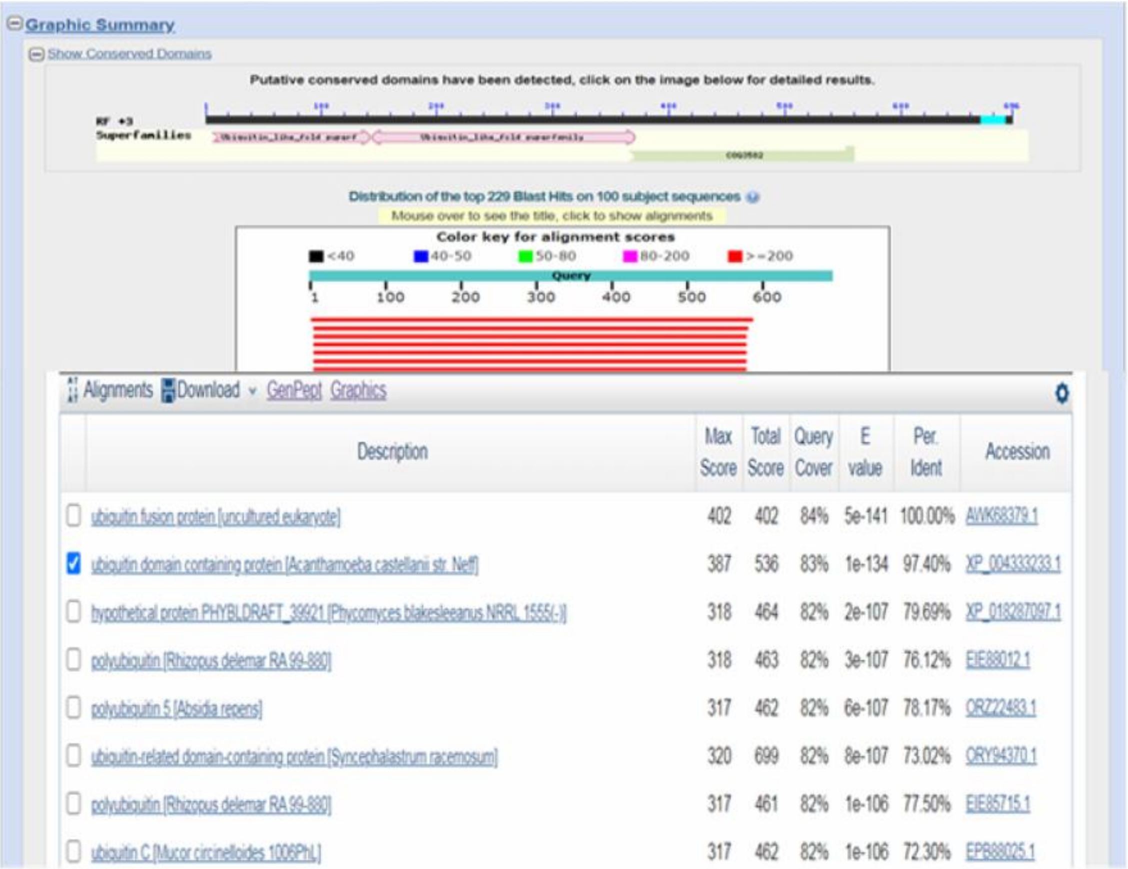
True Cd tolerant transformants carrying full length ORFs selected from above screening procedures i.e., heterologous complementation studies along with BLASTX analysis were further characterized to know their taxonomic profile; putative protein characteristics and functions; expression studies (in case of full-length ORFs) and their capacity to tolerate other metal toxicity. To achieve this, full length sequences with complete ORFs were first characterized by using various bioinformatic tools. The resulting polypeptide sequences were aligned by multiple sequence alignment to reveal their unique characteristics aligning to the genes already deposited in the GenBank databases. These potential proteins were analyzed for the presence of various potential sites such as O- and N-linked glycosylation sites. These proteins were further studied for their tolerance to other metals such as Cu, Zn and Co. In this study, we are reporting the sequence analysis, expression studies in presence of $CdSO_4$, metal tolerance towards other metals and metal accumulation inside respective yeast mutants for four full length cDNA sequences. These sequences were submitted to NCBI under accession numbers i.e., MH244348 (Ubiquitin fusion protein), MF770726 (von Willebrand factor type D domain), MK079355 (Heat shock protein type I like protein) and MK248844 (Gamma glutamylcysteinyl transferase).

4.3.1. PLBe1/Ubiquitin fusion protein

Screening of library B size range between 0.5 to 1.0 kb has given a ORF carrying cDNA PLBe1. Metatranscriptomic cDNA PLBe1 conferred tolerance towards toxic concentration of Cd therefore tested for their relative expression in presence of Cd and determined the characteristics of encoded polypeptide such as structure features and tolerance capabilities towards other metals used in the study.

4.3.1.1 Sequence analysis

The BLASTX analysis with the deduced nucleotide sequences of PLBe1 revealed 693 bp cDNA showing 97% identity (98% positives) with ubiquitin domain containing protein (UFP) fused with AN1 type Zn finger of *Acanthameoba castellani* (XP_004333233) (Figure 4.17). cDNA PLBe1 sequence possessed ORF of 447 bp encoding a polypeptide of 148 amino acids (Figure 4.18). The predicted molecular mass of peptide was 38.96 kDa (average mass) with pI of 4.77. Sequence obtained in this study was deposited at NCBI under accession number MH244348. The encoding polypeptide showed two prime characteristic features: presence of N-terminal 76 aa ubiquitin domain followed by an extension of AN1 type Zn finger (Zf_AN1) located at C-terminal. Amino acid sequence alignment of the deduced polypeptide with various other members of ubiquitin domain containing protein or zinc finger fusion protein showed the presence of conserved domains in the sequence, which advocates this polypeptide as a member of UFP (Figure. 4.19). As shown in Figure 4.20, presence of diglycine residues at the N- terminal of ubiquitin domain in fusion protein UFP signify the proteolytic processing of ubiquitin moiety (Walters *et al.* 2002), while lysine residue at 48 and 63 positions (K48 and K63) advocates the formation of polyubiquitin chains through chain elongation via K48 and K63 after posttranslational modification (Pickart and Fushman, 2004). K48 residue site corresponds as an attachment site for other ubiquitin molecules, which showed that the ubiquitin moiety of UFP after posttranslational modifications helps in binding of other ubiquitin molecules to form polyubiquitin chains for protein degradation while K63 corresponds to non-degradative signaling pathways for homeostasis. (Finley *et al.* 1987; Redman and Rechsteiner, 1989). In the present study, UFP has ubiquitin domain along with 66 amino acid long extension protein with characteristic zinc finger features (Figure 4.20).



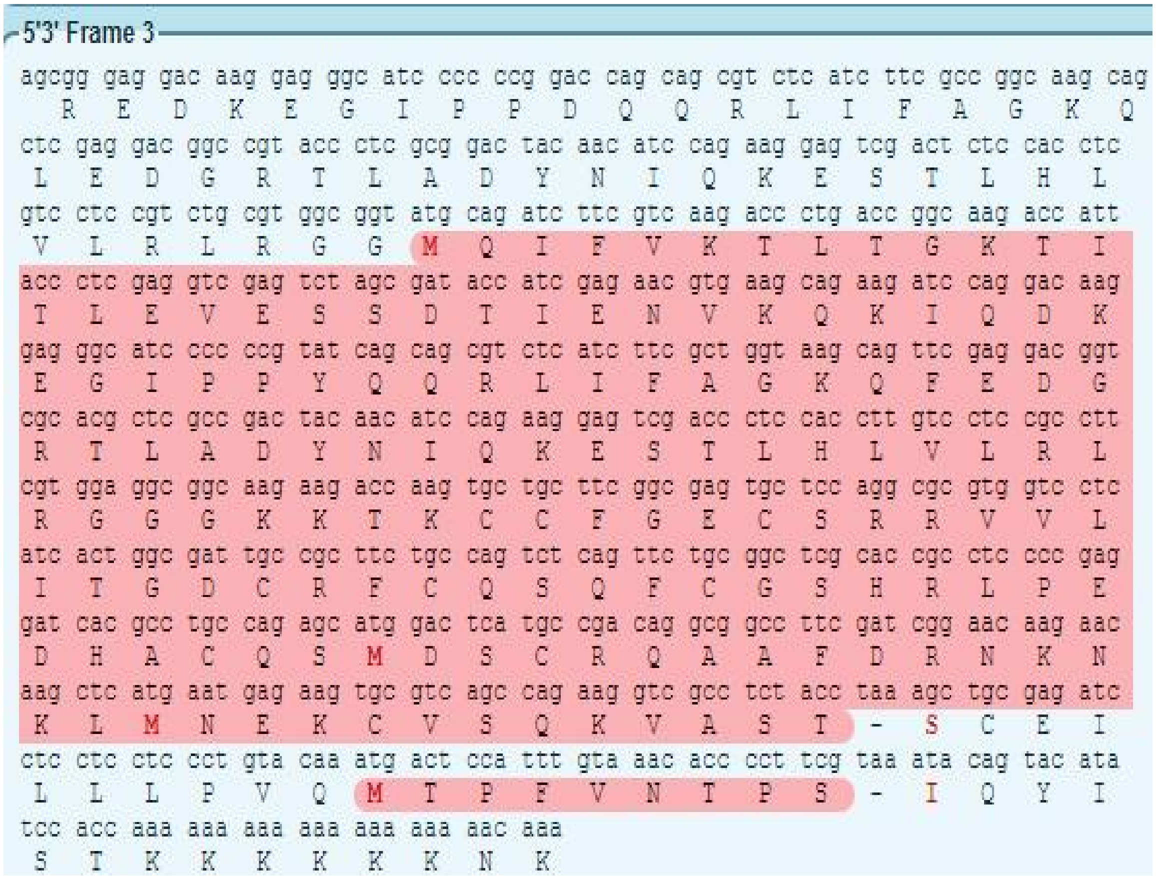
Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 229 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☐	ubiquitin fusion protein [uncultured eukaryote]	402	402	84%	5e-141	100.00%	AWK68379.1
☑	ubiquitin domain containing protein [Acanthamoeba castellanii str. Neff]	387	536	83%	1e-134	97.40%	XP_004333233.1
☐	hypothetical protein PHYBLDRAFT_39921 [Phycomyces blakesleeanus NRRL 1555(-)]	318	464	82%	2e-107	79.69%	XP_018287097.1
☐	polyubiquitin [Rhizopus delemar RA 99-880]	318	463	82%	3e-107	76.12%	EIE88012.1
☐	polyubiquitin 5 [Absidia repens]	317	462	82%	6e-107	78.17%	ORZ22483.1
☐	ubiquitin-related domain-containing protein [Syncephalastrum racemosum]	320	699	82%	8e-107	73.02%	ORY94370.1
☐	polyubiquitin [Rhizopus delemar RA 99-880]	317	461	82%	1e-106	77.50%	EIE85715.1
☐	ubiquitin C [Mucor circinelloides 1006PhL]	317	462	82%	1e-106	72.30%	EPB88025.1

Figure 4.17: Sequence homology of cDNA PLBe1 showing 97% identity (98% positives) with ubiquitin domain containing protein (UFP) fused with AN1 type Zn finger of *Acanthameoba castellani*.

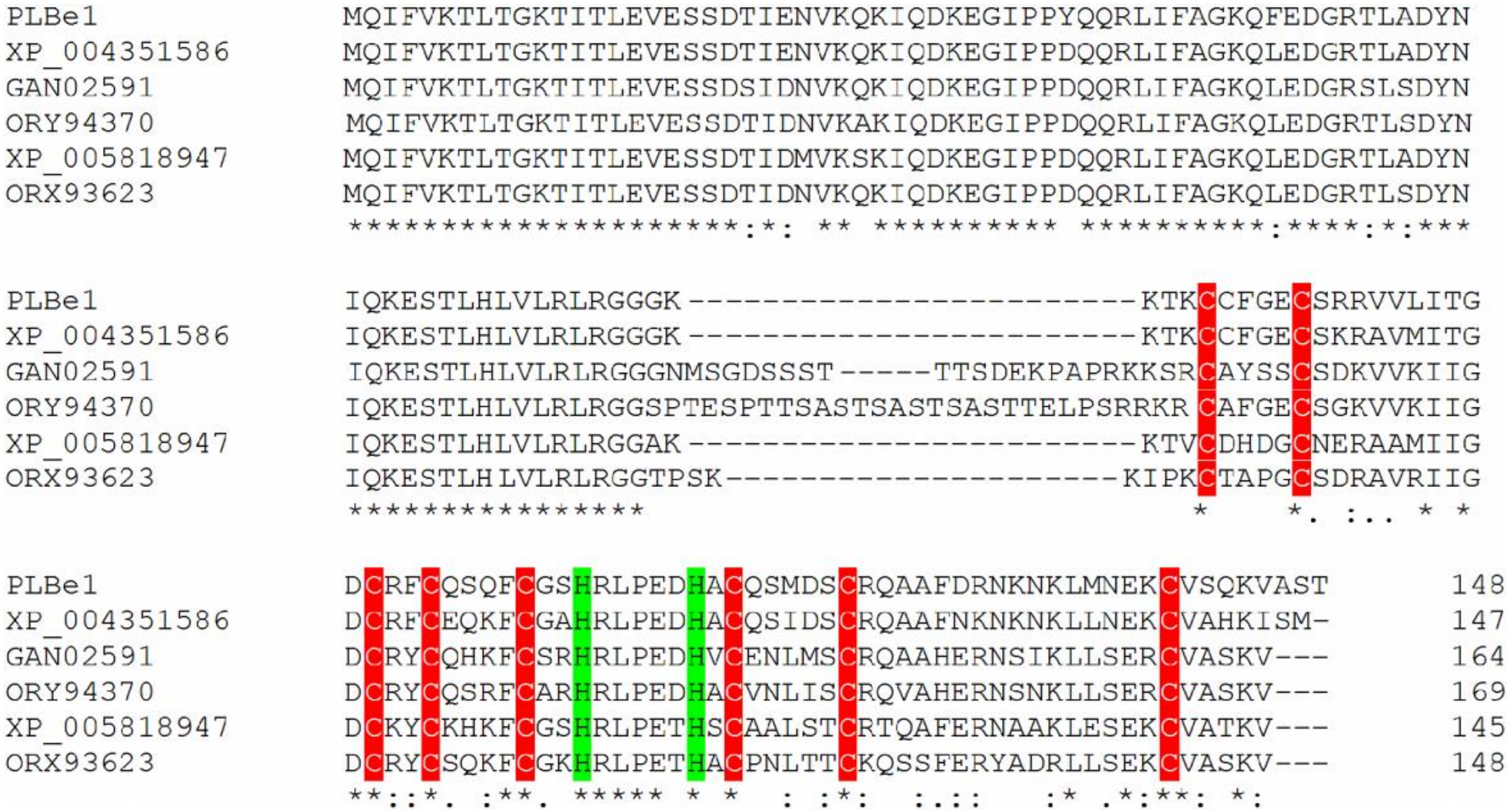


5'3' Frame 3

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

Figure 4.18: Translation of nucleotide sequence of cDNA PLBe1. Red highlighted region is full length ORF showing the nucleotide sequence and encoded amino acids.

These non-related proteins or extension proteins are either 52 aa in all species or varies from 76 aa to 80 aa in yeasts and human. The characteristic pattern of Zn finger (C-X2-C-X(9–12)-C-X(1–2)-C-X4-C-X2-H-X5-H-X-C) at the C terminal of this fusion protein corresponds to Zf_AN1 and structure of Zn ring finger showed two interleaved Zn binding sites (Freemont, 2000). Klug and Rhodes (1987) reported that UFP comprised of two domains where carboxyl group of ubiquitin domain can be linked to amino group of one or two small non-related proteins varying in length from 52 aa to 80 aa. Generally these C -terminal extension proteins are highly basic; contain 4 cysteine residue sequences in consensus with the Zn^{2+} finger domain present in



```
PLBe1               MQIFVKTLTGKTITLEVESSDTIENVKQKIQDKEGIPPYQQRLIFAGKQFEDGRTLADYN
XP_004351586        MQIFVKTLTGKTITLEVESSDTIENVKQKIQDKEGIPPDQQRLIFAGKQLEDGRTLADYN
GAN02591            MQIFVKTLTGKTITLEVESSDSIDNVKQKIQDKEGIPPDQQRLIFAGKQLEDGRSLSDYN
ORY94370            MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYN
XP_005818947        MQIFVKTLTGKTITLEVESSDTIDMVKSKIQDKEGIPPDQQRLIFAGKQLEDGRTLADYN
ORX93623            MQIFVKTLTGKTITLEVESSDTIDNVKQKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYN
                    *********************:*: ** ********** **********:****:*:***

PLBe1               IQKESTLHLVLRLRGGGK------------------------KTKCCFGECSRRVVLITG
XP_004351586        IQKESTLHLVLRLRGGGK------------------------KTKCCFGECSKRAVMITG
GAN02591            IQKESTLHLVLRLRGGGNMSGDSSST-----TTSDEKPAPRKKSRCAYSSCSDKVVKIIG
ORY94370            IQKESTLHLVLRLRGGSPTESPTTSASTSASTSASTTELPSRRKRCAFGECSGKVVKIIG
XP_005818947        IQKESTLHLVLRLRGGAK------------------------KTVCDHDGCNERAAMIIG
ORX93623            IQKESTLHLVLRLRGGTPSK---------------------KIPKCTAPGCSDRAVRIIG
                    ****************                             *    *. :.. * *

PLBe1               DCRFCQSQFCGSHRLPEDHACQSMDSCRQAAFDRNKNKLMNEKCVSQKVAST    148
XP_004351586        DCRFCEQKFCGAHRLPEDHACQSIDSCRQAAFNKNKNKLLNEKCVAHKISM-    147
GAN02591            DCRYCQHKFCSRHRLPEDHVCENLMSCRQAAHERNSIKLLSERCVASKV---    164
ORY94370            DCRYCQSRFCARHRLPEDHACVNLISCRQVAHERNSNKLLSERCVASKV---    169
XP_005818947        DCKYCKHKFCGSHRLPETHSCAALSTCRTQAFERNAAKLESEKCVATKV---    145
ORX93623            DCRYCSQKFCGKHRLPETHACPNLTTCKQSSFERYADRLLSEKCVASKV---    148
                    **::*. :**. ***** * *  : :*:  :.::   :* .*:**: *:
```

Figure 4.19: Multiple sequence alignment of PLBe1 gene with other homologous sequences retrieved by BLASTp analysis. Alignment was performed using the Clustal omega program and manually adjusted. Fully conserved cysteine residues (highlighted in red) and histidine residues (highlighted in green) of C– terminal zinc finger domain is shown. Accession numbers are *Mucor ambiguus* (GAN02591) *Syncephalastrum racemosum* (ORY94370), *Acanthameoba castellini* (XP_004351586), *Guillardia theta* CCMP2712 (XP_005818947), *Basidiobolus meristosporus* CBS 931.73 (ORX93623).

MQIFVKTLTGKTITLEVESSDTIENVKQKIQDKEGIPPYQQRLIFAGKQF 50

EDGRTLADYNIQKESTLHLVLRLRGGGKKTKCCFGECSRRVVLITGDCRF 100

CQSQFCGSHRLPEDHACQSMDSCRQAAFDRNKNKLMNEKCVSQKVAST 147

Figure 4.20: Amino acid sequence of UFP encoded by cDNA PLBe1: Highly conserved lysine residue at 48 position (green highlighted) and diglycine residue in the C-terminal end of ubiquitin moiety at 75 and 76 amino acid (highlighted in blue). Cysteine and histidine residues of C- terminal zinc finger moiety of the fusion protein are highlighted in red. Phosphorylation sites are highlighted in grey.

nucleic acid binding proteins (Kalderon *et al.* 1984). Linnen *et al.* (1993) characterized similar fusion proteins encoded by two isoforms of *An1* mRNA in *Xenopus laevis*. It had been reported that ubiquitin is available in the living beings either as polyubiquitin form or as a fusion protein at C-terminal and their post-translational modifications led to formation of ubiquitin monomers. Various studies reported the immediate cleavage of these fusion proteins encoded from genetically engineered ubiquitin fusion genes hence depicted the simultaneous *in vivo* processing of amino peptide linked ubiquitin (Bachmair *et al.* 1986; Bachmair and Varshavsky, 1989; Butt *et al.* 1988). Characterization of 147 aa long polypeptide also depicts various metabolically significant sites such as 14 potential sites for phosphorylation and three O-linked glycosylation sites Secondary structure of polypeptide encoded by cDNA PLBe1 showed the presence of different structural arrangements among the amino acids sequences (Figure 4.21). Phylogenetic analysis of cDNA PLBe1 with the similar genes corresponding to various taxons has clearly clustered this cDNA with phylum Amoebozoa which support the fact that this gene might have its origin from the kingdom Protista (Fig 4.22).

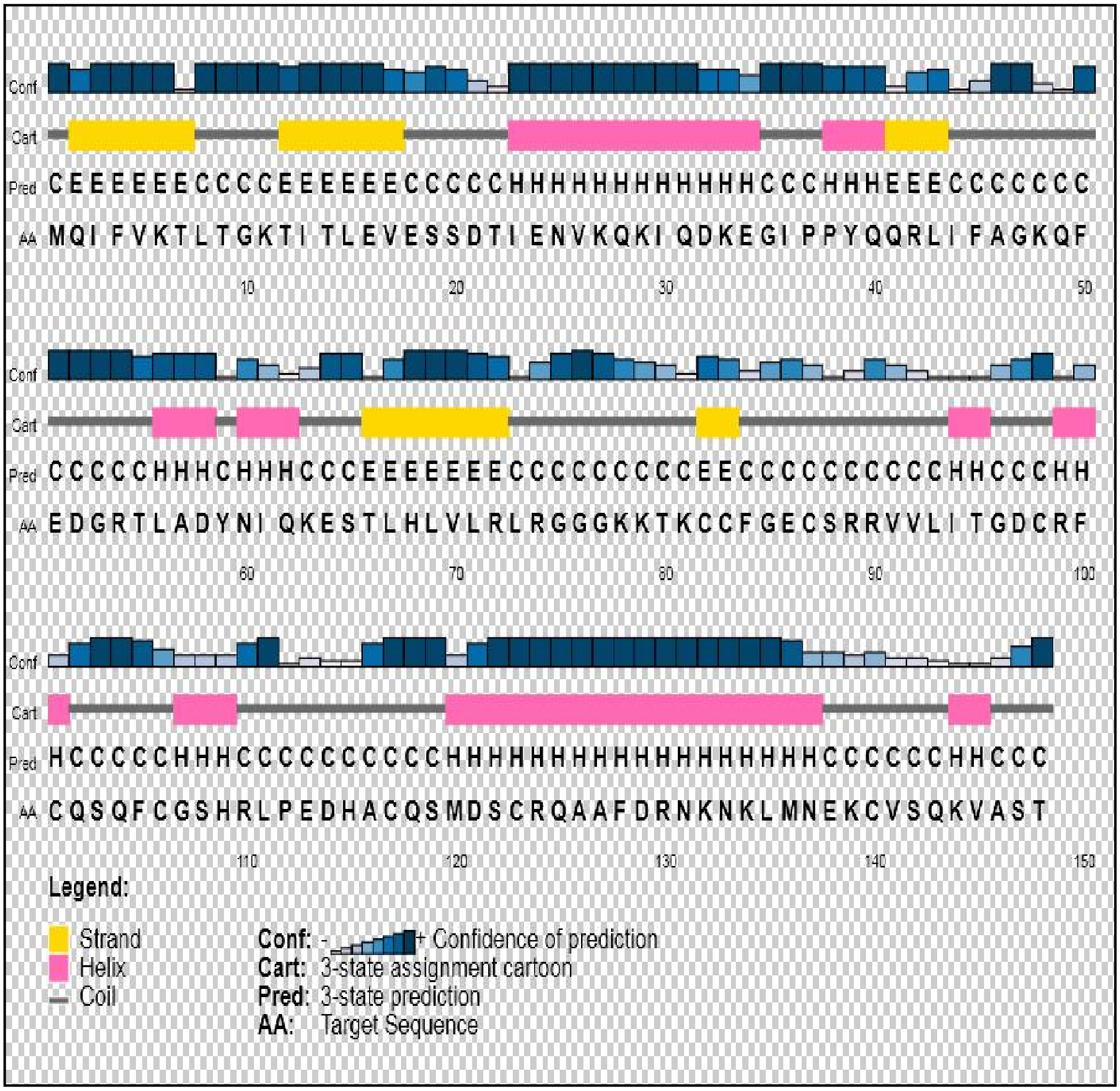


Figure 4.21: Secondary structure of deduced amino acid sequence encoded by cDNA PLBe1 showing various spatial arrangements of amino acids.

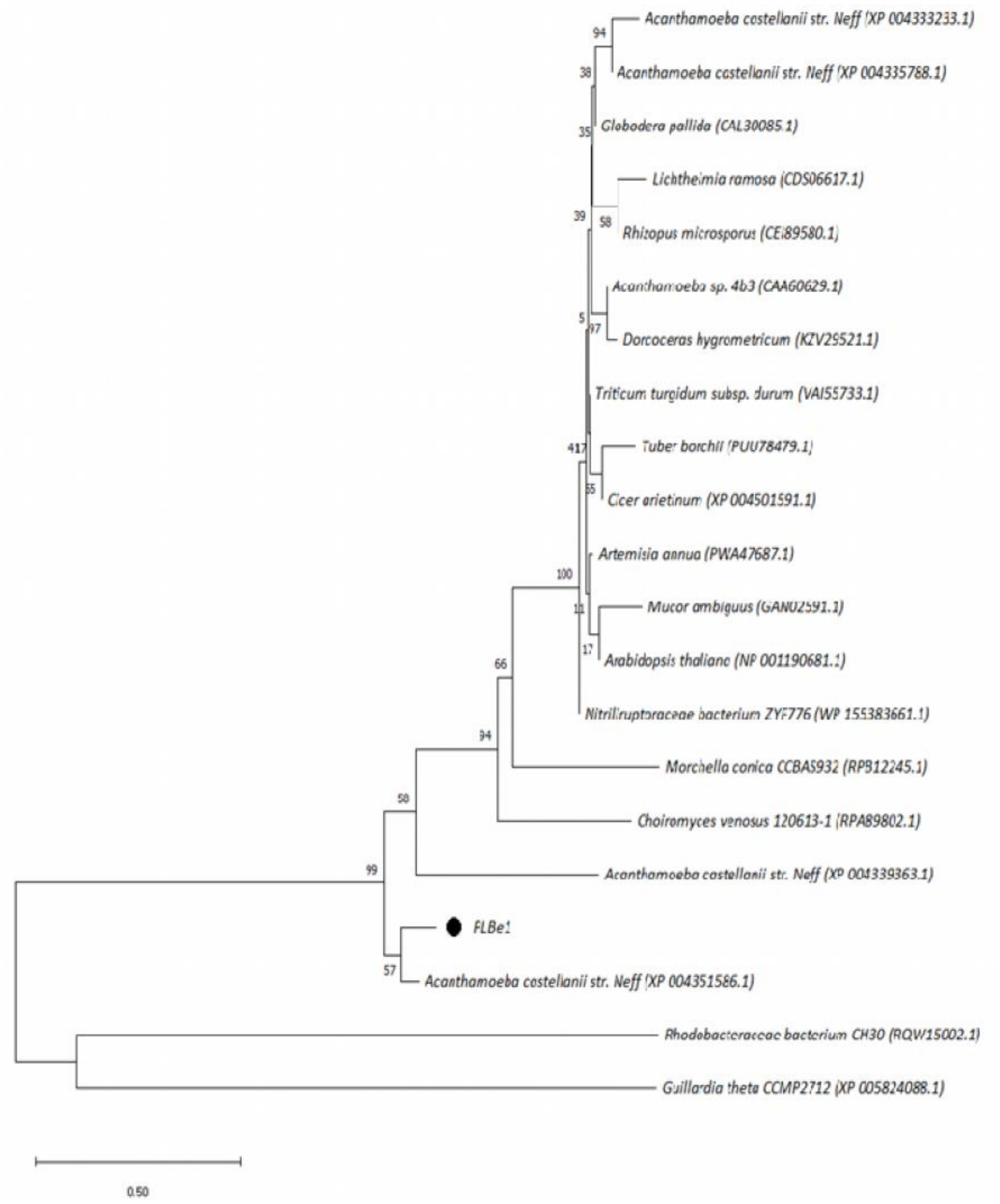


Figure 4.22: Phylogenetic relation of the transcript PLBe1 with other taxa corresponding to similar proteins constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side where as GenBank accession numbers are given in parentheses. Branch lengths are proportional to evolutionary distances.

4.3.1.2 Relative quantification of cDNA PLBe1

Quantity of mRNA accumulated inside the yeast cells carrying cDNA PLBe1 on stimulation by Cd metal was estimated by quantitative RT-PCR. Relative expression of mRNA increases with the increase in metal concentration (Table 4.7). At 40 µM of $CdSO_4$, the expression level of cDNA PLBe1 upraised to 27 folds which further increased with increase in metal concentration i.e., up to a level of 49 times at 60 µM of $CdSO_4$ with a highest level of expression at 80 µM of $CdSO_4$ where the fold of expression was reached up to eighty nine fold. Experimental values obtained at three different level of Cd concentration inferred that with the increase in metal concentration, the level of expression of cDNA PLBe1 also increased (Figure 4.23). Similar findings were reported for induction of cDNA *OjaUb* isolated from *Oryzias javanicusm* (Javanese ricefish) on exposure to potentially toxic metal ions Cd^{2+}, Cu^{2+}, Ag^{2+}, Zn^{2+}, Cr^{2+} and Ni^{2+} and observed that the cDNA *OjaUb* shows maximum level of expression in presence of Cd^{2+} metal ion (Woo *et al.* 2009).

Table 4.7: Effect of different cadmium concentrations on the relative expression of cDNA PLBe1

Cd Concentration (µM)	**Fold expression**
40	27.54 ± 1.34a
60	49.26 ± 1.39b
80	89.18 ± 2.47c

Values sharing a common letter within the column are not significant at $P<0.05$ (Mean±SD, n=3)

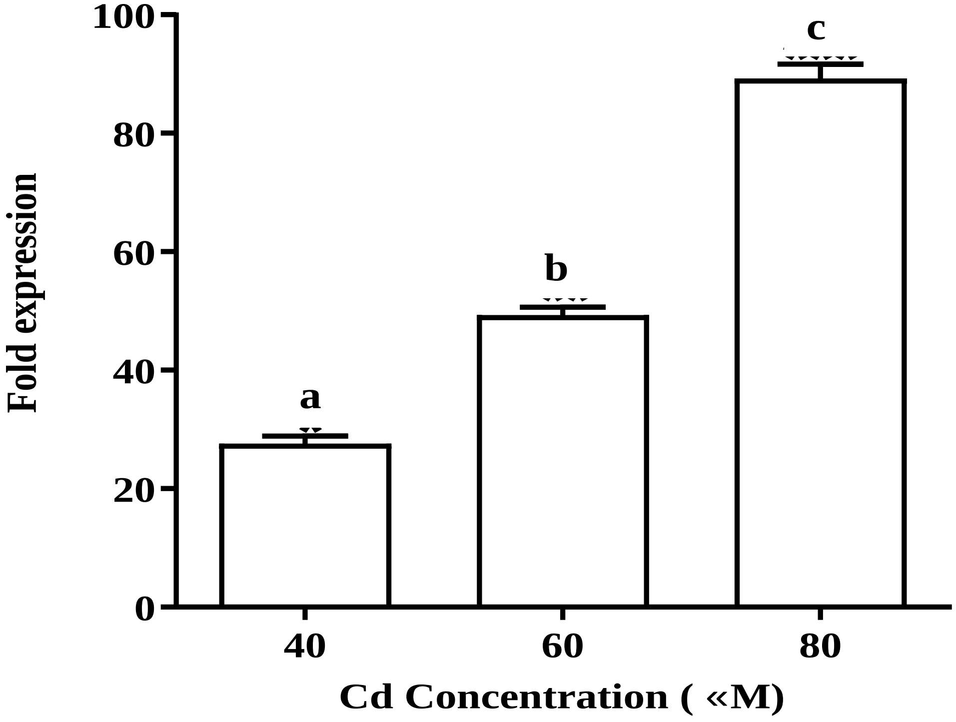


Figure 4.23: Fold increase in expression levels of cDNA PLBe1 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCe10 without metal treatment). Bars sharing a common letter are not significant at $P<0.05$ ($n=3$). Error bars are ±SD.

4.2.1.3. Multi-metal tolerance of PLBe1

To validate the role of PLBe1 towards the tolerance of metals such as Cd, Cu, Zn and Co, the recombinant plasmid carrying PLBe1 cDNA was transformed into *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$ mutant strains of *S. cerevisiae*. The cDNA PLBe1 was able to provide tolerance to Cd hypersensitive mutant *ycf1*$^{\Delta}$ at 40 µM of $CdSO_4$ in comparison to the wild strain and the mutant strain transformed with empty vector pFL61 (Figure 4.24a). Further, it was observed that the Cd tolerant cDNA PLBe1 also conferred tolerance to the hypersensitive yeast mutant *cup1*$^{\Delta}$ (DTY4) at 150 µM Cu while no growth was observed

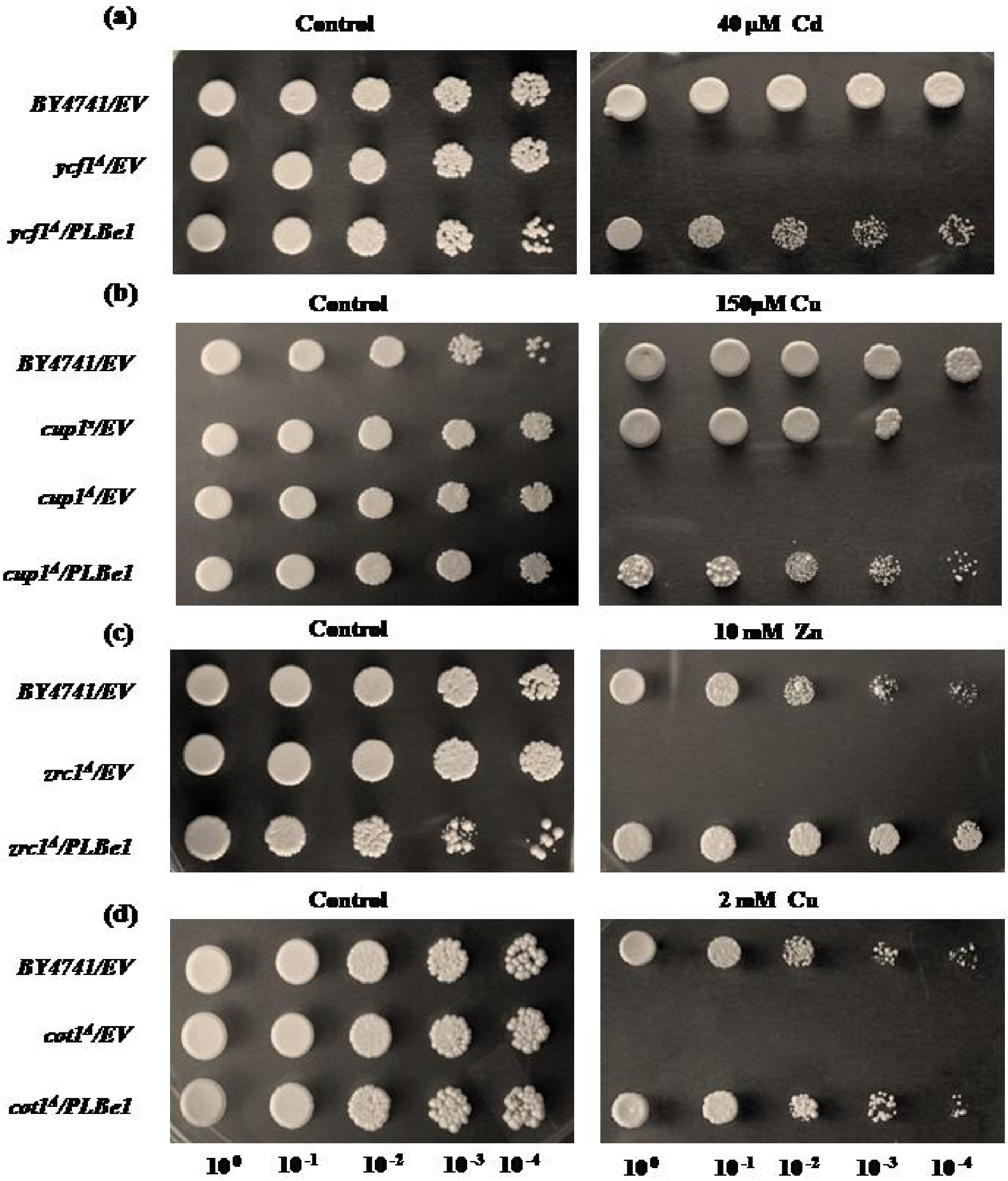


Figure 4.24: Drop assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLBe1 at various concentrations a) Cd^{2+} 40 µM, b) Cu^{2+} 150 µM, c) Zn^{2+} 10 mM and d) Co^{2+} 2 mM. Wild strain BY4741 and *cup1*^{s} are used as control and EV denotes to empty vector pFL61.

for yeast mutant *cup1*$^{\Delta}$ transformed with empty vector pFL61 (Figure 4.24b). In Zn amended (10 mM) media, prolific growth was observed for transformant *zrc1*$^{\Delta}$/PLBe1 in comparison to the mutant strain *zrc1*$^{\Delta}$pFL61 (Figure 4.24c). Similar results were obtained for Co metal (2 mM) while studying the restoring efficiency of cDNA PLBe1 towards cobalt toxicity. The transformant *cot1*$^{\Delta}$/PLBe1 grows at a similar rate to that of the wild strain BY4741 in comparison to *cot1*$^{\Delta}$pFL61 (Figure 4.24d). In general, for drop assay, a usual fall in growth was observed for all the yeast strains in metal amended media. To study the metal tolerant profiling of cDNA PLBe1 for various metal concentrations, the transformants were grown in liquid media supplemented with different concentrations of the metals and growth was monitored after 48 hrs. Growth assays of different transformants showed significant results on different range of concentration of metals (Table 4.8). It was observed that the screened transformant *ycf1*$^{\Delta}$/PLBe1 was able to tolerate high concentrations of Cd. The maximum growth of this transformant was observed at 40 µM $CdSO_4$. With increase in concentration of $CdSO_4$, reduction in the growth was observed. However, growth was significantly increased compared to the wild type strain. No growth was observed when *ycf1*$^{\Delta}$ transformed with empty pFL61 in presence of Cd (Figure 4.25a). Similarly, the transformant *cup1*$^{\Delta}$/PLBe1 was able to grow up to 300 µM $CuSO_4$ observing the maximum growth at 150 µM. It was observed that *cup1*$^{\Delta}$/PLBe1 conferred approximately four times more tolerance compared to the *cup1*^{s} (DTY3) which possess both copies of CUP1 gene at 150 µM $CuSO_4$ concentration. The growth of *cup1*$^{\Delta}$/PLBe1 was significantly higher than its wild type and *cup1*$^{\Delta}$ transformed with empty vector (Figure 4.25b). The growth of *zrc1*$^{\Delta}$/PLBe1 also increased its tolerance

Table 4.8: Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLBe1 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant strain with empty vector pFL61 (EV).

cDNA PLBe1			
O.D. 600			
Cadmium (µM)	***ycf1***$^{\Delta}$***/EV***	***BY4741/EV***	***ycf1***$^{\Delta}$***/* PLBe1**
40	0.09 ± 0.01a	1.42 ± 0.116b	4.90 ± 0.37c
60	0.06 ± 0.01a	0.75 ± 0.29b	2.61 ± 0.44c
80	0.04 ± 0.006a	0.39 ± 0.18b	0.69 ± 0.20b
O.D. 600			
Copper (µM)	***cup1***$^{\Delta}$***/EV***	***cup1***^{s}***/EV***	***cup1***$^{\Delta}$***/* PLBe1**
150	0.24 ± 0.032a	0.56 ± 0.01b	1.94 ± 0.23c
300	0.07 ± 0.007a	0.19 ± 0.03b	1.27 ± 0.08c
500	0.05 ± 0.001a	0.04 ± 0.001b	1.50 ± 0.15c
O.D. 600			
Zinc (mM)	***zrc1***$^{\Delta}$***/EV***	***BY4741/EV***	***zrc1***$^{\Delta}$***/* PLBe1**
10	0.34 ± 0.07a	1.38 ± 0.19b	2.52 ± 0.11c
11	0.14 ± 0.03a	0.96 ± 0.03b	1.1 ± 0.11c
12	0.084 ± 0.014a	0.48 ± 0.09b	0.9 ± 0.04c
O.D. 600			
Cobalt (mM)	***cot1***$^{\Delta}$***/EV***	***BY4741/EV***	***cot1***$^{\Delta}$***/* PLBe1**
2	0.26 ± 0.07a	1.49 ± 0.19b	2.58 ± 0.28c
3	0.07 ± 0.007a	0.97 ± 0.11b	1.45 ± 0.17c
4	0.05 ± 0.009a	0.14 ± 0.02b	1.14 ± 0.11c

Values sharing common letters for a particular metal concentration between samples are not significantly different at $P<0.05$ (O.D. 600 values are mean ± SD, n=3).

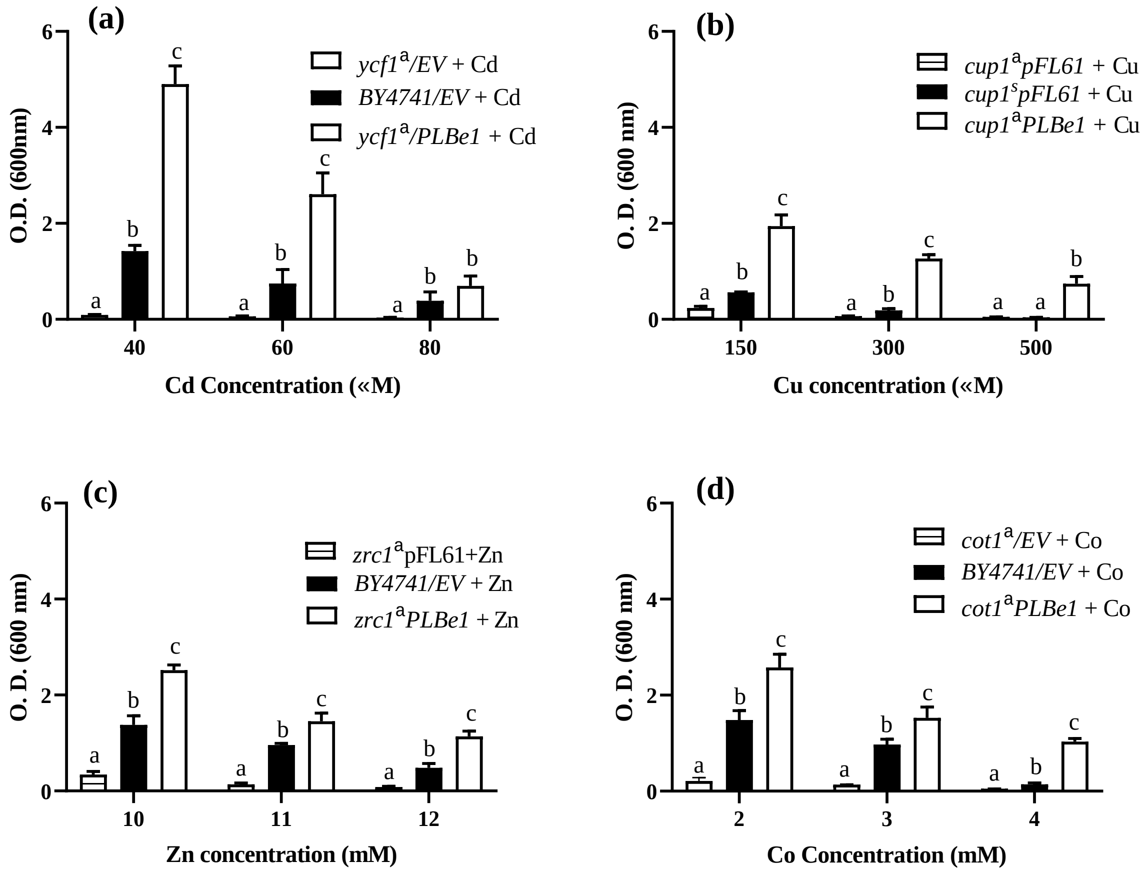


Figure 4.25: Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLBe1 in SD-Ura medium supplemented with (a) Cd 40-80 µM, (b) Cu 150-500 µM, (c) Zn 10-12 mM, (d) Co 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3).

to high concentrations of $ZnSO_4$ (12 mM) showing maximum growth at 10 mM of $ZnSO_4$. It was observed that the growth of transformant *zrc1*$^{\Delta}$/PLBe1 at 10 mM concentration of $ZnSO_4$ was seven times more when compared to the wild strain BY4741. The growth of *zrc1*$^{\Delta}$/PLBe1 was significantly higher than its wild type and *zrc1*$^{\Delta}$ transformed with empty vector (Figure 4.25c). The cobalt sensitive yeast transformant carrying cDNA PLBe1 (*cot1*$^{\Delta}$pFL61) had shown high tolerance to Co with maximum growth at 2 mM of $CoCl_2$ concentration (Figure 4.25d). The growth assay had shown that *cot1*$^{\Delta}$/PLBe1 was ten times more capable of tolerating 2 mM of $CoCl_2$ concentration in comparison to yeast mutant *cot1*$^{\Delta}$pFL61. The growth of *cot1*$^{\Delta}$/PLBe1 was much higher than its wild type and mutant strain transformed with empty vector in all the concentrations of $CoCl_2$ tested. Therefore, ubiquitin fusion protein encoded by cDNA PLBe1 showed immense tolerance towards all the four metals in the study in an order: $Zn^{2+} > Co^{2+} > Cu^{2+} > Cd^{2+}$.

4.3.1.4. Yeast metal uptake by cDNA PLBe1

Different hypersensitive yeast mutants transformed with cDNA PLBe1 was further tested for level of metal accumulation inside the cells by estimation of acid digested product of yeast cells through ICP-MS. While estimating the metal accumulation efficiency of yeast cells carrying cDNA PLBe1, it was observed that cDNA PLBe1 helped in accumulating higher range of metals in comparison to the wild type strain and hypersensitive mutant strains (Table 4.9). For Cd metal, the level of accumulation was maximum at 40 µM of $CdSO_4$, which further decreased with increase in the Cd metal concentration (Figure 4.26a). For Cu metal, the level of accumulation increased up to 300 µM of $CuSO_4$ while shown a very low accumulation at 500 µM of $CuSO_4$ among the Cu concentrations tested in this study (Figure 4.26b). Accumulation study for Zn metal, maximum accumulation was observed at 10 mM of $ZnSO_4$ which further decreases with increasing Zn metal concentration with minimum accumulation reported at 12 mM of $ZnSO_4$ (Figure 4.26c). For Co metal, metal accumulation decreased with increase in concentration of $CoCl_2$ and cDNA PLBe1 had shown minimum metal accumulation tendency for Co^{2+} metal ion among all four metals (Figure 4.26d). Therefore, cDNA PLBe1 encoding ubiquitin fusion protein showed good amount of metal accumulation towards all the four metals in the study in an order: $Zn^{2+} > Co^{2+} > Cu^{2+} > Cd^{2+}$.

Table 4.9: Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLBe1 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV).

cDNA PLBe1			
Cadmium uptake (mg^{-1}g dry weight)			
Cd (µM)	***ycf1*$^{\Delta}$/ PLBe1**	***BY4741/EV***	***ycf1*$^{\Delta}$/*EV***
40	0.92 ± 0.019a	0.15 ± 0.012b	0.015 ± 0.001c
60	0.64 ± 0.05a	0.25 ± 0.002b	0.025 ± 0.010c
80	0.027 ± 0.019a	0.05 ± 0.00b	0.009 ± 0.001c
Copper uptake (mg^{-1}g dry weight)			
Cu (µM)	***cup1*$^{\Delta}$/ PLBe1**	***BY4741/EV***	***cup1*$^{\Delta}$/*EV***
150	0.50 ± 0.006a	0.36 ± 0.013b	0.026 ± 0.001c
300	0.66 ± 0.023a	0.22 ± 0.005b	0.022 ± 0.001c
500	0.36 ± 0.027a	0.16 ± 0.009b	0.008 ± 0.001c
Zinc uptake (mg^{-1}g dry weight)			
Zn (mM)	***zrc1*$^{\Delta}$/ PLBe1**	***BY4741/EV***	***zrc1*$^{\Delta}$/*EV***
10	1.05 ± 0.05a	0.76 ± 0.066b	0.05 ± 0.005c
11	0.82 ± 0.05a	0.31 ± 0.006b	0.03 ± 0.001c
12	0.49 ± 0.009a	0.21 ± 0.01b	0.02 ± 0.001c
Cobalt uptake (mg^{-1}g dry weight)			
Co (mM)	***cot1*$^{\Delta}$/ PLBe1**	***BY4741/EV***	***cot1*$^{\Delta}$/*EV***
2	0.41 ± 0.012a	0.04 ± 0.001b	BDL
3	0.27 ± 0.011a	0.05 ± 0.001b	BDL
4	0.17 ± 0.027a	0.013 ± 0.001b	BDL

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

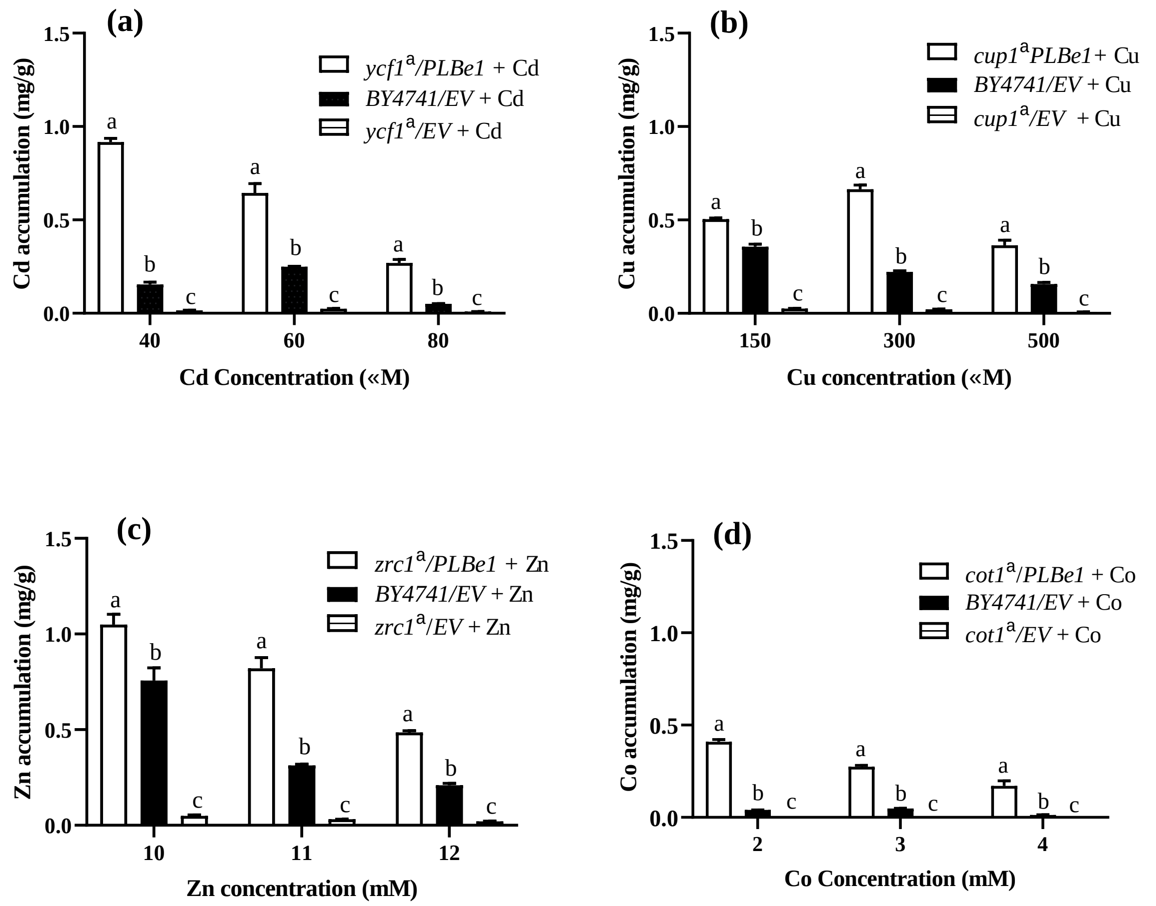


Figure 4.26: Accumulation of a) Cd^{2+}, b) Cu^{2+}, c) Zn^{2+} and d) Co^{2+} in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLBe1. Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3).

Accumulation of metals inside the yeast transformants was in a correlation with the tolerance level of particular yeast transformants towards these metals. Tolerance towards these metals might be due to increase in growth of yeast mutants in the presence of cDNA PLBe1, which further results into a good amount of metal accumulation.

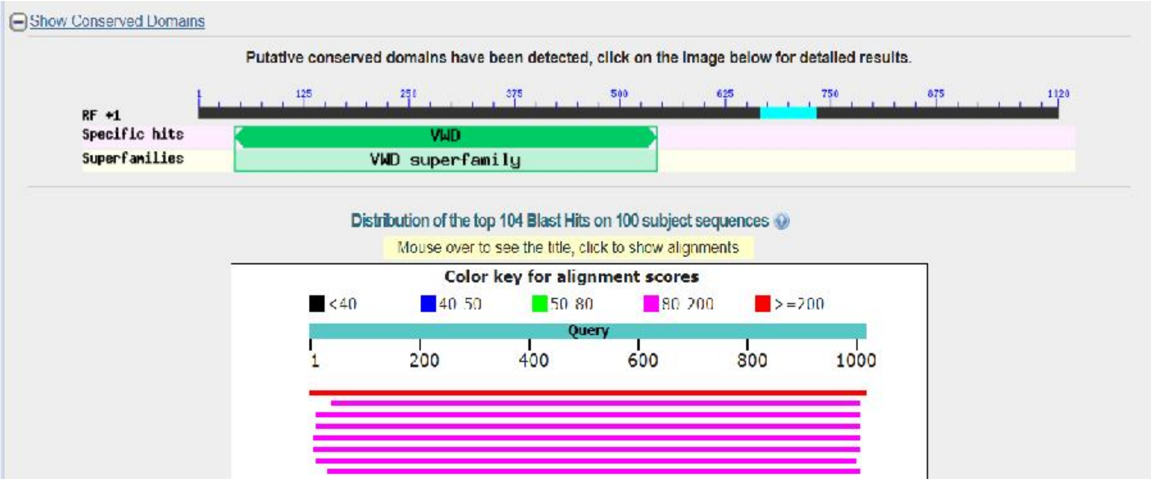
Ubiquitin fusion protein (UFP) is a eukaryotic protein playing an important role in selective degradation of aberrant proteins translated due to presence of potentially toxic metals (Pena *et al.* 2007; Forzani *et al.* 2002) or due to other stress conditions like osmostress, chemicals and heat shock (Leach *et al.* 2011; Peng *et al.* 2017; Adler *et al.* 2017). It had been reported that ubiquitin is available in the living beings either as polyubiquitin form or as a fusion protein at C-terminal where the post-translational modifications further led to formation of ubiquitin monomers (Finley *et al.* 1987; Redman and Rechsteiner, 1989). These fusion proteins are linked either to amino group of another ubiquitin molecule or to one of the three additional amino acids or to one of the two small protein moieties. In fungi, higher plants and animals, a type of UFP is found which consists of tandem head to tail repeats of the ubiquitin coding region with two or three additional amino acids at the terminal which are absent in mature proteins. Another type of UFP encodes an ubiquitin monomer followed by one or two non-related proteins and was first identified in yeast and human (Callis *et al.* 1990). It has been observed that these fusion proteins are highly conserved which itself suggest their significant role in functioning or synthesis of ubiquitin among eukaryotes (Schlesinger and Bond, 1987; Callis and Vierstra, 1989). Initial translation products of both types of UFPs must be precisely processed to release C-terminal carboxyl group to produce functional ubiquitin protein (Hershko *et al.* 1981) and the enzyme that cleave ubiquitin moiety from other linked proteins have been identified in yeast, mammals as well as in higher plants (Callis *et al.* 1990). Presence of diglycine residue at N-terminal end and lysine residues at 48 and 63 positions in UFPs assure the proteolytic processing of ubiquitin protein moiety and formation of polyubiquitin chains after posttranslational modifications, hence support these ubiquitin moieties released from UFP cleavage as potential cofactors for ubiquitin proteosome system (UPS) (Walters *et al.* 2002; Pickart and Fushman, 2004; Sá-Moura *et al.* 2013). UPS is a network present in eukaryotes that helps in degradation of ubiquitin targeted aberrant proteins generated mainly due to stress conditions such as metal toxicity in eukaryotes (Finley *et al.* 1987; Amerik *et al.* 2005; Ravid and Hochstrasser, 2008). Similarly, biological functions of extension protein moiety of UFP are partially elucidated and it has been observed that both of these extension proteins co-sediments with ribosomes, which indicate that they are constituents of mature active ribosomes (Finley *et al.* 1989). It has been observed that 52 aa extension protein from UFP co-sediment with 60S subunit of 80S ribosomes and 76 aa extension protein with 40S subunit of 80S ribosomes except in *Dictyostelium* where it was identified as a 40S protein (Müller-Taubenberger *et al.* 1989). Similarly, in mammals, 80 aa protein moiety has been localized on 40S ribosomal subunit as a ribosomal protein (Redman and Rechsteiner, 1989). Various studies have observed that these UFPs play significant role in providing tolerance towards potentially toxic metals present in an environment and are considered as stress-associated proteins. Woo *et al.* (2009) observed that ubiquitin encoding cDNA *OjaUB* showed elevated expression level in order to remove numerous aberrant proteins synthesized after exposure to different concentrations of Cd, Cu, Zn, Ni, Cr and Ag. Similarly, Leach *et al.* (2011) reported that ubiquitination plays an important role in growth, stress responses and metabolic adaptation of *Candida albicans*. Proteins with Zn-finger are also among the potential members of stress-associated proteins in both plants (Mukhopadhyay *et al.* 2004) and animals (Linnen *et al.* 1993; Evans *et al.* 2004; Huang *et al.* 2004; Heyninck and Beyaert, 2005; Hishiya *et al.* 2006). Thus, these findings interpret that various metals used in present study induced the expression of UFP encoded by cDNA PLBe1 in response to degradation of numerous proteins damaged due to metal toxicity by ubiquitination process of proteasome - dependent pathway as observed in a study by Woo *et al.* (2009). Therefore, this study contributes UFP as a prospective member of metal tolerant proteins in response to metal contamination in soil environment.

4.3.2. PLCe10/ von Willebrand factor type D Domain

One of the Cd tolerant transformants, *ycf1*$^{\Delta}$/PLCe10 screened from largest sized cDNA library C, showed the presence of full length ORF during BLAST analysis. As cDNA PLCe10 has shown tolerance towards toxic concentration of Cd thus selected for further gene characterization with respect to sequence analysis and tolerance potential towards other metals used in the study.

4.3.2.1. Sequence analysis

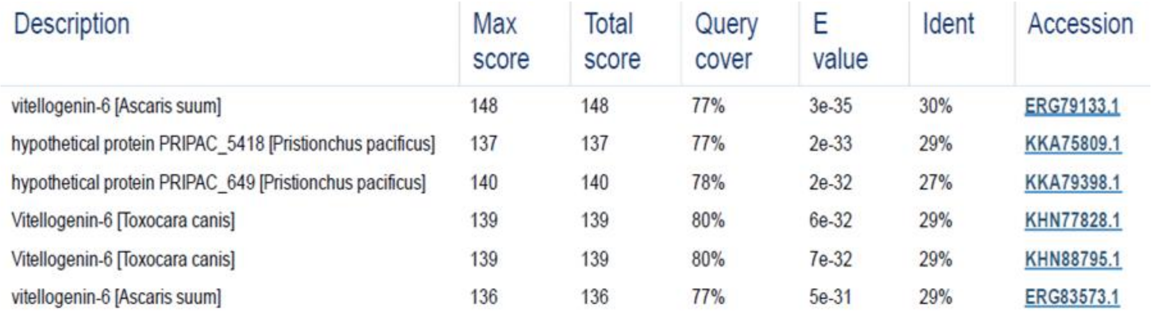
The BLASTX analysis with the deduced nucleotide sequences of PLCe10 revealed 30% identity (50% similarity) to von Willebrand factor type D (VWD) of vitellogenin-6 of nematode *Ascaris suum* with the open reading frame of 1017 bp, encoding 338 amino acids (Figure 4.27 & 4.28). The predicted molecular mass of peptide was 38.96 kDa (average mass) with pI of 4.77. The deduced peptide showed homology with amino acid sequence of *A. suum* from 1428 to 1719 amino acids denoted C-terminal of the Vg-6 which is an egg yolk protein. It had been reported that in *Paracylopina nana*, C-terminal region of Vg1 from amino acid sequence 1571-1739 and the Vg2 i.e., from 1481 to 1658 amino acids codes for von Willebrand factor type D domain (VWD).



Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 104 Blast Hits on 100 subject sequences

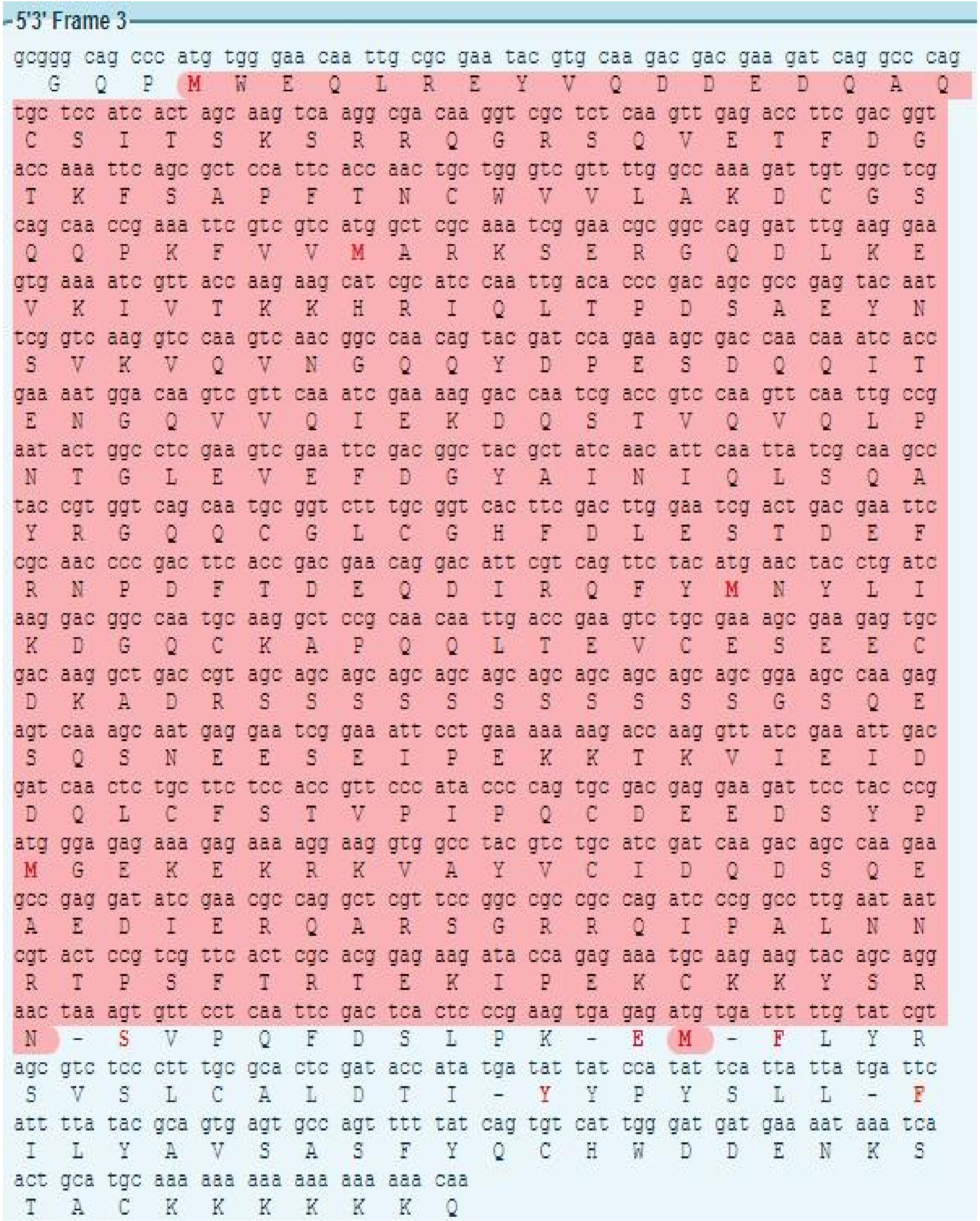
Mouse over to see the title, click to show alignments



Description	Max score	Total score	Query cover	E value	Ident	Accession
vitellogenin-6 [Ascaris suum]	148	148	77%	3e-35	30%	ERG79133.1
hypothetical protein PRIPAC_5418 [Pristionchus pacificus]	137	137	77%	2e-33	29%	KKA75809.1
hypothetical protein PRIPAC_649 [Pristionchus pacificus]	140	140	78%	2e-32	27%	KKA79398.1
Vitellogenin-6 [Toxocara canis]	139	139	80%	6e-32	29%	KHN77828.1
Vitellogenin-6 [Toxocara canis]	139	139	80%	7e-32	29%	KHN88795.1
vitellogenin-6 [Ascaris suum]	136	136	77%	5e-31	29%	ERG83573.1

Figure 4.27: Sequence homology of cDNA PLCe10 showing revealed 30% identity (50% positives) to von Willebrand factor type D (VWD) of vitellogenin-6 of nematode *Ascaris suum*

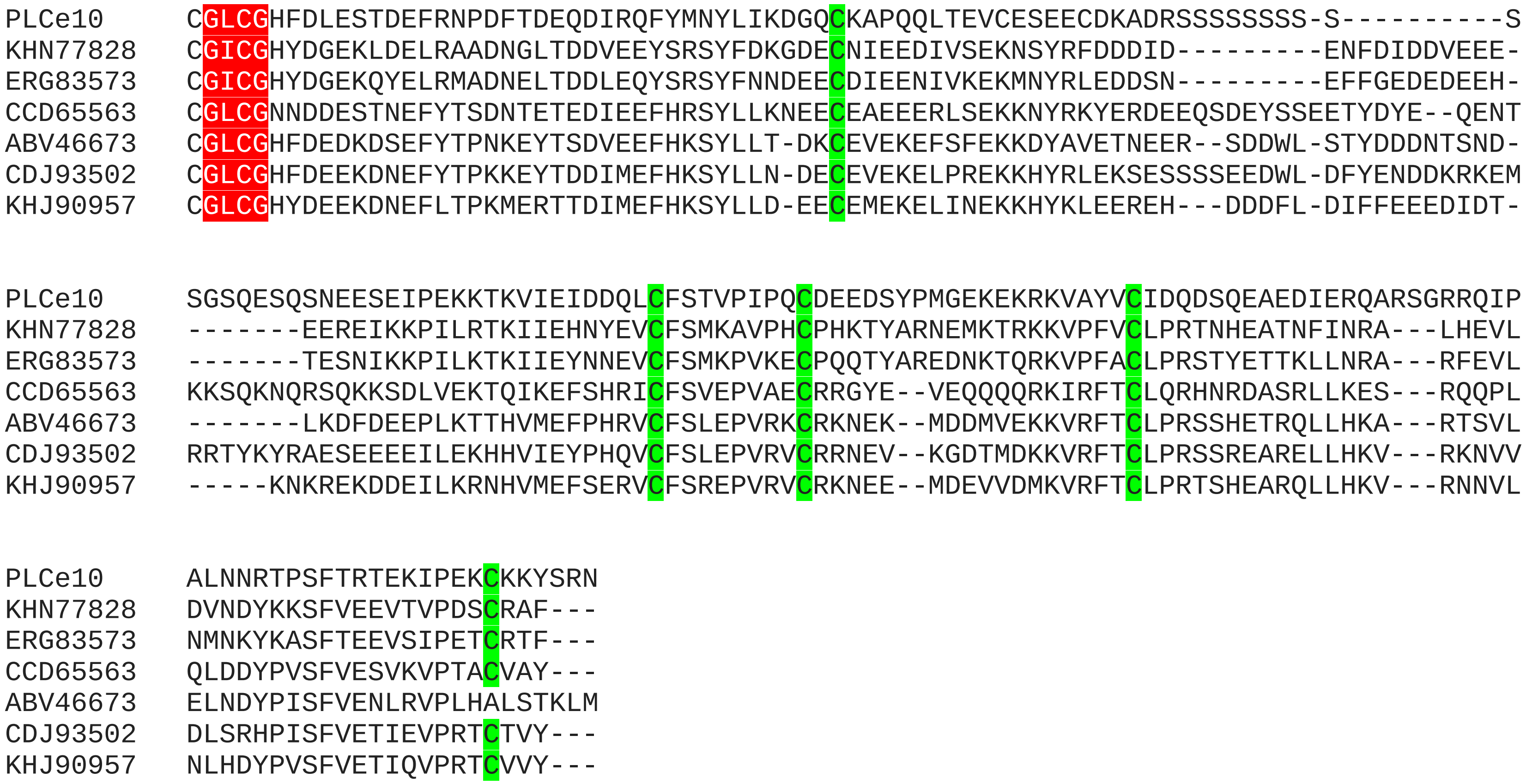
Hwang *et al.* (2010) reported that the BLAST analysis with the inferred amino acid sequences of *Paracyclopina nana* Vg1 and Vg2 had shown that in amino acid residues strongly matched with the intertidal copepod Vg1 (35% identities/56% positives) and *Lepeophtheirus salmonis* Vg2 (31% identities/52% positives), respectively. Multiple sequence alignment of homologous sequences showed that the 1017 bp gene belongs to protein family smart00216 which is also named as VWD (Figure 4.29). The deduced polypeptide identified as VWD like protein, contained characteristic phosphorylation sites at 32 amino acid sequences and 14 potential motifs for O-linked glycosylation sites



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

Figure 4.28: Translation of nucleotide sequence of cDNA PLCe10. Red highlighted region is an ORF of the nucleotide sequence encoding for von Willebrand factor type D domain.

were also found while no N-linked glycosylation sites were observed (Figure 4.30). Finn (2007) observed that before exporting Vg from hepatocytes, phosphorylation of precursor protein takes place in the Golgi bodies and the presence of phosphorylated serine residues in the Vg polypeptide are assumed to avert premature degradation and assist the solubility of Vg molecules in the blood. Various phosphorylation sites within the predicted VWD like polypeptide sequence indicate that the wide majority of serine residues present in the polypeptide of VWD like domain may be phosphorylated along with threonine and tyrosine. In addition to this, comparison of VWD like polypeptide to other vitellogenin showed a high degree of conservation; contain two RXXR consensus cleavage sites each at C-terminal and N-terminal of the polypeptide, a well characterized GL/ICG motif and conserved polycysteine residues at C-terminal (Figure 4.30). Tufail *et al.* (2010) characterized Vg from insect *Nilaparvata lugens* and found high degree of conservation for the motif GL/ICG. In a study, Shu *et al.* (2009) aligned Vg of *Spodoptera litura* with Vg of other species and characterized it by presence of highly conserved cleavage signal RXXR at N terminal, in the C-terminal GL/ICG motif followed by cysteine residues. Salmela *et al.* (2016) characterised three Vg like proteins and found that all of them share structural similarities with Vg but each of them also has unique gene/protein architecture and distinct pattern of positive selection, which are indicative of divergence of functions. Presence of hexameric polyadenylation signal (PAS) AATAAA at 12 bp upstream the poly-A tail in ORF PLCe10, also characterized this cDNA as ORF with a potential for coding a protein as described by Beaudoing *et al.* (2000).

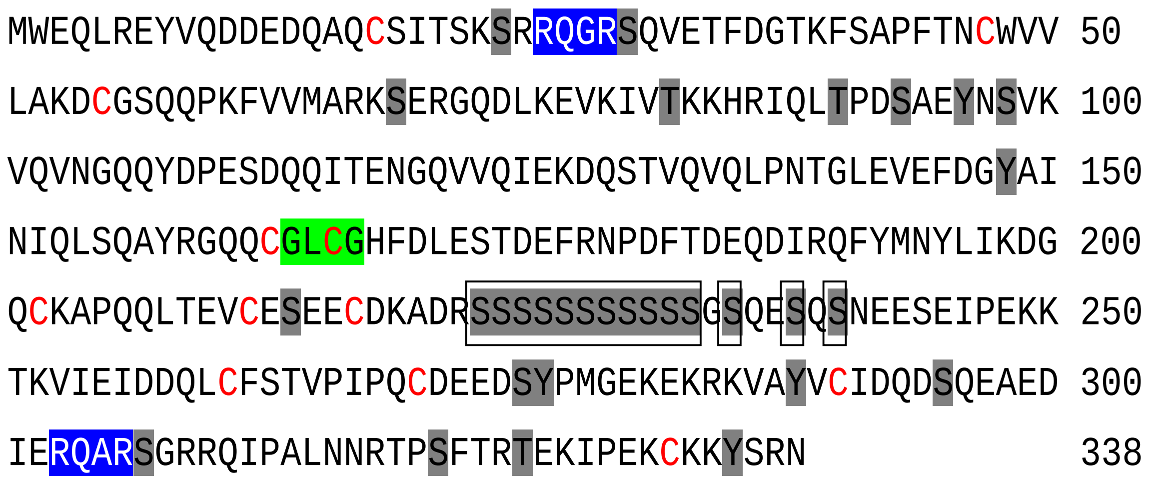


```
PLCe10     CGLCGHFDLESTDEFRNPDFTDEQDIRQFYMNYLIKDGQCKAPQQLTEVCESEECDKADRSSSSSSSS-S----------S
KHN77828   CGICGHYDGEKLDELRAADNGLTDDVEEYSRSYFDKGDECNIEEDIVSEKNSYRFDDDID---------ENFDIDDVEEE-
ERG83573   CGICGHYDGEKQYELRMADNELTDDLEQYSRSYFNNDEECDIEENIVKEKMNYRLEDDSN---------EFFGEDEDEEH-
CCD65563   CGLCGNNDDESTNEFYTSDNTETEDIEEFHRSYLLKNEECEAEEERLSEKKNYRKYERDEEQSDEYSSEETYDYE--QENT
ABV46673   CGLCGHFDEDKDSEFYTPNKEYTSDVEEFHKSYLLT-DKCEVEKEFSFEKKDYAVETNEER--SDDWL-STYDDDNTSND-
CDJ93502   CGLCGHFDEEKDNEFYTPKKEYTDDIMEFHKSYLLN-DECEVEKELPREKKHYRLEKSESSSSEEDWL-DFYENDDKRKEM
KHJ90957   CGLCGHYDEEKDNEFLTPKMERTTDIMEFHKSYLLD-EECEMEKELINEKKHYKLEEREH---DDDFL-DIFFEEEDIDT-

PLCe10     SGSQESQSNEESEIPEKKTKVIEIDDQLCFSTVPIPQCDEEDSYPMGEKEKRKVAYVCIDQDSQEAEDIERQARSGRRQIP
KHN77828   -------EEREIKKPILRTKIIEHNYEVCFSMKAVPHCPHKTYARNEMKTRKKVPFVCLPRTNHEATNFINRA---LHEVL
ERG83573   -------TESNIKKPILKTKIIEYNNEVCFSMKPVKECPQQTYAREDNKTQRKVPFACLPRSTYETTKLLNRA---RFEVL
CCD65563   KKSQKNQRSQKKSDLVEKTQIKEFSHRICFSVEPVAECRRGYE--VEQQQQRKIRFTCLQRHNRDASRLLKES---RQQPL
ABV46673   -------LKDFDEEPLKTTHVMEFPHRVCFSLEPVRKCRKNEK--MDDMVEKKVRFTCLPRSSHETRQLLHKA---RTSVL
CDJ93502   RRTYKYRAESEEEEILEKHHVIEYPHQVCFSLEPVRVCRRNEV--KGDTMDKKVRFTCLPRSSREARELLHKV---RKNVV
KHJ90957   -----KNKREKDDEILKRNHVMEFSERVCFSREPVRVCRKNEE--MDEVVDMKVRFTCLPRTSHEARQLLHKV---RNNVL

PLCe10     ALNNRTPSFTRTEKIPEKCKKYSRN
KHN77828   DVNDYKKSFVEEVTVPDSCRAF---
ERG83573   NMNKYKASFTEEVSIPETCRTF---
CCD65563   QLDDYPVSFVESVKVPTACVAY---
ABV46673   ELNDYPISFVENLRVPLHALSTKLM
CDJ93502   DLSRHPISFVETIEVPRTCTVY---
KHJ90957   NLHDYPVSFVETIQVPRTCVVY---
```

Figure 4.29: Alignment of amino acid sequences of different organisms showing highly conserved GL/ICG motif (red highlighted) to the C-terminal ends. Accession numbers are in parenthesis *Toxocara canis* (KHN77828), *Ascaris suum* (ERG83573), *Caenorhabditis elegans* (CCD65563), *Angiostrongylus vasorum* (ABV46673), *Haemonchus contortus* (CDJ93502) and *Oesophagostomum dentatum* (KHJ90957). Cysteine residues are highlighted with green color.



```
MWEQLREYVQDDEDQAQCSITSKSRRQGRSQVETFDGTKFSAPFTNCWVV  50
LAKDCGSQQPKFVVMARKSERGQDLKEVKIVTKKHRIQLTPDSAEYNSVK 100
VQVNGQQYDPESDQQITENGQVVQIEKDQSTVQVQLPNTGLEVEFDGYAI 150
NIQLSQAYRGQQCGLCGHFDLESTDEFRNPDFTDEQDIRQFYMNYLIKDG 200
QCKAPQQLTEVCESEECDKADRSSSSSSSSSSSGSQESQSNEESEIPEKK 250
TKVIEIDDQLCFSTVPIPQCDEEDSYPMGEKEKRKVAYVCIDQDSQEAED 300
IERQARSGRRQIPALNNRTPSFTRTEKIPEKCKKYSRN             338
```

Figure 4.30: Deduced amino acid sequence of VWD protein: Highly conserved GL/ICG motif (highlighted in green) to the C-terminal end. Cysteine residues are given in bold red color while RXXR cleavage sites are highlighted in blue box with white font. Bold and boxed amino acids are sites for O-linked glycosylation and grey highlighted are phosphorylation sites.

Secondary structure of polypeptide sequence showed the localized organization among the amino acids due to various assumed different spatial arrangements (Figure 4.31). Phylogenetic analysis of cDNA PLCe10 with the similar genes corresponding to various taxon have clearly clustered cDNA PLCe10 with nematodes, which support the fact that this gene might have belonged to phylum Nematoda (Figure 4.32). Hence, from bioinformatics analysis, we assumed that the cDNA sequence PLCe10 obtained in this study encode for VWD like domain of putative protein Vg. Jubeaux *et al.* (2012) demonstrated the benefit of moderate phylogenetic conservation of Vg like proteins such as Vg proteotypic peptides, to directly exploit them in developing biomarkers to identify more or less close phylogenetically related species. The cDNA sequence was deposited at NCBI database with the accession number MF770726.

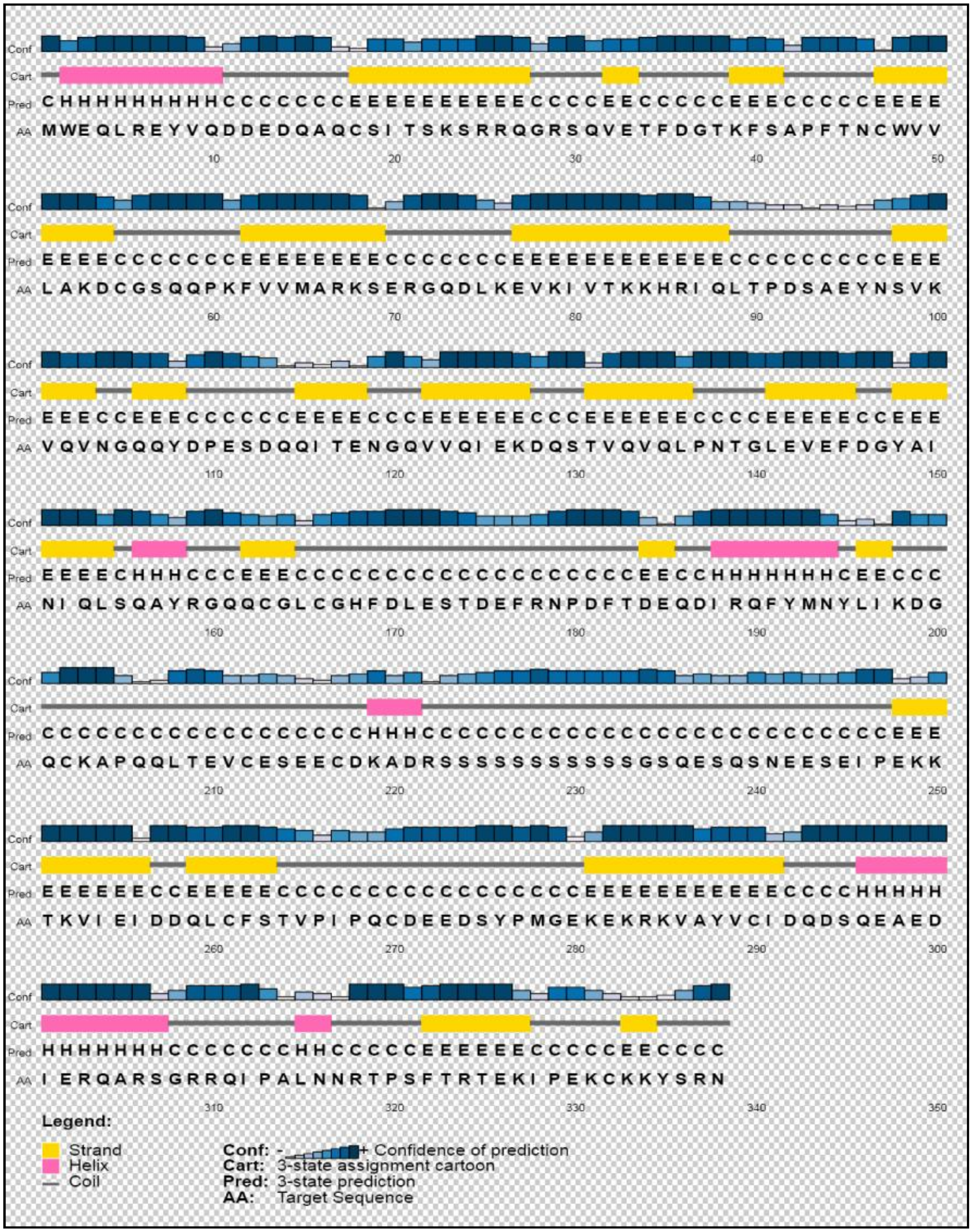


Figure 4.31: Secondary structure of deduced amino acid sequence encoded by metatranscriptomic cDNA PLCe10 showing different possible spatial arrangements among the amino acids.

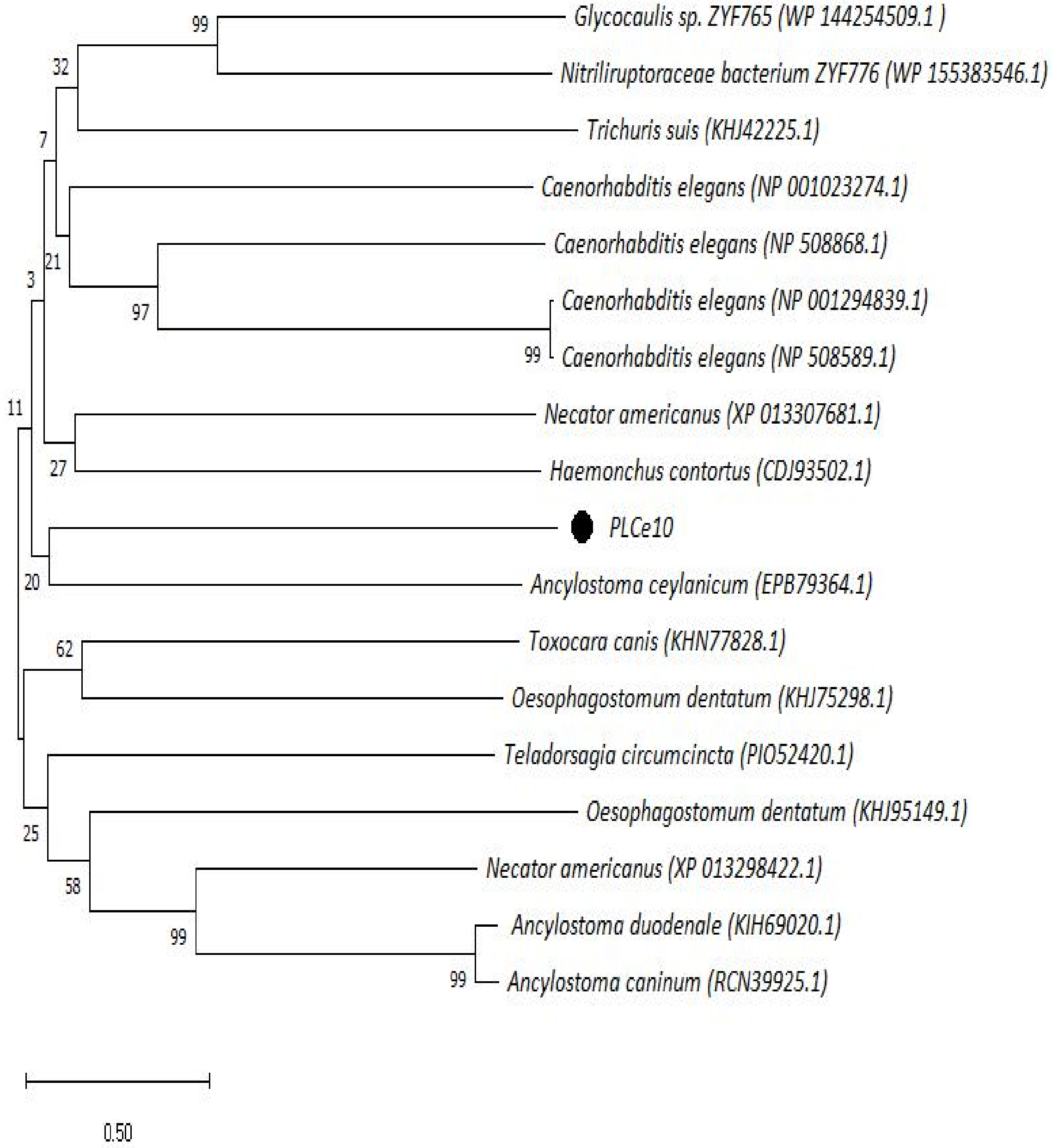


Figure 4.32: Phylogenetic relation of the transcript PLCe10 with other taxa corresponding to similar proteins constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side with GenBank accession numbers in parentheses. Branch lengths are proportional to evolutionary distances.

4.3.2.2. qPCR analysis

Induction of cDNA PLCe10 encoding VWD like protein in presence of Cd was studied to know the expression level in stressful condition (Table 4.10). Induction levels of PLCe10 increased significantly due to Cd stress with maximum mRNA accumulation observed at 60 µM of $CdSO_4$, which was 32 fold higher than control while minimum (approximately 10 times) at 40 µM of $CdSO_4$. Further, increase in Cd concentration increased the level of Cd toxicity hence resulted into decline for RNA expressed i.e., 17 times at 80 µM of $CdSO_4$ (Figure 4.33).

Table 4.10: Effect of different cadmium concentrations on the relative expression of cDNA PLCe10

Cd Concentration (µM)	Fold expression
40	9.88 ± 2.8a
60	32.60 ± 2.7b
80	17.18 ± 1.9c

Values sharing a common letter within the column are not significant at $P<0.05$ (Mean ± SD n=3)

Previous studies have suggested that PTMs regulate Vg mRNA expression levels in various vertebrates such as cope pod *Tigriopus japonicus* and mosquito fish *Gambusia affinis* (Lee *et al.* 2008: Huang *et al.* 2014). It was reported that the change in expression level of Vg mRNA is a strong indication of endocrine disrupting activities, which are generally due to endocrine disrupting chemicals such as PTMs (Martín-Díaz *et al.* 2008). It has been observed that PTMs such as Zn, Cd and Pb induce the Vg mRNA expression in mosquito fish *Gambusia affinis,* which suggest these metals have estrogenic activity in

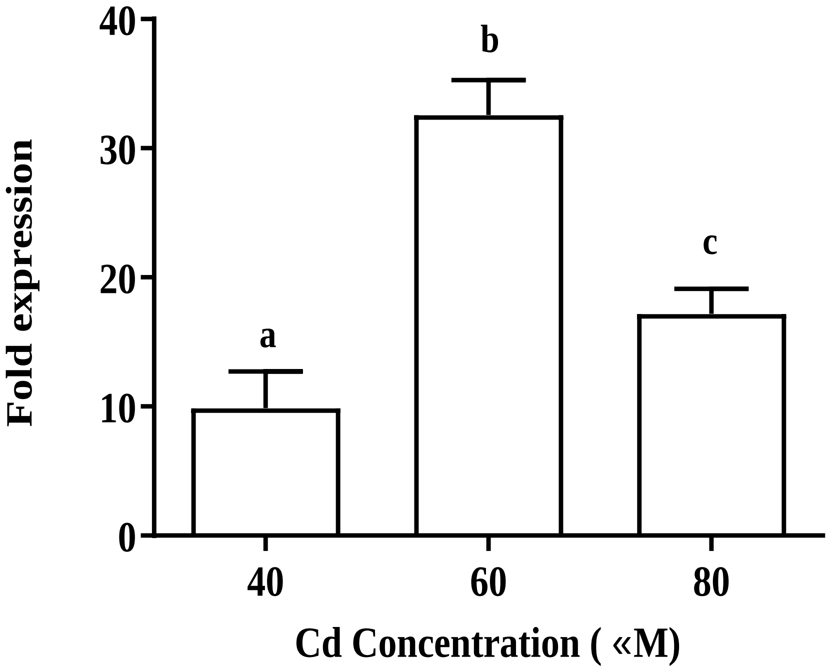


Figure 4.33: Fold increase in expression levels of cDNA PLCe10 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCe10 without metal treatment). Bars sharing a common letter are not significant at P<0.05 (n=3).Error bars are ±SD.

mosquito fish (Huang *et al.* 2014). Significant induction of Vg mRNA expression in the present study is further supported by previous similar findings obtained with primary hepatocytes of *Gadus morhua* (Atlantic cod), exposed to Cd (Søfteland *et al.* 2010) and mature intermoult female *Carcinu smaenas* exposed to Pb and Zn (Martín-Díaz *et al.* 2008).

4.3.2.3. Multi-metal tolerance of PLCe10

The cDNA PLCe10 was tested for its tolerance potential of cDNA towards potentially toxic metals (PTMs) such as Cd, Cu, Zn and Co by over-expressing the cDNA in various hypersensitive yeast mutants of *S. cerevisiae* i.e., *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$. Metal sensitive yeast mutants transformed with PLCe10 were spotted on metal amended media and allowed to grow in optimum conditions. The cDNA PLCe10 has shown significant tolerance towards all the four metals used in the study in contrast to the mutants as well as wild strains carrying empty expression vector (Figure 4.34). As shown in Figure 4.34a, Cd sensitive yeast mutant *ycf1*$^{\Delta}$ has shown prolific growth at 40 µM $CdSO_4$ in contrast to the mutant strain with empty vector. Similarly, Cu sensitive mutant *cup1*$^{\Delta}$ with empty pFL61 was sensitive to 150 µM $CuSO_4$ while *cup1*$^{\Delta}$ cells carrying plasmid PLCe10 (*cup1*$^{\Delta}$/PLCe10) were showing tolerance to a recognizable level in comparison to *cup1*^{s}pFL61 which carries single copy of CUP1 gene only (Figure 4.34b). Likewise, *zrc1*$^{\Delta}$ cells carrying PLCe10 plasmid were tolerant at 10 mM $ZnCl_2$ whereas this concentration was toxic for *zrc1*$^{\Delta}$pFL61 as shown in Figure 4.34c. Similarly, 2 mM $CoCl_2$ inhibited the growth of *cot1*$^{\Delta}$pFL61 where as *cot1*$^{\Delta}$ cells carrying plasmid PLCe10 (*cot1*$^{\Delta}$/PLCe10) were showing tolerance to Co as in case of wild strain BY4741pFL61 (Figure 4.34d). For drop assay, a general decrease in growth was observed for all the yeast strains in metal amended media. Tolerance potential of cDNA PLCe10 was further authenticated by performing growth assays for different metals at various concentrations. The growth assays of different transformants showed positive results on different range of concentration of metals (Table 4.11). The Cd tolerant transformant *ycf1*$^{\Delta}$/PLCe10 showed significantly higher growth in all the concentrations compared to the wild type strain transformed with empty vector. With increase in $CdSO_4$ concentration, reduction in the growth was observed both in *ycf1*$^{\Delta}$/PLCe10 and wild type strain (Figure 4.35a).

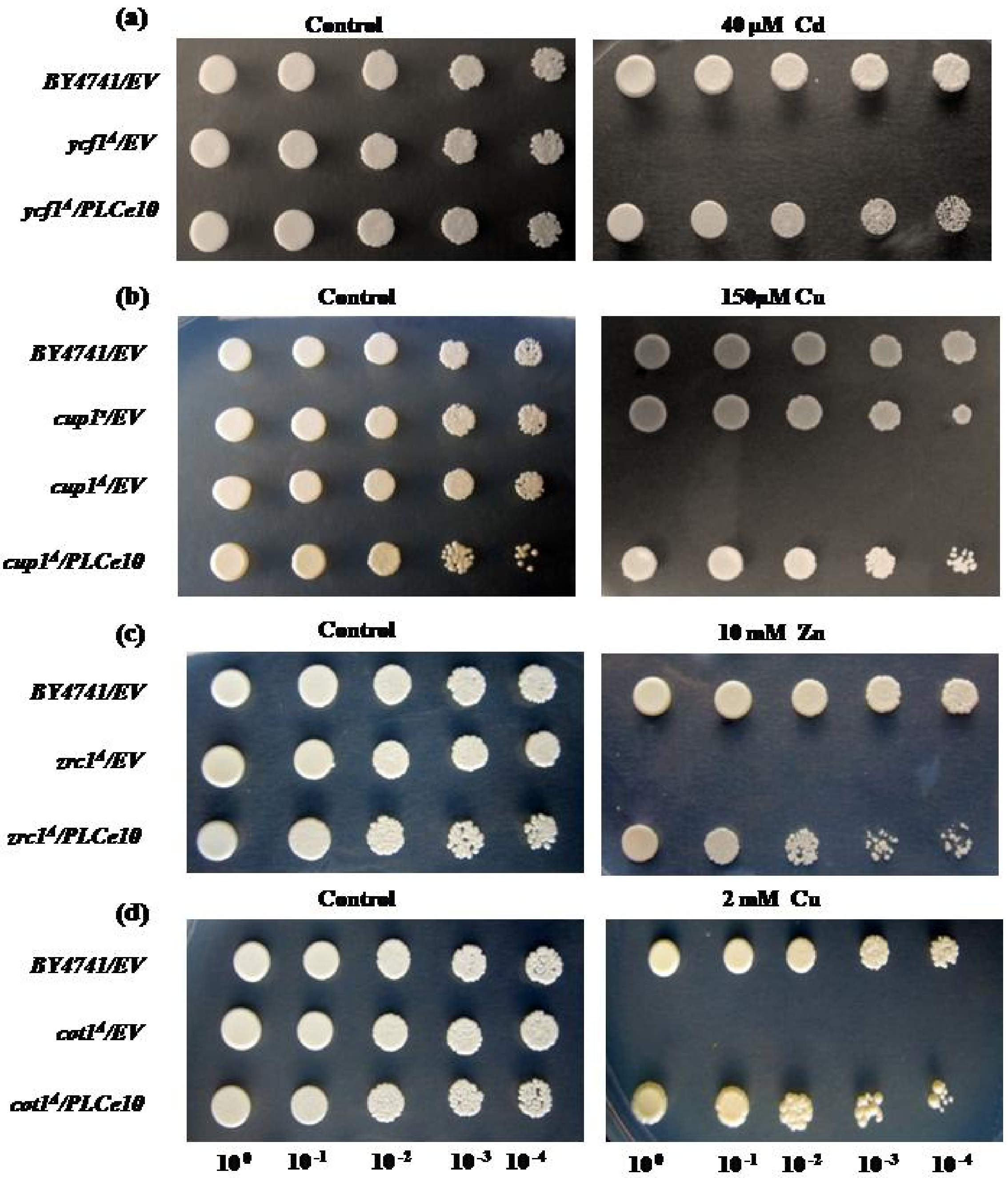


Figure 4.34: Drop assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLCe10 at various concentrations a) $CdSO_4$ 40 µM, b) $CuSO_4$ 150 µM, c) $ZnCl_2$ 10 mM and d) $CoCl_2$ 2 mM. Wild strain BY4741 and *cup1*^{s} were used as control and EV denote to empty vector pFL61.

When screened for Cu, it was observed that *cup1*$^{\Delta}$/PLCe10 cells were able to grow up to 500 µM compared to mutant strain *cup1*$^{\Delta}$pFL61 and isotopic strain *cup1*^{s}pFL61 (Figure 4.35b). In context to Cu, it was reported that estrogen-induced vitellogenin was tolerant to oxidation reactions induced by Cu and chelate low concentrations of Cu ion. Antioxidant property of vitellogenin protected the copper-stimulated oxidation of very low-density lipoprotein (VLDL). Vitellogenin believed to be served as transition metals binding lipoprotein, which results into extensive depressing of free radical reactions in the oocytes (Ando and Yanagida, 1999). While studying tolerance for Zn metal tolerance, it was found that transformant has shown more tolerance towards Zn (10 mM to 12 mM $ZnCl_2$) in comparison to both wild strain BY4741pFL61 and mutant strain *zrc1*$^{\Delta}$pFL61 (Figure 4.35c). Vitellogenin mediates the transfer of Zn from the liver to the maturing oocytes in the turkey hen, ultimately resulting in Zn deposition into yolk of the newly formed egg (Richards, 1989). It was reported that Vg is a macromolecule phospholipid glycoprotein containing Ca and Zn ligands (Wallace, 1985; Montorzi *et al.* 1994; Denslow *et al.* 1999). Furthermore, it was reported that among honeybee workers and queens, Vg plays role as a Zn carrier that protects them from oxidative stress (Amdam *et al.* 2004; Seehuus *et al.* 2006) and contributes to queen longevity (Corona *et al.* 2007). Similarly in cobalt, the gene provided tolerance to mutant yeast from range of 2 mM to 4 mM of $CoCl_2$. Increase in tolerance was observed with the increase in range of $CoCl_2$ concentrations (Figure 4.35d). Henceforth, this study showed that cDNA PLCe10 was solely responsible for providing metal tolerance phenotype to hypersensitive yeast mutants towards all the four metals.

Table 4.11: Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLCe10 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant strain with empty vector pFL61 (EV).

	cDNA PLCe10		
	O.D. 600		
Cadmium (µM)	***ycf1***$^{\Delta}$***/EV***	***BY4741/EV***	***ycf1***$^{\Delta}$***/* PLCe10**
40	0.09 ± 0.01a	1.09 ± 0.08b	1.74 ± 0.09c
60	0.06 ± 0.01a	0.38 ± 0.06b	1.12 ± 0.07c
80	0.04 ± 0.006a	0.39 ± 0.18b	0.65 ± 0.10b
	O.D. 600		
Copper (µM)	***cup1***$^{\Delta}$***/EV***	***cup1***^{s}***/EV***	***cup1***$^{\Delta}$***/* PLCe10**
150	0.24 ± 0.032a	0.56 ± 0.01b	3.60 ± 0.32c
300	0.07 ± 0.007a	0.19 ± 0.03b	2.39 ± 0.35c
500	0.05 ± 0.001a	0.04 ± 0.001b	1.50 ± 0.11c
	O.D. 600		
Zinc (mM)	***zrc1***$^{\Delta}$***/EV***	***BY4741/EV***	***zrc1***$^{\Delta}$***/* PLCe10**
10	0.34 ± 0.07a	1.38 ± 0.19b	1.45 ± 0.06c
11	0.14 ± 0.03a	0.96 ± 0.03b	1.1 ± 0.11c
12	0.084 ± 0.014a	0.48 ± 0.09b	0.9 ± 0.04c
	O.D. 600		
Cobalt (mM)	***cot1***$^{\Delta}$***/EV***	***BY4741/EV***	***cot1***$^{\Delta}$***/* PLCe10**
2	0.26 ± 0.07a	1.49 ± 0.19b	2.67 ± 0.18c
3	0.07 ± 0.00a	0.97 ± 0.11b	2.38 ± 0.34c
4	0.05 ± 0.009a	0.14 ± 0.02b	2.15 ± 0.04c

Values sharing common letters for a particular metal concentration between samples are not significantly different at $P<0.05$ (O.D. 600 values are mean ± SD, n=3).

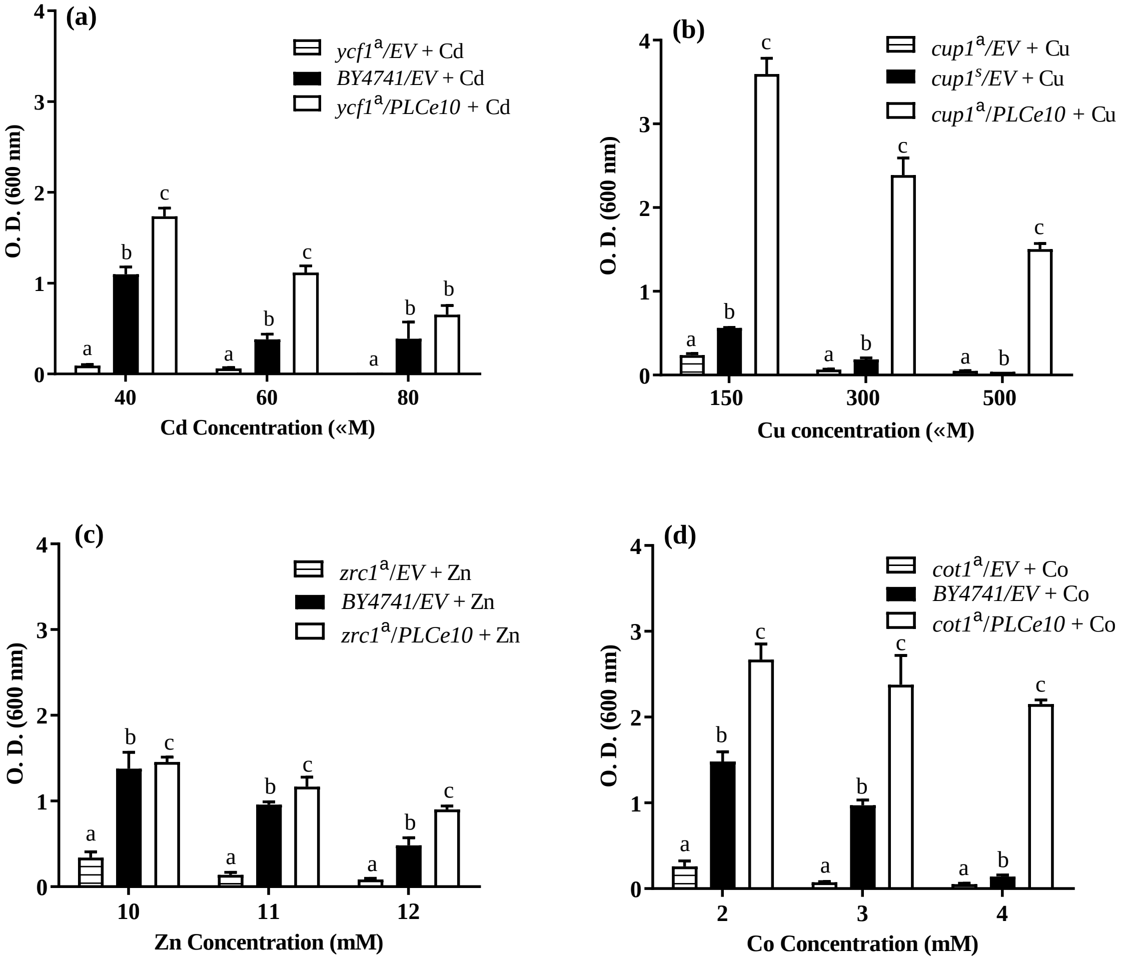


Figure 4.35: Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLCe10 in SD-Ura medium supplemented with (a) $CdSO_4$ 40-80 µM, (b) $CuSO_4$ 150-500 µM, (c) $ZnCl_2$ 10-12 mM and (d) $CoCl_2$ 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at P<0.05 (n=3).

4.3.2.4. Yeast metal uptake of PLCe10

Hypersensitive yeast mutants *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying tolerant cDNA PLCe10 was further tested for the level of accumulation of respective metal used in the study. To achieve this, specific yeast transformant cultures were allowed to grow for 48 hrs and acid digested product of their cell pellet (1 gm) was estimated for quantity of metal available by ICP-MS (Element XR, Thermo Fisher Scientific, Germany). This study revealed that cDNA PLCe10 accumulate a higher range of metals (Cd, Cu, Zn and Co) in comparison to wild strain BY4741 transformed with empty vector pFL61 (Table 4.12). The *ycf1*$^{\Delta}$/PLCe10 cells accumulated different levels of Cd when grown in presence of various $CdSO_4$ concentrations. It has been observed that the amount of Cd metal accumulated increased up to 60 μM of $CdSO_4$ concentration and decreased at 80 μM $CdSO_4$ (Figure 4.36a). For Cu, the level of metal accumulation increased with increase in concentration as shown in Figure 4.36b, while for Zn and Co metal, the level of accumulation increases up to 11 mM of $ZnCl_2$ and 3 mM $CoCl_2$ respectively and decreases with further increase in Zn or Co concentration (Figure 4.36c & d). Such variation in order of tolerance and accumulation of heavy metals could potentially be an explanation for distinct metal-binding abilities of the deduced VWD like protein. In general, increased growth and metal accumulation was observed in mutant yeast cells carrying PLCe10 compared to the respective wild type strains. The increase in growth might be due to protection of PLCe10 to the yeast mutants under metal stress and the accumulation of metal in the yeast strains also might be due to increased growth.

It is well known that PTMs in an environment cause toxicity and significantly engendering oxidative stress inside the cells, thus inhibiting the growth and development of living organisms (Huang *et al.* 2012; Kodrík *et al.* 2015; Suganya *et al.* 2016). Potentially toxic metals as Cd and lead (Pb) cause various toxic effects, including reproductive toxicity at the cellular and molecular level (Scaps *et al.* 1998; Cervera *et. al.,* 2005; Shu *et. al.,* 2009). Previous studies suggested that PTMs had endocrine disrupting effects among different vertebrate and invertebrate species resulted into induction of Vg among aquatic animals on exposure to these PTMs (Cervera *et al.* 2006; Hwang *et al.* 2010; Annabi *et al.* 2012; Lee-Pow *et al.* 2016).

Table 4.12: Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCe10 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV).

cDNA PLCe10			
Cadmium uptake (mg^{-1}g dry weight)			
Cd (µM)	***ycf1*$^{\Delta}$/ PLCe10**	***BY4741/EV***	***ycf1*$^{\Delta}$*/EV***
40	0.19 ± 0.002a	0.15 ± 0.012b	0.015 ± 0.001c
60	0.28 ± 0.010a	0.25 ± 0.002b	0.025 ± 0.010c
80	0.065 ± 0.003a	0.05 ± 0.00b	0.009 ± 0.001c
Copper uptake (mg^{-1}g dry weight)			
Cu (µM)	***cup1*$^{\Delta}$/ PLCe10**	***BY4741/EV***	***cup1*$^{\Delta}$*/EV***
150	0.42 ± 0.016a	0.36 ± 0.013b	0.026 ± 0.001c
300	0.48 ± 0.017a	0.22 ± 0.005b	0.022 ± 0.001c
500	0.59 ± 0.010a	0.16 ± 0.009b	0.008 ± 0.001c
Zinc uptake (mg^{-1}g dry weight)			
Zn (mM)	***zrc1*$^{\Delta}$/ PLCe10**	***BY4741/EV***	***zrc1*$^{\Delta}$*/EV***
10	1.07 ± 0.02a	0.76 ± 0.066b	0.05 ± 0.005c
11	1.13 ± 0.14a	0.31 ± 0.006b	0.03 ± 0.001c
12	0.84 ± 0.02a	0.21 ± 0.01b	0.02 ± 0.001c
Cobalt uptake (mg^{-1}g dry weight)			
Co (mM)	***cot1*$^{\Delta}$/ PLCe10**	***BY4741/EV***	***cot1*$^{\Delta}$*/EV***
2	0.14 ± 0.009a	0.04 ± 0.001b	BDL
3	0.24 ± 0.011a	0.05 ± 0.001b	BDL
4	0.09 ± 0.005a	0.013 ± 0.001b	BDL

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

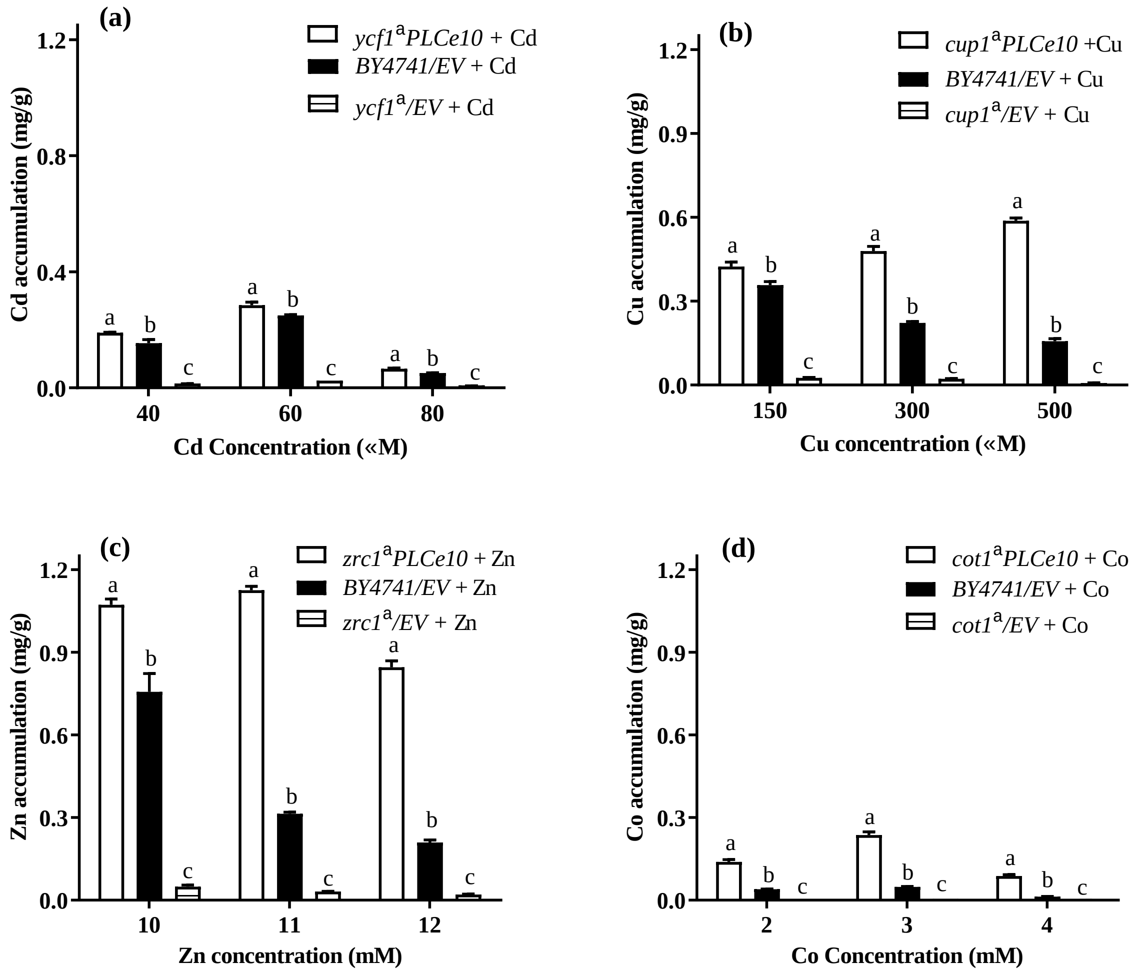


Figure 4.36: Accumulation of a) Cd^{2+}, b) Cu^{2+}, c) Zn^{2+} and d) Co^{2+} in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCe10 Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3).

Hwang et al. (2010) observed that even a short term exposure of PTMs induce the process of vitellogenesis among the oviparous vertebrates and invertebrates. Some studies showed that Vg protect the insects such as honey bees against the damage due to oxidative stress and reported that the survival rate of insects with elevated level of Vg protein is higher than those with low Vg phenotypes (Kodrik *et al.* 2015). Various studies demonstrated that PTMs such as Cd modulates Vg synthesis in invertebrates such as insects (Cervera *et al.* 2005, 2006), centipedes (Scaps *et al.* 1998), nematodes (Novillo *et al.* 2005) and crustaceans (Martin-Diaz *et al.* 2005; Lee *et al.* 2007; Hwang *et al.* 2009). In aquatic environments, Vg has been widely used as a biomarker to assess metal contamination and estrogenic contamination in various environments (Matozzo *et al.* 2008; Hwang *et al.* 2009). Potentially toxic metals as well as endocrine disrupting chemicals that mimic estrogens also induce the expression of Vg genes in male animals, particularly in fish, which significantly supports the use of Vg as a biomarker in aquatic environment. The study on *Paracyclopina nana* also provided the evidence in support of Vg as a potential indicator for monitoring and assessing the toxic effect of potentially toxic metals like Cd, Cu, and As (Hwang *et al.* 2009).

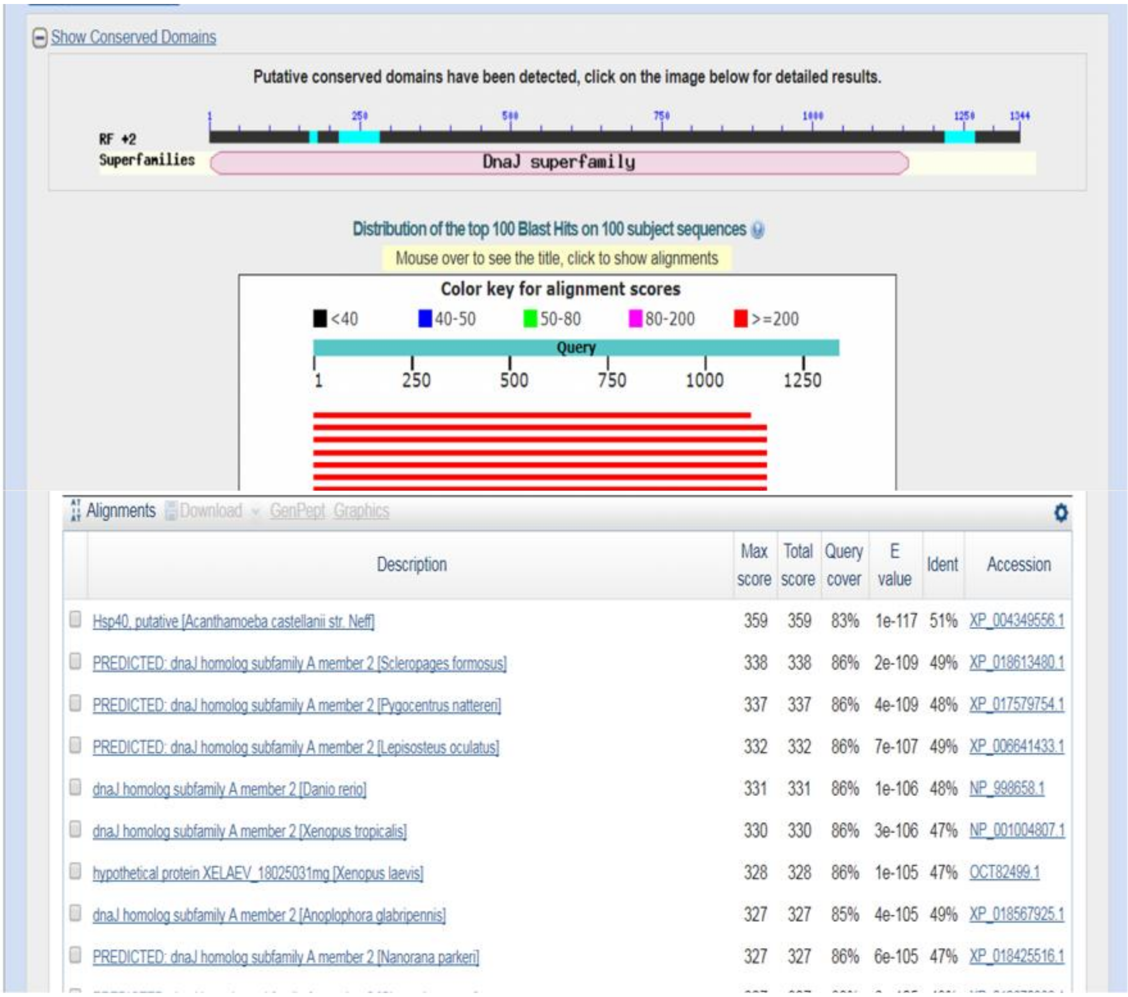
Generally, Vg contains three conserved domains, LPD N (also known as vitellogenin N or LLT) which is identified in the N-terminus of large lipid transfer protein (LLTP) members, domain of unknown function 1943 (DUF1943) and von Willebrand factor type D domain (VWD) which is located at the C-terminus and distributed over a wide range of proteins. In 2000, Lee *et al.* observed that the presence of GL/ICG motif followed by number of cysteine residues at conserved locations in the VWD was entitled as a general trait among various invertebrate and vertebrate species (Lee *et al.* 2000). Bioinformatics analysis of cDNA PLCe10 also reported the presence of GL/ICG conserved domain in the deduced polypeptide. Consequently, there could be a possibility to gain benefit of this peptide motif conservation to develop biomarkers applicable for groups of phylogenetically related species. Hence, these findings advocated the metal tolerant cDNA PLCe10 encoding VWD like protein as a potential member of metal tolerant gene family of eukaryotic community which could be further exploited to increase the possibility of using such peptides as biomarker for metal contamination.

4.3.3. PLCc43/Heat shock protein type I like protein

Screening of library C observed another Cd tolerant transformant carrying a cDNA with full length ORF, showing partial homology to the native Hsp40 Type I proteins. This cDNA was characterized and tested for its tolerance to other metals.

4.3.3.1. Sequence analysis of PLCc43

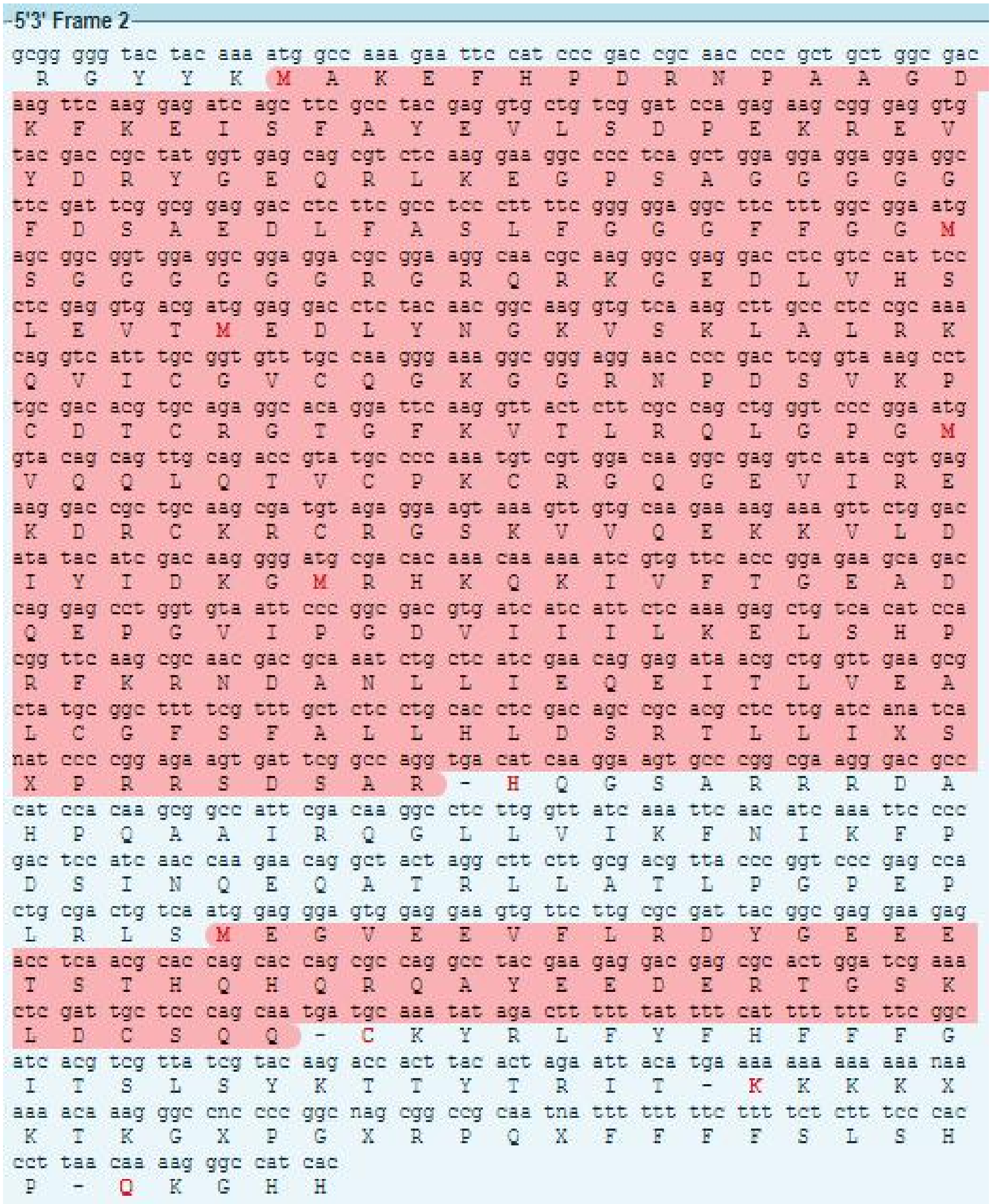
Sequence analysis of cDNA PLCc43 revealed that it consisted of 1269 bp cDNA which showed 51% homology (65% positives) with Hsp40 Type I protein of *Acanthameoba castellanii* (Figure 4.37). cDNA PLCc43 consisted of 855 bp ORF encoding a putative protein of 284 amino acids with theoretical molecular mass of 31.2 KDa and pI of 8.99 (Figure 4.38). Generally, Hsp40 proteins are categorised based on presence of different combinations of three domains, namely: J domain that helps in stimulating the ATPase activity of Hsp70; long stretch of G/F amino acid rich region and cysteine rich region that contains zinc-binding CXXCXGXG motif repeats and a carboxy-terminal substrate-binding domain (Caplan and Douglas 1991; Qiu *et al.* 2006). Putative PLCc43p may display the fundamental structural features of Hsp40 Type I proteins i.e., a partial J domain with the conserved tri-peptide HPD (Histidine, Proline and Aspartic acid) motif, a stretch of Glycine (G) Phenylalanine (P) rich region and four zinc finger domains although the final G amino acid in the fourth stretch CXXCXGXG was missing (Figure



	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Hsp40, putative [Acanthamoeba castellanii str. Neff]	359	359	83%	1e-117	51%	XP_004349556.1
☐	PREDICTED: dnaJ homolog subfamily A member 2 [Scleropages formosus]	338	338	86%	2e-109	49%	XP_018613480.1
☐	PREDICTED: dnaJ homolog subfamily A member 2 [Pygocentrus nattereri]	337	337	86%	4e-109	48%	XP_017579754.1
☐	PREDICTED: dnaJ homolog subfamily A member 2 [Lepisosteus oculatus]	332	332	86%	7e-107	49%	XP_006641433.1
☐	dnaJ homolog subfamily A member 2 [Danio rerio]	331	331	86%	1e-106	48%	NP_998658.1
☐	dnaJ homolog subfamily A member 2 [Xenopus tropicalis]	330	330	86%	3e-106	47%	NP_001004807.1
☐	hypothetical protein XELAEV_18025031mg [Xenopus laevis]	328	328	86%	1e-105	47%	OCT82499.1
☐	dnaJ homolog subfamily A member 2 [Anoplophora glabripennis]	327	327	85%	4e-105	49%	XP_018567925.1
☐	PREDICTED: dnaJ homolog subfamily A member 2 [Nanorana parkeri]	327	327	86%	6e-105	47%	XP_018425516.1

Figure 4.37: Sequence homology of cDNA PLCc43 showing 51% homology (65% positives) with Hsp40 Type I protein of *Acanthameoba castellanii*.

4.39). Further, polypeptide analysis of PLCc43 coding for Hsp40 Type I like protein also marked the presence of 22 metabolically important phosphorylation sites (Figure 4.39). Prediction of secondary structure of the deduced polypeptide marked the presence of different spatial arrangements among the amino acids (Figure 4.40). Multiple sequence alignment of amino acid sequence of deduced polypeptide with various other members of Hsp40 Type I protein showed the presence of these conserved domains in sequences which advocates this polypeptide as a member of Hsp40 Type I family (Figure 4.41).



-5'3' Frame 2

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

Figure 4.38: Translation of nucleotide sequence of cDNA PLCc43 showing the nucleotide sequence and encoded amino acids. Full length ORF is highlighted red region.

MAKEFHPDRNPAAGDKFKEISFAYEVLSDPEKREVYDRYGEQRLKEGPSA 50

GGGGGFDSAEDLFASLFGGGFFGGMSGGGGGGRGRQRKGEDLVHSLEVTM 100

EDLYNGKVSKLALRKQV**ICGVCQGKG**GRNPDSVKP**CDTCRGTG**FKVTLRQ 150

LGPGMVQQLQTV**CPKCRGQG**EVIREKDR**CKRCRGS**KVVQEKKVLDIYIDK 200

GMRHKQKIVFTGEADQEPGVIPGDVIIILKELSHPRFKRNDANLLIEQEI 250

TLVEALCGFSFALLHLDSRTLLIXSXPRRSDSAR 284

Figure 4.39: Deduced amino acid sequence of Hsp40 type I like protein encoded by cDNA PLCc43: Highly conserved tri-peptide HPD (green highlighted) and G/F rich region (boxes in bold) is shown. Four cysteine rich domains CXXCXGXG (yellow highlighted bold letters) are underlined in the polypeptide sequence. Putative phosphorylation sites are highlighted in red.

The sequence characterization of cDNA PLCc43 clone thus evidenced features supporting that the encoded polypeptide may belong to the class of Hsp40 Type I proteins. The sequence of PLCc43 evidenced in this study was deposited at NCBI database under the accession number MK079355. Furthermore, phylogenetic analysis of cDNA PLCc43 with the similar genes corresponding to various taxa have clearly clustered cDNA PLCc43 with kingdom Animalia that supported the fact that this gene might have originated from higher animals (Figure 4.42).

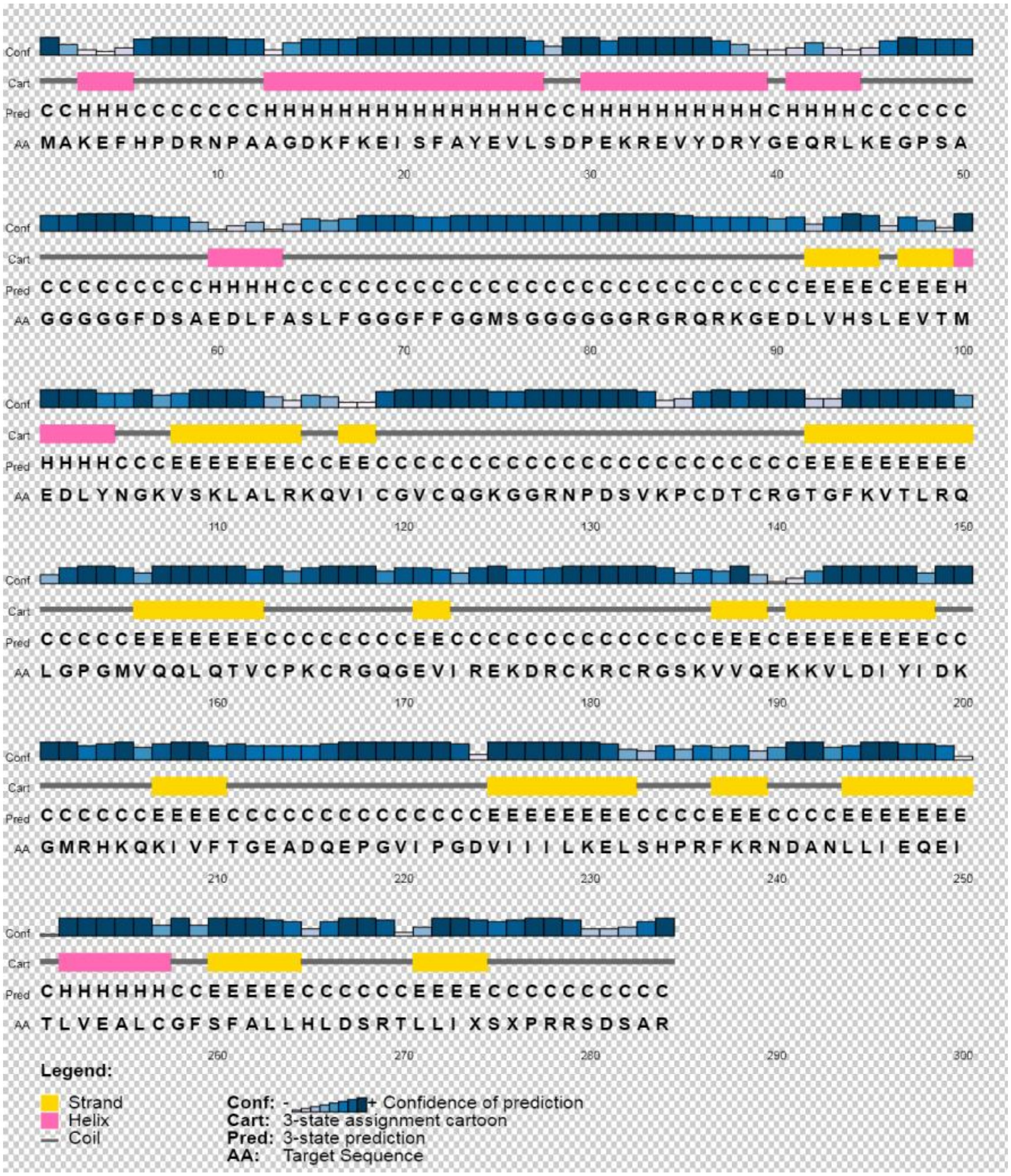


Figure 4.40: Prediction of secondary structure of deduced amino acid sequence of metatranscriptomic cDNA PLCc43 with the confidence of prediction for various spatial arrangements among amino acids.

Figure 4.41: Alignment of sequences of Hsp40 type I like proteins. Alignment was performed using the Clustal omega program. The fully conserved tripeptide HPD and the cysteine rich domains CXXCXGXG are highlighted in yellow and red colors respectively. Hsp40 type I from *Acanthameoba castellini* (XP_004349556) and DnaJ homolog members from *Danio rerio* (NP_998658) and *Xenopus tropicalis* (NP_001004807) are shown.

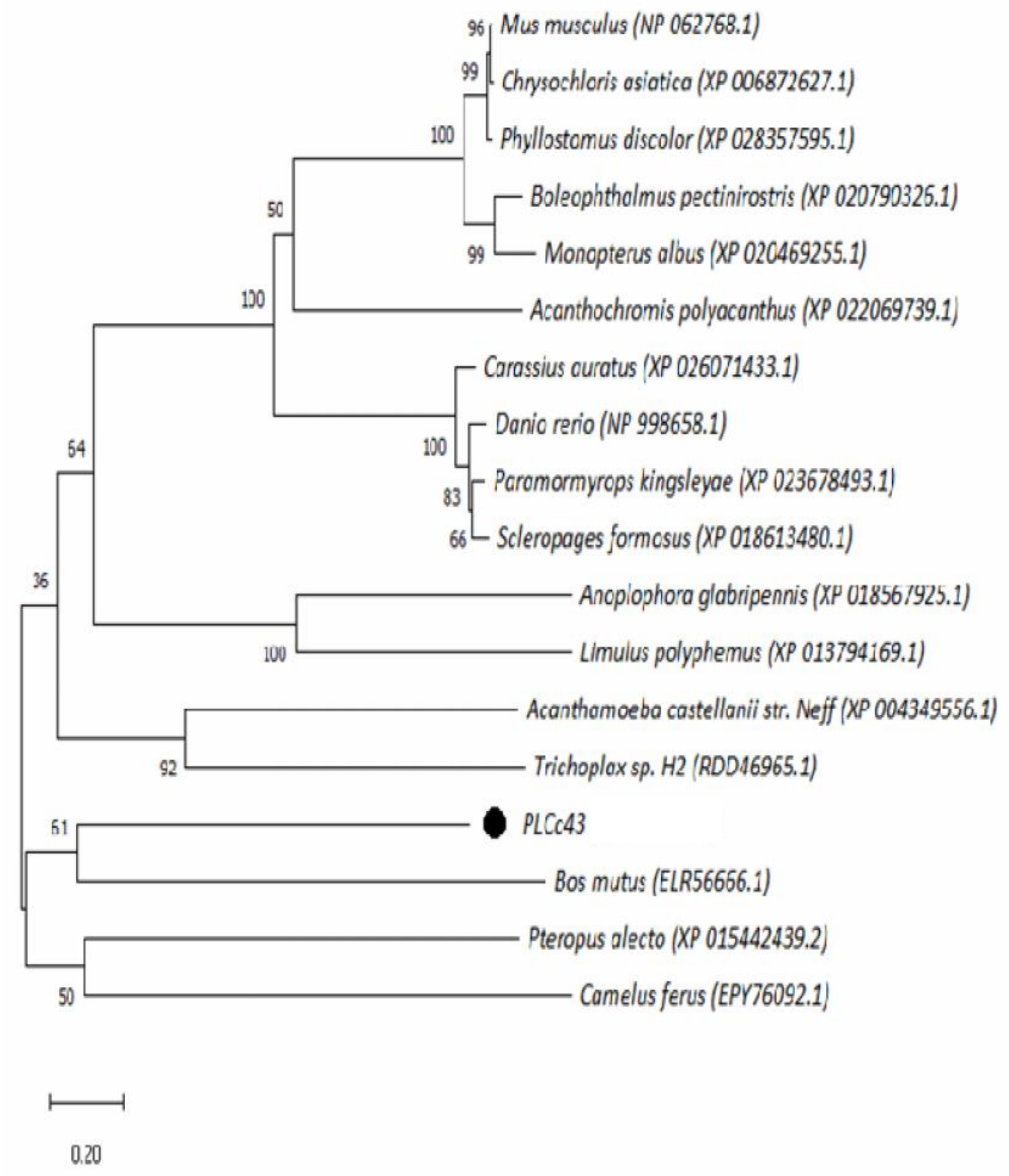


Figure 4.42: Phylogenetic relation of the transcript PLCc43 with other taxa corresponding to similar proteins. Phylogenetic tree was constructed using neighbor-joining method and the bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side where as GenBank Accession numbers are given in parentheses. Branch lengths are proportional to evolutionary distances.

Hsp40 proteins are highly conserved proteins that play an important role in protein degradation, translocation, folding and unfolding of protein by stimulating the ATPase activity of another 70 kDa chaperone, Hsp70, which is induced by stress (Caplan and Douglas 1991; Cyr *et al.* 1994; Qiu *et al.* 2006). Biochemical and genetic studies have reported the functional redundancy between *E. coli* DnaJ proteins and eukaryotic Hsp40 proteins; hence, the concept of regulation of Hsp70 by Hsp40 *via* stimulating ATPase activity with the help of J domain as reported in *E. coli* has also been conserved in higher organisms as well (Cyr *et al.* 1994). Similarly, the tri-peptide motif, HPD has also been evidenced to play crucial role in functional regulation of Hsp70 as mutations in this conserved motif block the activity of the Hsp70 ATPase (Wall *et al.* 1994; Tsai and Douglas, 1996). Although J domain is considered as an essential feature for functioning of Hsp40 proteins, studies have identified Hsp40 proteins that do not need J domain to perform all their functions (Hageman *et al.* 2010; Sahi *et al.* 2010). This further suggests that Hsp40 proteins can perform J domain independent functions and not all the functions of Hsp40 proteins are Hsp70 dependent. For instance, DnaJA1, a class I Hsp40 protein in mammals, which acts as host factor for influenza A virus replication could associate itself with the PA and PB2 subunits of RNA polymerase by C-terminal substrate binding domain, which lacks J domain (Cao *et al.* 2014). Further, cysteine rich zinc finger domains of polypeptide encoded by cDNA PLCc43 possess same sequence pattern as in type I Hsp40 protein of *S. cerevisiae* Ydj1 (Caplan and Douglas 1991).When analyzing the structure of Hsp40 proteins, Banecki et al. (1996) reported that the four-cysteine rich zinc finger domains combine by pairs; each of pairs binding to one unit of Zn^{2+} ion, hence, allowing the Hsp40 proteins to complex two Zn^{2+} ions. Zinc finger domains known to prevent the aggregation of denatured proteins, is also a part of peptide binding domain, which is responsible for chaperone function of Hsp40 proteins (Banecki *et al.* 1996; Szabo *et al.* 1996). The fourth domain found in carboxyl-terminal part of the protein was less conserved and plays a significant role in substrate binding (Banecki *et al.* 1996). Lu and Cyr (1998) reported that the mutant Ydj1$_{179\text{-}384}$ protein with completely lacking J domain, G/F region and a large part of cysteine rich zinc finger region, was able to suppress the aggregation of rhodanese *in vitro* thus arguing that some functions relies on the C-terminal substrate binding fragment only. Similarly, Sahi et al. (2013) advocates that C-terminal region of Hsp40 Class I proteins is most evolved region of Hsp40 proteins, which results into a more diverse and specific Hsp40 class I protein family with binding capability towards different polypeptides. Thus, all these structural similarities between the screened cDNA PLCc43 and already reported Hsp40 Type I gene/proteins supported the metatranscriptomic cDNA as a potential member of Hsp40 Type I like protein family.

4.3.3.2. Relative quantification of cDNA PLCc43

The cDNA PLCc43, which was responsible for Cd tolerance phenotype in hypersensitive yeast mutant *ycf1*$^{\Delta}$ was further tested for its relative expression inside these yeast cells on Cd exposure. Expression level of cDNA PLCc43 increased with increase in Cd concentration and the maximum expression was observed at 80 μM of $CdSO_4$, which was sixty eight fold higher to control cells (Table 4.13) (Figure 4.43). Eighteen fold increase at 40 μM and 55 fold increases at 60 μM $CdSO_4$ was recorded respectively in *ycf1*$^{\Delta}$/PLCc43 transformed cells compared to control cells.

Table 4.13: Effect of different cadmium concentrations on the relative expression of cDNA PLCc43

Cd Concentration (µM)	**Fold expression**
40	18.02 ± 1.35a
60	55.36 ± 3.67b
80	68.64 ± 1.65c

Values sharing a common letter within the column are not significant at $P<0.05$ (Fold expression values are Mean±SD, n=3)

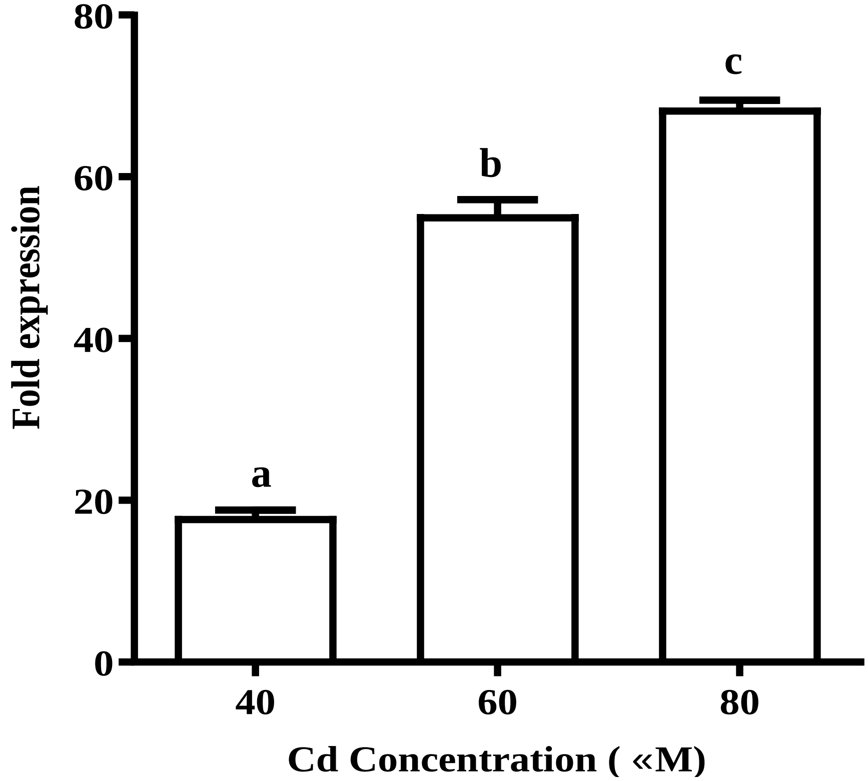


Figure 4.43: Fold increase in expression levels of cDNA PLCc43 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCc43 without metal treatment). Bars sharing a common letter are not significant at $P<0.05$ (n=3).

Several studies have reported the over-expression of Hsp40 Type I proteins along with other heat shock proteins and their role in degradation and refolding of proteins in response to various accumulated metals such as Cd, Cu and Hg in various living organisms (Barque *et al.* 1996; Cai and Zhang, 2000; Chai *et al.* 2000; Fusco *et al.* 2005).

4.3.3.3. Tolerance to metal toxicity and metal content

The role of cDNA PLCc43 in conferring metal tolerance when over-expressed within different hypersensitive mutants of the baker's yeast *S. cerevisiae* was further studied by drop assay and growth assay in specific concentrations of different metals. Following drop tests, the heterologous expression of cDNA PLCc43 within all hypersensitive yeast mutant strains used (*ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$) was reported to allow significant tolerance towards Cd, Cu, Zn and Co respectively, in contrast to yeast strains carrying empty expression vector (Figure 4.44). The yeast mutant *ycf1*$^{\Delta}$ transformed with cDNA PLCc43 has shown tremendous growth on the Cd amended media plate as compared to the mutant strain carrying empty vector (Figure 4.44a). Similarly, when cDNA was tested for its behaviour in Cu amended media, the surplus growth of transformed cells on the Cu amended media has further supported this cDNA as a copper tolerant affiliate (Figure 4.44b). Similar tolerant phenotypes were conferred to both Zn sensitive as well as Co sensitive yeast mutants carrying cDNA PLCc43 on exposure to metals such as Zn and Co (Figure 4.44c & d).

The rescue of sensitivity towards the four metals used in this study due to over expression of cDNA PLCc43 was further authenticated by growing the transformed yeast mutants into metal amended liquid media (Table 4.14). It was observed that the transformed

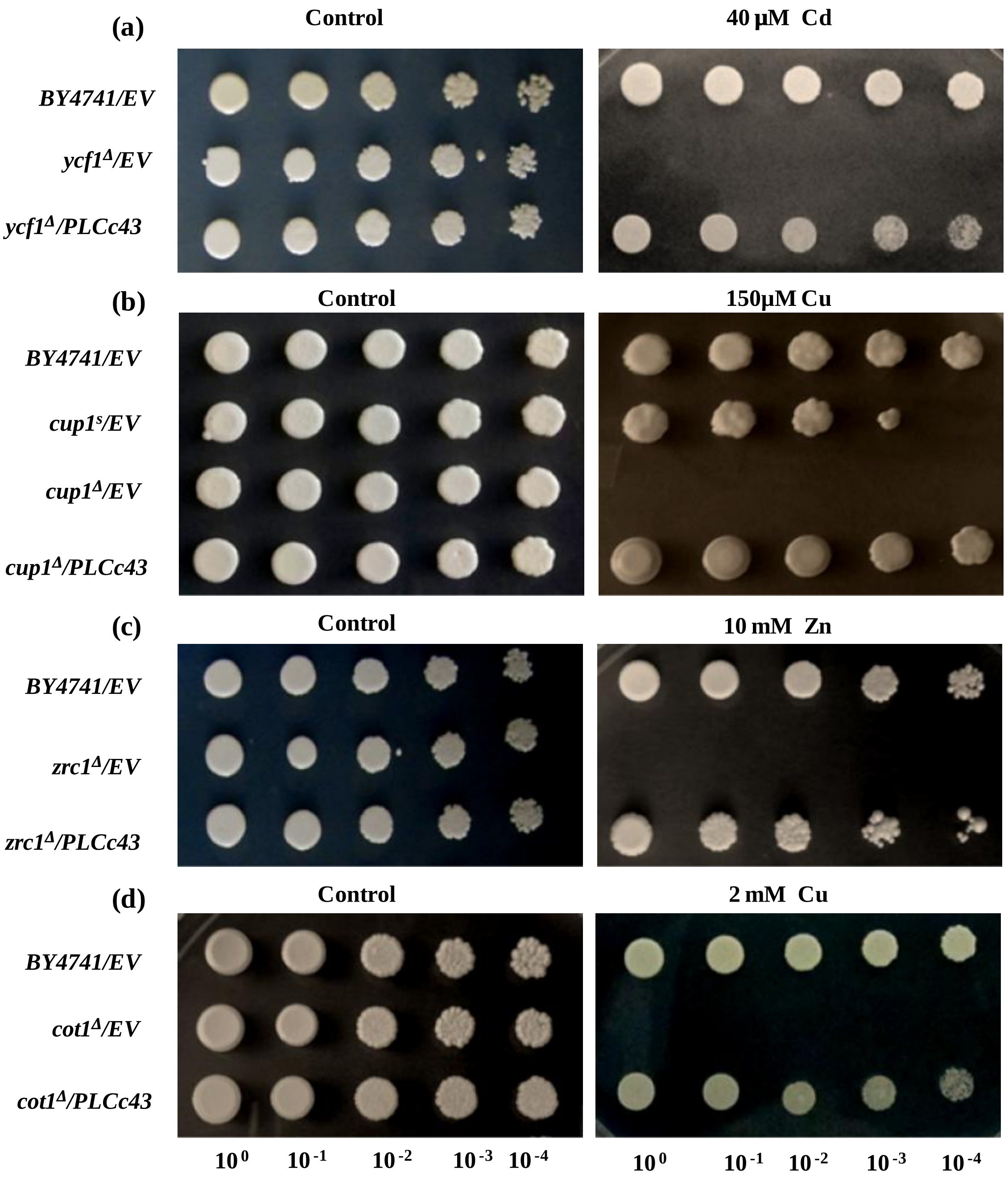


Figure 4.44: Drop assay of metal sensitive mutant strains *$ycf1^{\Delta}$*, *$cup1^{\Delta}$*, *$zrc1^{\Delta}$*and *$cot1^{\Delta}$* expressing PLCc43 at various concentrations (a) $CdSO_4$ 40 µM, (b) $CuSO_4$ 150 µM, (c) $ZnSO_4$ 10 mM and (d) $CoCl_2$ 2 mM. Wild strain BY4741/*EV*, *$cup1^{s}$/EV* serves as positive controls and *$ycf1^{\Delta}$*/EV, *$cup1^{\Delta}$/EV*, *$zrc1^{\Delta}$/EV* and *$cot1^{\Delta}$/EV* as negative controls respectively. EV stands for empty vector.

yeast mutant *ycf1*$^{\Delta}$/PLCc43 has shown 15 times more tolerance towards Cd as compared to original hypersensitive yeast mutant *ycf1*$^{\Delta}$ when 40 µM of $CdSO_4$ was used. Tolerance level of yeast strain *ycf1*$^{\Delta}$/PLCc43 was elevated to 60 times those observed for strain *ycf1*$^{\Delta}$/EV at 60 µM of $CdSO_4$. Tolerance profile of yeast mutant strain expressing cDNA PLCc43 shows twelve times increase in tolerance at 80 µM $CdSO_4$ compared to 40 µM exposition, thus supporting PLCc43 as prospective member of gene providing tolerance to Cd (Figure 4.45a). For Cu metal, expression of cDNA PLCc43 when carried in the mutant strain *cup1*$^{\Delta}$ led to significant growth increase of the recipient strain exposed to either concentrations of $CuSO_4$ tested, ranging from 15 to 59 times those observed for the sensitive strain *cup1*$^{\Delta}$ harboring empty vector (Figure 4.45b). Similarly, after 48 hrs of incubation with different range of Zn concentrations, the growth of *zrc1*$^{\Delta}$ cells carrying cDNA PLCc43 showed 18-fold increase at 10 mM $ZnSO_4$ in comparison to the mutant strain with empty vector. Likewise, cDNA PLCc43 restored tolerance towards Zn toxicity with increasing Zn concentration i.e., 11 mM and 12 mM (Figure 4.45c). In case of cobalt, over expression of cDNA PLCc43 in hypersensitive yeast mutant *cot1*$^{\Delta}$ provided 22 times more tolerance against Co toxicity and helped mutant strain *cot1*$^{\Delta}$ carrying PLCc43 to sustain even at higher range of Co concentration till 4 mM of $CoCl_2$ (Figure 4.45d). Henceforth, these tolerance phenotypes offered to different metal sensitive yeast mutants by the presence of cDNA PLCc43 advocate the role of encoded Hsp40 Type I like protein in detoxifying the lethal effects of the PTMs.

Table 4.14: Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLCc43 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant strain with empty vector pFL61 (EV).

		cDNA PLCc43	
		O.D. 600	
Cadmium (µM)	***ycf1*$^{\Delta}$*/EV***	***BY4741/EV***	***ycf1*$^{\Delta}$/ PLCc43**
40	0.09 ± 0.01a	1.42 ± 0.116b	5.25 ± 0.07c
60	0.06 ± 0.01a	0.75 ± 0.29b	3.8 ± 0.29c
80	0.04 ± 0.006a	0.39 ± 0.18b	1.10 ± 0.12c
		O.D. 600	
Copper (µM)	***cup1*$^{\Delta}$*/EV***	***cup1*^{s}*/EV***	***cup1*$^{\Delta}$/ PLCc43**
150	0.24 ± 0.032a	0.56 ± 0.012b	3.60 ± 0.32c
300	0.07 ± 0.007a	0.19 ± 0.03b	2.39 ± 0.35c
500	0.05 ± 0.001a	0.04 ± 0.001b	1.50 ± 0.11c
		O.D. 600	
Zinc (mM)	***zrc1*$^{\Delta}$*/EV***	***BY4741/EV***	***zrc1*$^{\Delta}$/ PLCc43**
10	0.34 ± 0.07a	1.38 ± 0.19b	6.18 ± 0.25c
11	0.14 ± 0.03a	0.96 ± 0.03b	4.27 ± 0.15c
12	0.084 ± 0.014a	0.48 ± 0.09b	2.69 ± 0.38c
		O.D. 600	
Cobalt (mM)	***cot1*$^{\Delta}$*/EV***	***BY4741/EV***	***cot1*$^{\Delta}$/ PLCc43**
2	0.26 ± 0.07a	1.49 ± 0.19b	4.76 ± 0.34c
3	0.14 ± 0.007a	0.97 ± 0.11b	2.69 ± 0.11c
4	0.05 ± 0.009a	0.14 ± 0.02b	1.91 ± 0.11c

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

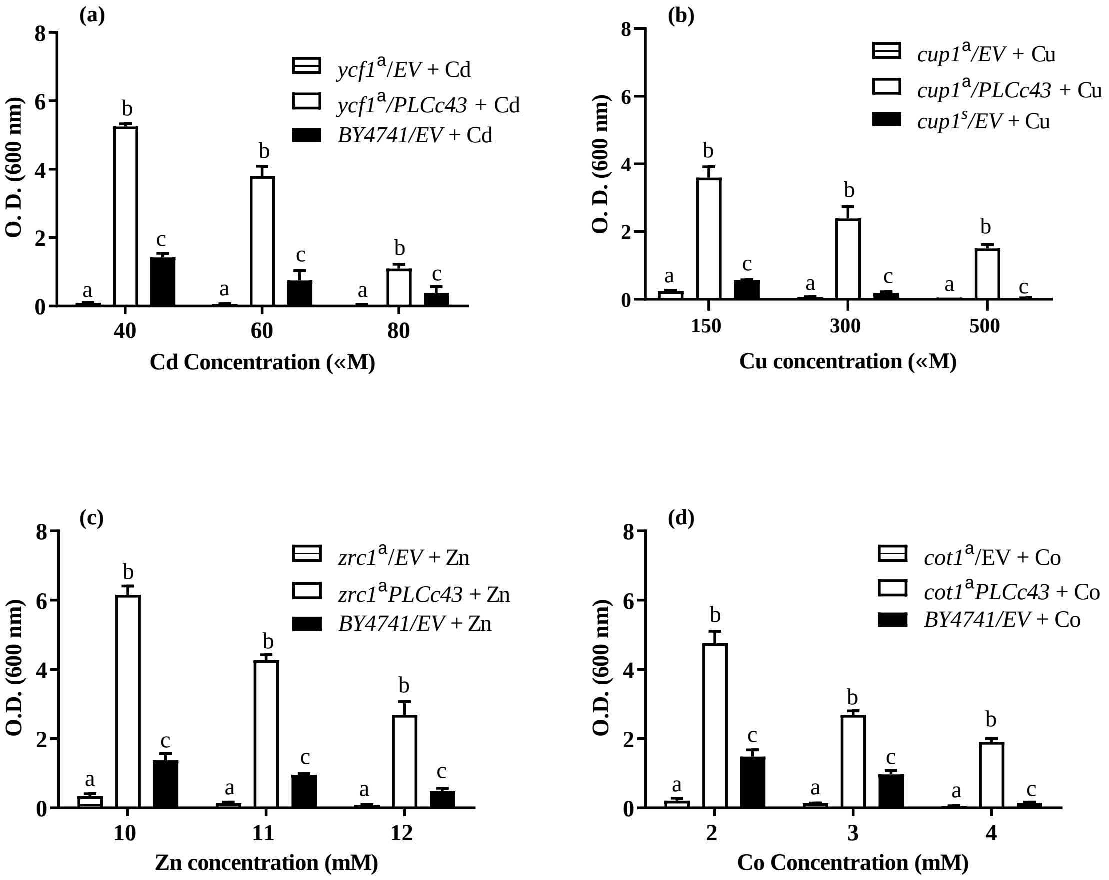


Figure 4.45: Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLCc43 in SD-Ura medium supplemented with (a) $CdSO_4$ 40-80 µM, (b) $CuSO_4$ 150-500 µM, (c) $ZnSO_4$ 10-12 mM, (d) $CoCl_2$ 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3).

4.3.3.4. Yeast metal uptake by cDNA PLCc43

Accumulation of metal by respective yeast mutant strain carrying cDNA PLCc43 was further estimated by ICP-MS and observed that cDNA PLCc43 played a significant role in accumulating higher range of metals in comparison to the wild type strain and hypersensitive mutant strains (Table 4.15). An increase in metal accumulation up to 60 μM of $CdSO_4$ followed by a decrease at 80 μM of $CdSO_4$ was observed. The metal accumulation levels were significantly higher in *ycf1*$^{\Delta}$/PLCc43 compared to wild type BY4741/EV and also *ycf1*$^{\Delta}$/*EV* (Figure 4.46a). Similarly, the level of Cu accumulation increased with copper concentrations up to 300 μM of $CuSO_4$ after which a slight but not significant decrease was observed at 500 μM of $CuSO_4$ concentration (Figure 4.46b). In case of Zn and Co metals, maximum metal accumulation was observed for 10 mM of $ZnSO_4$ and 2 mM of $CoCl_2$ respectively, after which a sharp decrease was observed for either concentration of metal used (Fig.4.46c & d). Overall, metal accumulation levels were significantly higher in yeast mutants expressing cDNA PLCc43 than in either mutants or wild-type strain transformed with empty vector. Therefore, increase in metal accumulation efficiency of various hypersensitive yeast mutants when transformed with cDNA PLCc43 supported this soil metatranscriptomic cDNA as potential member of metal tolerant gene family. In 1991, Nover reported that the divalent cations of transition metal series such as Cd, Cu, Zn, and Hg act as heat stress-like chemical stressors, which induce the production of various stress combating proteins such as Hsp40. It was observed that Hsp40 in association with Hsp70 prevents aggregation of proteins and assist in refolding of non-native proteins under both normal and stressed environments (Hartl, 1996; Frydman, 2001). Degradation of non-native proteins or proteolytic inactivation of regulatory proteins is major cellular response of living organisms in metal contaminated environments. Hsp40 proteins are known to regulate the complex formation between the aberrant proteins and Hsp70 by mainly three mechanisms. First mechanism involves binding of aberrant proteins to the unique class of polypeptide domains (PPDs) present on Hsp40 which have capability to deliver these special clients (aberrant proteins) to Hsp70 (Cyr *et al.* 1994; Cheetham and Caplan 1998).

Table 4.15: Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCc43 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV).

	cDNA PLCc43		
	Cadmium uptake (mg^{-1}g dry weight)		
Cd (µM)	***ycf1*$^{\Delta}$/ PLCc43**	***BY4741/EV***	***ycf1*$^{\Delta}$/EV**
40	0.36 ± 0.015a	0.15 ± 0.012b	0.015 ± 0.001c
60	0.43 ± 0.017a	0.25 ± 0.002b	0.025 ± 0.010c
80	0.18 ± 0.019a	0.05 ± 0.00b	0.009 ± 0.001c
	Copper uptake (mg^{-1}g dry weight)		
Cu (µM)	***cup1*$^{\Delta}$/ PLCc43**	***BY4741/EV***	***cup1*$^{\Delta}$/EV**
150	0.51 ± 0.005a	0.36 ± 0.013b	0.026 ± 0.001c
300	0.67 ± 0.018a	0.22 ± 0.005b	0.022 ± 0.001c
500	0.59 ± 0.013a	0.16 ± 0.009b	0.008 ± 0.001c
	Zinc uptake (mg^{-1}g dry weight)		
Zn (mM)	***zrc1*$^{\Delta}$/ PLCc43**	***BY4741/EV***	***zrc1*$^{\Delta}$/EV**
10	0.89 ± 0.05a	0.76 ± 0.066b	0.05 ± 0.005c
11	0.82 ± 0.05a	0.31 ± 0.006b	0.03 ± 0.0005c
12	0.49 ± 0.009a	0.21 ± 0.01b	0.02 ± 0.001c
	Cobalt uptake (mg^{-1}g dry weight)		
Co (mM)	***cot1*$^{\Delta}$/ PLCc43**	***BY4741/EV***	***cot1*$^{\Delta}$/EV**
2	0.26 ± 0.008a	0.04 ± 0.001b	BDL
3	0.59 ± 0.024a	0.05 ± 0.001b	BDL
4	0.36 ± 0.021a	0.013 ± 0.001b	BDL

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

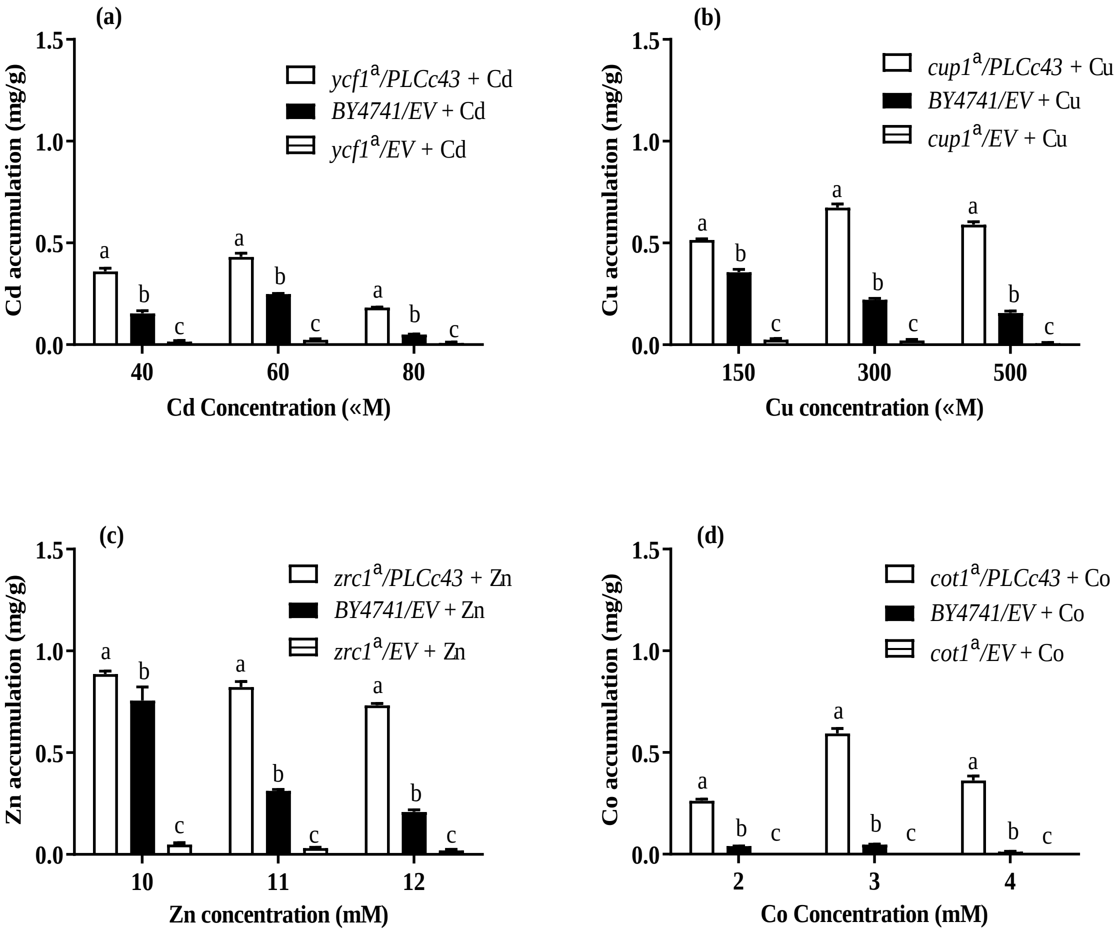


Figure 4.46: Accumulation of a) Cd^{2+}, b) Cu^{2+}, c) Zn^{2+} and d) Co^{2+} in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCc43. Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3).

Secondly, Hsp40 stabilizing the Hsp70-polypeptide complex by dephosphorylation of ATP complex to adenosine diphosphate (ADP) complex (Liberek *et al.* 1991; Cyr *et al.* 1992; Langer *et al.* 1992). Third mechanism involves the specialized members of Hsp40 family, which are localized to different sites in the same cellular compartment and interaction between Hsp70 and differently localized Hsp40s enables different Hsp70-Hsp40 pairs to bind unique clients especially aberrant proteins at these sites (Brodsky and

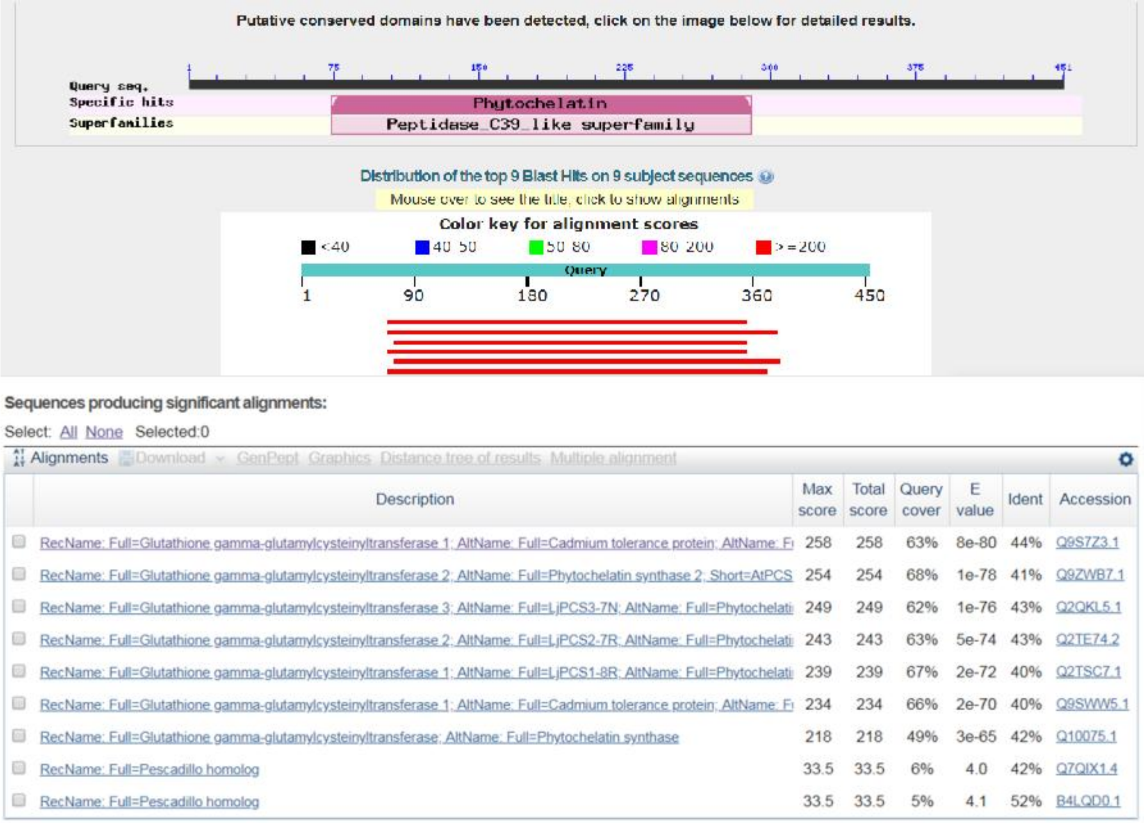
Schekman 1993; Cyr and Neupert 1996; Shen *et al.* 2002). Meacham et al. (1999) observed that membrane localized human Hsp40 protein Hdj2 along with Hsp70 and co-chaperone C-terminus of heat-shock cognate 70 stress protein-interacting protein (CHIP) promotes the ubiquitination of misfolded proteins such as the cystic fibrosis transmembrane conductance regulator. It has been reported that both Hsp40 and Hsp40 like proteins are involved in maintaining the cell homeostasis through translocation, refolding or degradation of aberrant proteins during cellular stress such as metal toxicity (Suzuki *et al.* 2001; Cyr *et al.* 1994). In a study, a total of twenty-one cDNAs including Hsp40 were reported in birch *Betula pendula* playing significant role in maintaining the homeostasis and in alleviating the toxic effects of accumulated Cu (Keinänen *et al.* 2007). Several studies have reported the over-expression of Hsp40 proteins along with other heat shock proteins and their role in degradation and refolding of proteins in various living organisms in response to various accumulated metals (Barque *et al.* 1996; Cai and Zhang, 2000; Fusco *et al.* 2005). Chai *et al.* (2000) reported that Hsp40 like protein encoding gene is highly expressed in bean when exposed to mercuric chloride. The mechanism behind this might be direct binding and trafficking, which has been observed experimentally with copper (Field *et al.* 2002), or by refolding of denatured proteins due to the stress conditions (Suzuki *et al.* 2001; Shen and Hendershot 2005). Hasan *et al.* (2017) reported that in many plant species, Hsps are highly expressed protein under a variety of metal stress. The expression of Hsp70 limits the proteotoxic symptoms of metal ions and helps in detoxification and sequestration of these ions by metallothioneins (Haap *et al.* 2016). Hsp70 chaperones, along with Hsp40, prevent accumulation of newly synthesized proteins as aggregates and assure the appropriate folding of protein during their transfer to the destination (Park and Seo, 2015). Overall, these findings support the role of Hsp40 Type I like protein encoded by cDNA PLCc43 in combating the toxicity caused by excessive metals inside the yeast cells, hence supported this soil cDNA as a potential member of metal tolerant gene family that could be beneficial for rejuvenating the metal polluted sites.

4.3.4. PLCd43/ γ-Glutamylcysteinyl transferase

Cadmium tolerant transformant *ycf1*$^{\Delta}$/PLCd43 screened from cDNA library C, showed presence of full length ORF by BLAST analysis. This cDNA was further analyzed for gene characterization with respect to sequence analysis and tolerance potential towards other metals used in the study.

4.3.4.1. Sequence analysis

Sequence analysis of cDNA PLCd43 showed that 1393 bp cDNA consisted an ORF of 1356 bp encoding a polypeptide of 451 amino acids. It showed 44% identity (62% positives) with γ-Glutamylcysteinyl transferase enzyme of *Arabidopsis thaliana* with predicted molecular mass and isoelectric point of 51 kDa and 6.8 respectively (Figure 4.47 & 4.48). γ-Glutamylcysteinyl synthase is an enzyme with a trivial name phytochelatin synthase (PC synthase) which catalyzes the non-ribosomal synthesis of phytochelatins, a metal binding peptides, from glutathione molecules by transpeptidation (Clemens and Peršoh, 2009). Sequence alignment of deduced polypeptide encoded by cDNA PLCd43 highlighted the elementary features possessed by a PC synthase (Figure 4.49).



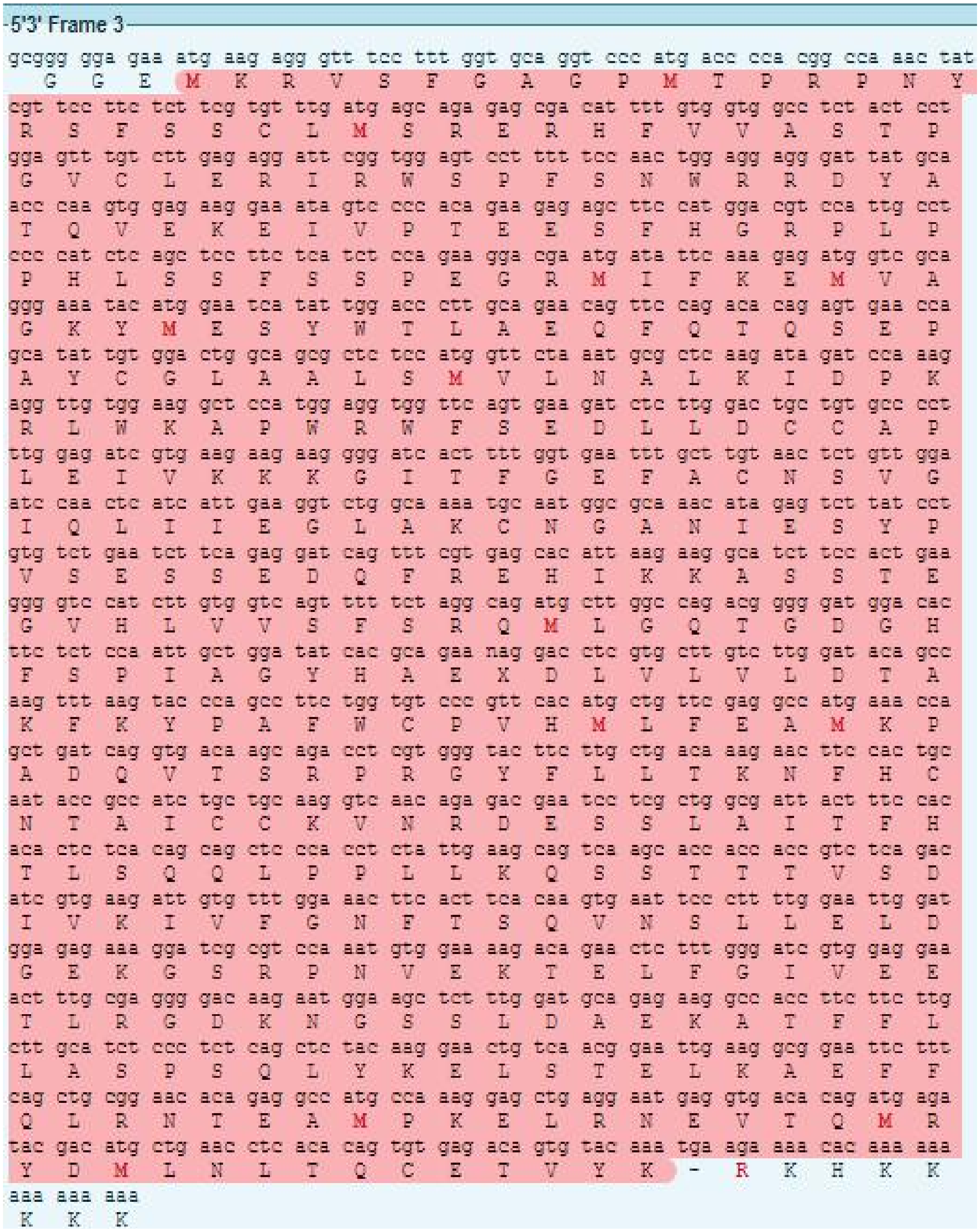
Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	RecName: Full=Glutathione gamma-glutamylcysteinyltransferase 1; AltName: Full=Cadmium tolerance protein; AltName: Fı	258	258	63%	8e-80	44%	Q9S7Z3.1
☐	RecName: Full=Glutathione gamma-glutamylcysteinyltransferase 2; AltName: Full=Phytochelatin synthase 2; Short=AtPCS	254	254	68%	1e-78	41%	Q9ZWB7.1
☐	RecName: Full=Glutathione gamma-glutamylcysteinyltransferase 3; AltName: Full=LjPCS3-7N; AltName: Full=Phytochelati	249	249	62%	1e-76	43%	Q2QKL5.1
☐	RecName: Full=Glutathione gamma-glutamylcysteinyltransferase 2; AltName: Full=LjPCS2-7R; AltName: Full=Phytochelati	243	243	63%	5e-74	43%	Q2TE74.2
☐	RecName: Full=Glutathione gamma-glutamylcysteinyltransferase 1; AltName: Full=LjPCS1-8R; AltName: Full=Phytochelati	239	239	67%	2e-72	40%	Q2TSC7.1
☐	RecName: Full=Glutathione gamma-glutamylcysteinyltransferase 1; AltName: Full=Cadmium tolerance protein; AltName: Fı	234	234	66%	2e-70	40%	Q9SWW5.1
☐	RecName: Full=Glutathione gamma-glutamylcysteinyltransferase; AltName: Full=Phytochelatin synthase	218	218	49%	3e-65	42%	Q10075.1
☐	RecName: Full=Pescadillo homolog	33.5	33.5	6%	4.0	42%	Q7QIX1.4
☐	RecName: Full=Pescadillo homolog	33.5	33.5	5%	4.1	52%	B4LQD0.1

Figure 4.47: Sequence homology of cDNA PLCd43 showing 44% identity with γ-Glutamylcysteinyl transferase (PC synthase) enzyme of *Arabidopsis thaliana.*

Cobbett and Goldsbrough, (2002) studied a group of PC synthases and reported the presence of various full length or partial cDNAs encoding PC synthase with molecular weight varied between 42 kD to 70 kD. They observed that N-terminal region of PC synthase is more conserved i.e., 40-50% similar in yeast, plants and animals whereas in C-terminal region, the sequences are less conserved. Putative PLCd43p polypeptide may display the fundamental features possessed by PC synthase i.e., a highly conserved N-terminal domain and a variable C- terminal domain (Figure 4.50) (Fan *et al.* 2018). The N-terminal domain is highly conserved domain marked with the presence of C^{120}, H^{237} and D^{256} residues and five highly conserved cysteine residues that at position C^{120}, C^{154},



5'3' Frame 3

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

Figure 4.48: Translation of nucleotide sequence of cDNA PLCd43. Red highlighted region is full length ORF showing the nucleotide sequence and encoded amino acids.

C^{155}, C^{173} and C^{188}. The cysteine, histidine and aspartic acid residues in N-terminal domain have been reported to be essential for catalysis in *AtPCS1* while the remaining five cysteine residues played an important role in recognition of metal ions in all the eukaryotes (Vivares *et al.* 2005; Romanyuk *et al.* 2006). The variable C-terminal domain of PC synthase like enzyme has five cysteine residues and is considered as the only identifiable feature of C-terminal region. A variable cysteine rich region in C-terminal of PC synthase was reported in many species and observed to play a significant role in metal sensing ability and enzyme stability (Cobbett and Goldsbrough, 2002; Ruotolo *et al.* 2004). There are around 52 potential sites for phosphorylation in the deduced PC synthase like enzyme and the presence of majority of serine residues in the deduced polypeptide indicated that PC synthase like enzyme might be phosphorylated along with threonine and tyrosine (Figure 4.50). Secondary structure of deduced polypeptide was predicted with the various spatial arrangements among the amino acids that could be possible in a stable polypeptide (Figure 4.51). Phylogenetic analysis of cDNA PLCd43 with the similar genes corresponding to various taxon have clearly clustered cDNA with the family *Brassicaceae,* which support the fact that this gene might have originated from the plants (Figure 4.52). Hence, bioinformatic analyses have categorised cDNA PLCd43 as a PC synthase like protein and exploited further to study its multi metal tolerance phenotype.

```
PLCd43          MKRVSFGAGPMTPRPNYRSFSSCLMSRERHFVVASTPGVCLERIRWSPFSNWRRDYATQV 60
PCS1_WHEAT      ------------------------------------------------------------ 0
PCS1_ARATH      ------------------------------------------------------------ 0
PCS1_LOTJA      ------------------------------------------------------------ 0


PLCd43          EKEIVPTEESFHGRPLPPHLSSFSSPEGRMIFKEMVAGKYMESYWTLAEQFQTQSEPAYC 120
PCS1_WHEAT      ----MEVASLYRRVLPSPPAVEFASAEGKRLFAEALQGGTMEGFFNLISYFQTQSEPAFC 56
PCS1_ARATH      ----MAMASLYRRSLPSPPAIDFSSAEGKLIFNEALQKGTMEGFFRLISYFQTQSEPAYC 56
PCS1_LOTJA      ----MAMAGLYRRLLPSPPAVDFASSQGKQLFLEAVQNGTMESFYRLVSYFQTQSEPAFC 56
                    :     ::     *   .*:* :*: :* * :    **.:: * . ********:*

PLCd43          GLAALSMVLNALKIDPKRLWKAPWRWFSEDLLDCCAPLEIVKKKGITFGEFACNSVGIQL 180
PCS1_WHEAT      GLASLSVVLNALAIDPGRPWKGPWRWFDESMLDCCEPLHKVKAEGITFGKVVC------- 109
PCS1_ARATH      GLASLSVVLNALSIDPGRKWKGPWRWFDESMLDCCEPLEVVKEKGISFGKVVC------- 109
PCS1_LOTJA      GLASLSMVLNALAIDPGRKWKGPWRWFDESMLDCCEPLDKIKARGISFGKLVC------- 109
                ***:**:***** *** * **.*****.*.:**** **. :* .**:**:..*

PLCd43          IIEGLAKCNGANIESYPVSESSEDQFREHIKKASSTEGVHLVVSFSRQMLGQTGDGHFSP 240
PCS1_WHEAT      ----LAHCAGARVQSFRADQTTIHDFRAHLTRCASSQDCHLISSYHRSPFKQTGTGHFSP 165
PCS1_ARATH      ----LAHCSGAKVEAFRTSQSTIDDFRKFVVKCTSSENCHMISTYHRGVFKQTGTGHFSP 165
PCS1_LOTJA      ----LAHCAGAKVEAFHASHSSIDHFRKYVMKCSTSDDCHVISSYHREALKQTGTGHFSP 165
                    **:* **.:::: ...:: ..** .: :.::::. *:: :: *  : *** *****

PLCd43          IAGYHAEXDLVLVLDTAKFKYPAFWCPVHMLFEAMKPADQVTSRPRGYFLLTKNFHCNTA 300
PCS1_WHEAT      IGGYHAEKDMALILDVARFKYPPHWVPLTLLWDAMNTTDEATGLLRGFMLVSRRSSAPSL 225
PCS1_ARATH      IGGYNAERDMALILDVARFKYPPHWVPLKLLWEAMDSIDQSTGKRRGFMLISRPHREPGL 225
PCS1_LOTJA      IGGYHAGKDMALILDVARFKYPPHWIPLTHLWEGMNYVDESTGKTRGFMLISRPHREPGM 225
                *.**:*  *:.*:**.*:**** .* *:  *::.*.  *: *.  **::*:::
```

Figure 4.49: Alignment of amino acid sequence of cDNA PLCd43 encoding PC synthase like protein. Alignment was performed using the Clustal omega program and manually adjusted. Accession numbers are *Triticum aestivum* (Q9SWW5.1), *Arabidopsis thaliana* (Q9S7Z3), *Lotus japonicus* (Q2TSC7.1).

MKRV**S**FGAGPM**T**PRPNYRSF**S**S**C**LM**S**RERHFVVAS**T**PGV**C**LERIRW**S**PFS 50

NWRRDYA**T**QVEKEIVPTEESFHGRPLPPHLS**S**F**SS**PEGRMIFKEMVAGK**Y** 100

ME**SY**WTLAEQFQ**T**QSEPAY**C**GLAALSMVLNALKIDPKRLWKAPWRWF**S**ED 150

LLD**CC**APLEIVKKKGI**T**FGEFA**C**N**S**VGIQLIIEGLAK**C**NGANIE**S**YPV**S**E 200

SSEDQFREHIKKA**SST**EGVHLVV**S**FSRQMLGQTGDG**H**F**S**PIAGYHAEXDL 250

VLVL**DT**AKFK**Y**PAFW**C**PVHMLFEAMKPADQV**TS**RPRGYFLL**T**KNFH**C**NTA 300

I**CC**KVNRDE**SS**LAITFH**T**L**S**QQLPPLLKQ**S**S**T**T**T**VSDIVKIVFGNFT**S**QV 350

N**S**LLELDGEKG**S**RPNVEKTELFGIVEE**T**LRGDKNG**SS**LDAEKATFFLLA**S** 400

P**S**QLYKELS**T**ELKAEFFQLRN**T**EAMPKELRNEVTQMRYDMLNL**T**Q**C**E**T**VYK 451

Figure 4.50: Deduced amino acid sequence of PC synthase like protein encoded by cDNA PLCd43: Highly conserved C^{120}, H^{237} and D^{255} are highlighted in green while five conserved cysteine residues present at N-terminal of the polypeptide are boxed. Phosphorylation sites are underlined.

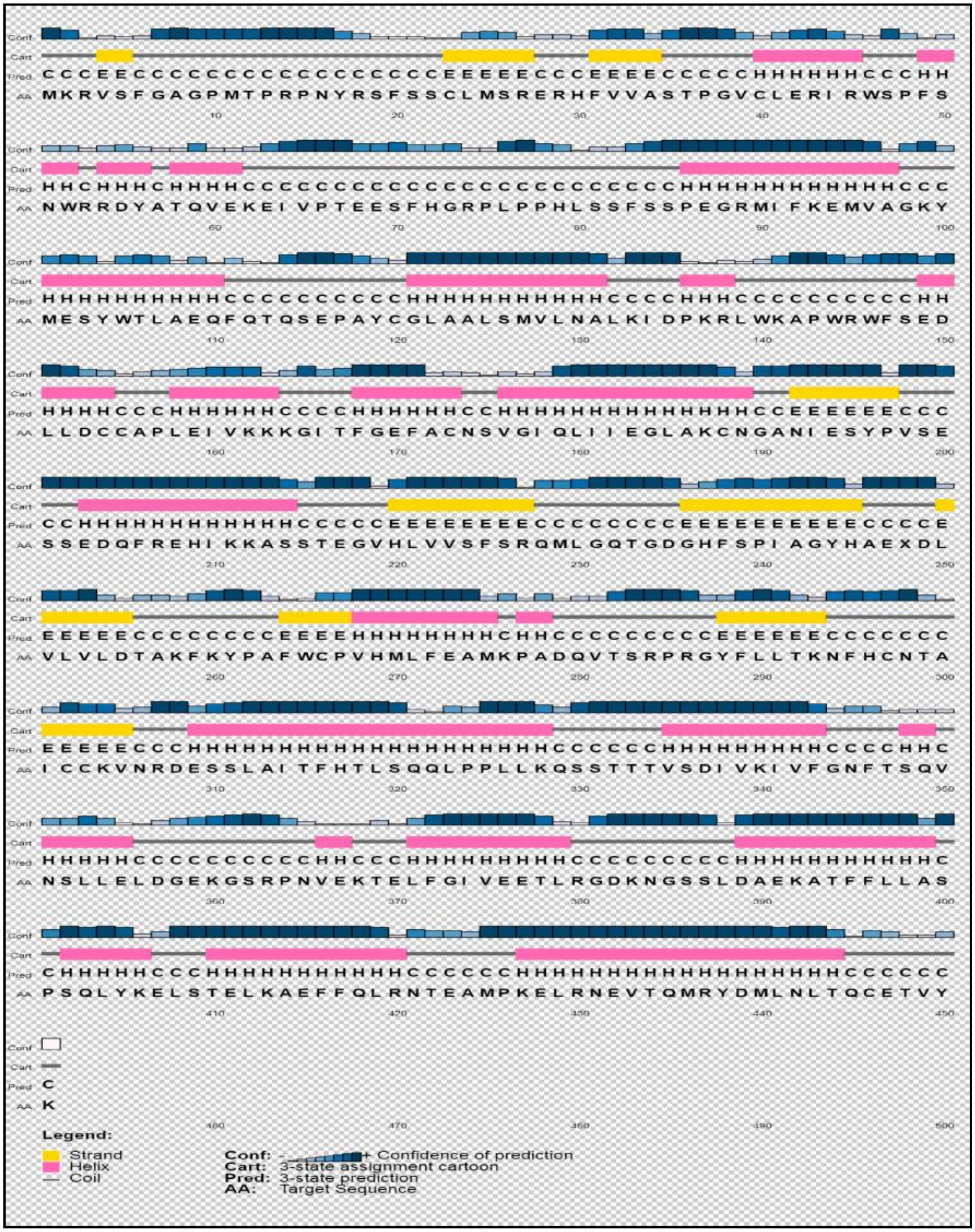


Figure 4.51: Secondary structure of deduced amino acid sequence encoded by metatranscriptomic cDNA PLCd43 showing different possible spatial arrangements among the amino acids along with confidence of prediction.

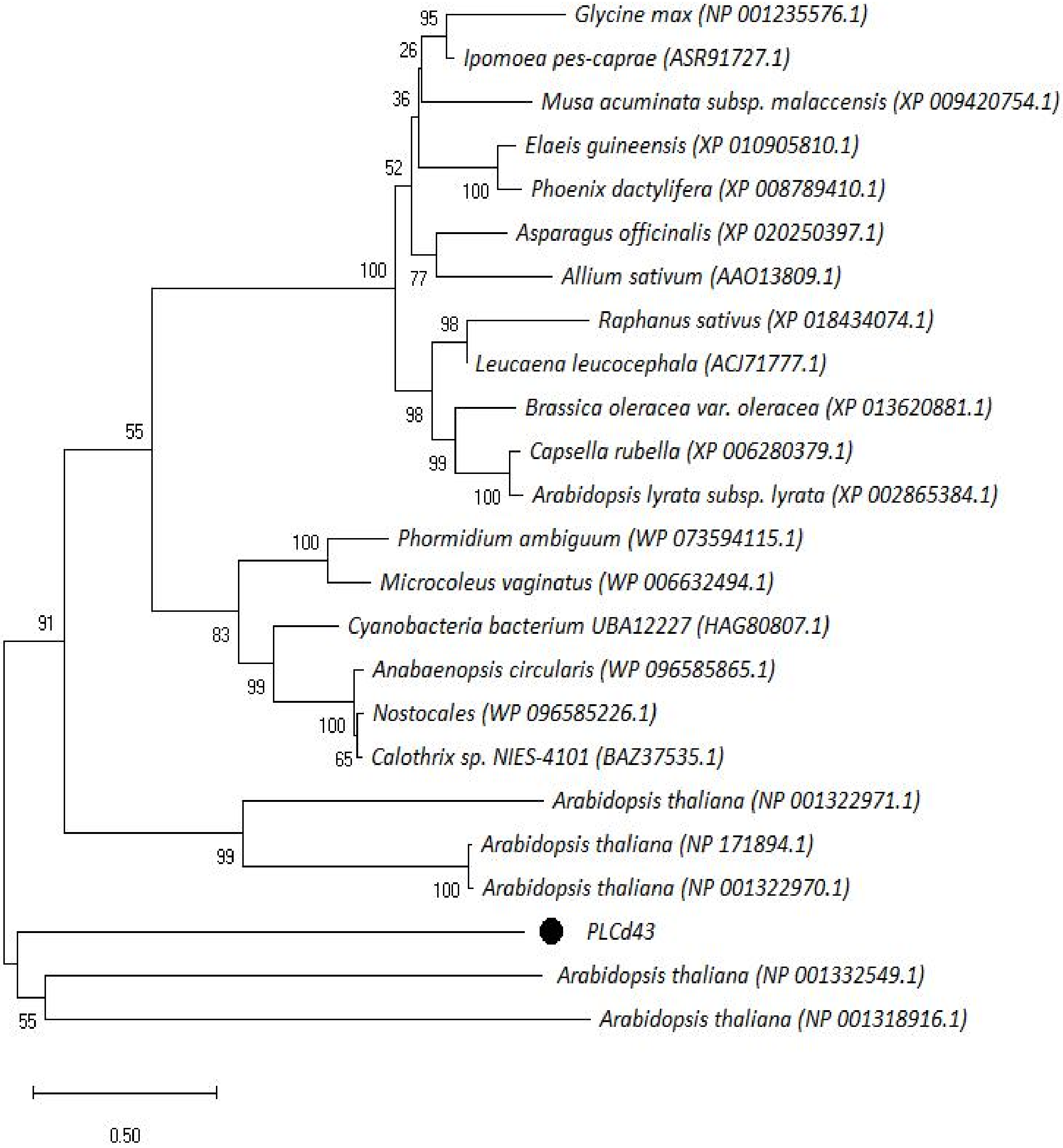


Figure 4.52: Phylogenetic relation of the transcript PLCd43 with PC synthase enzymes corresponding to other taxa constructed using neighbor-joining method. Bootstrap values were obtained with 1000 replicates are shown at the nodes. Species names are denoted on the right side where as GenBank Accession numbers are given in parentheses. Branch lengths are proportional to evolutionary distances.

4.3.4.2. Induction of cDNA PLCd43

The level of mRNA accumulated inside the transformed yeast cells *ycf1*$^{\Delta}$ on induced by Cd metal was estimated by quantitative RT-PCR. With the increase in metal concentration, the level of expression of cDNA PLCd43 has also increased (Table 4.16). At 40 µM of $CdSO_4$, PLCd43 showed 10 fold increase, which further increased to 24 folds at 60 µM and reaches up to 41fold at 80 µM of $CdSO_4$ (Figure 4.53). Clemens *et al.* (1999) reported similar findings for cDNA *TaPCS1* showing increase in fold expression of *TaPCS1* in presence of Cd and also observed increase amount of Cd accumulation in the yeast cells due to expression of cDNA *TaPCS1*. Fan *et al.* (2018) characterised two PC synthase encoding genes *MnPCS1* and *MnPCS2* from heavy metal tolerant tree *Morus notabilis* and reported an elevated level of expression in the presence of potentially toxic metals like Cd and Zn. Similar observations in our study interpreted as evidence in support of role of cDNA PLCd43 encoding PC synthase like enzyme in metal chelation or sequestration.

Table 4.16: Effect of different cadmium concentrations on the relative expression of cDNA PLCd43

Cd Concentration (µM)	**Fold expression**
40	10.13 ± 0.66a
60	24.00 ± 1.21b
80	40.82 ± 1.24c

Values sharing a common letter within the column are not significant at $P<0.05$ (Fold expression values are Mean±SD, n=3)

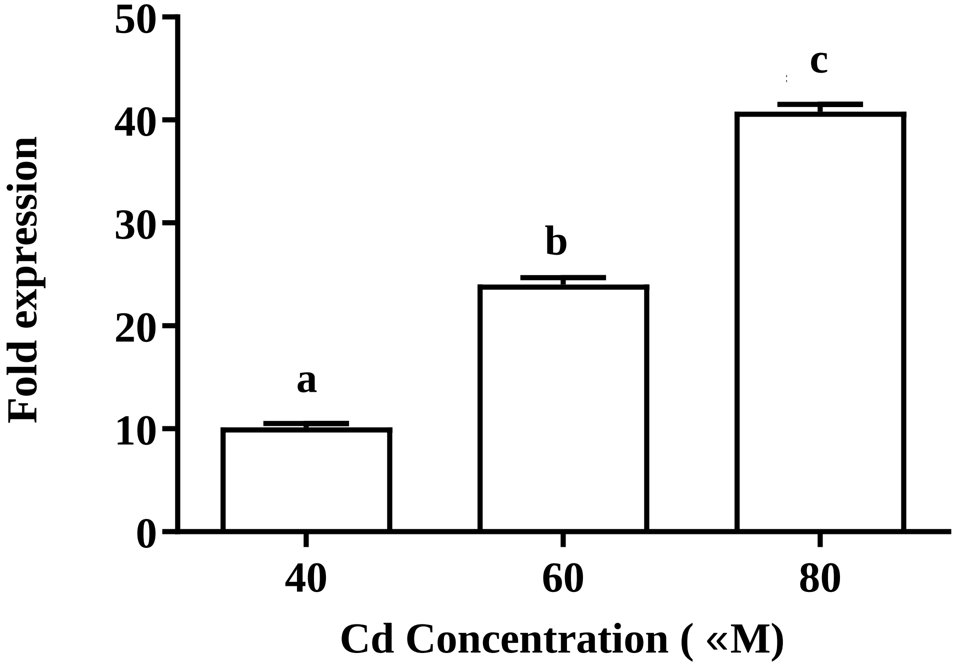


Figure 4.53: Fold increase in expression levels of cDNA PLCd43 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of $CdSO_4$. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCd43 without metal treatment). Bars sharing a common alphabet are not significant at P<0.05 (n=3). Error bars are ±SD

4.3.4.3. Multi-metal tolerance profiling of PLCd43

Role of cDNA PLCd43 in conferring tolerance to heavy metals such as Cd, Cu, Zn and Co was tested by heterologous complementation of PLCd43 into various metal hypersensitive mutant strains of *S. cerevisiae* i.e., *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$. Drop assay performed for specific tolerant transformants i.e., *ycf1*$^{\Delta}$/PLCd43, *cup1*$^{\Delta}$/PLCd43, *cot1*$^{\Delta}$/PLCd43 and *zrc1*$^{\Delta}$/PLCd43 had shown that PLCd43 provide tolerance to these yeast mutants against toxicity of potentially toxic metals used in this study (Figure 4.54).

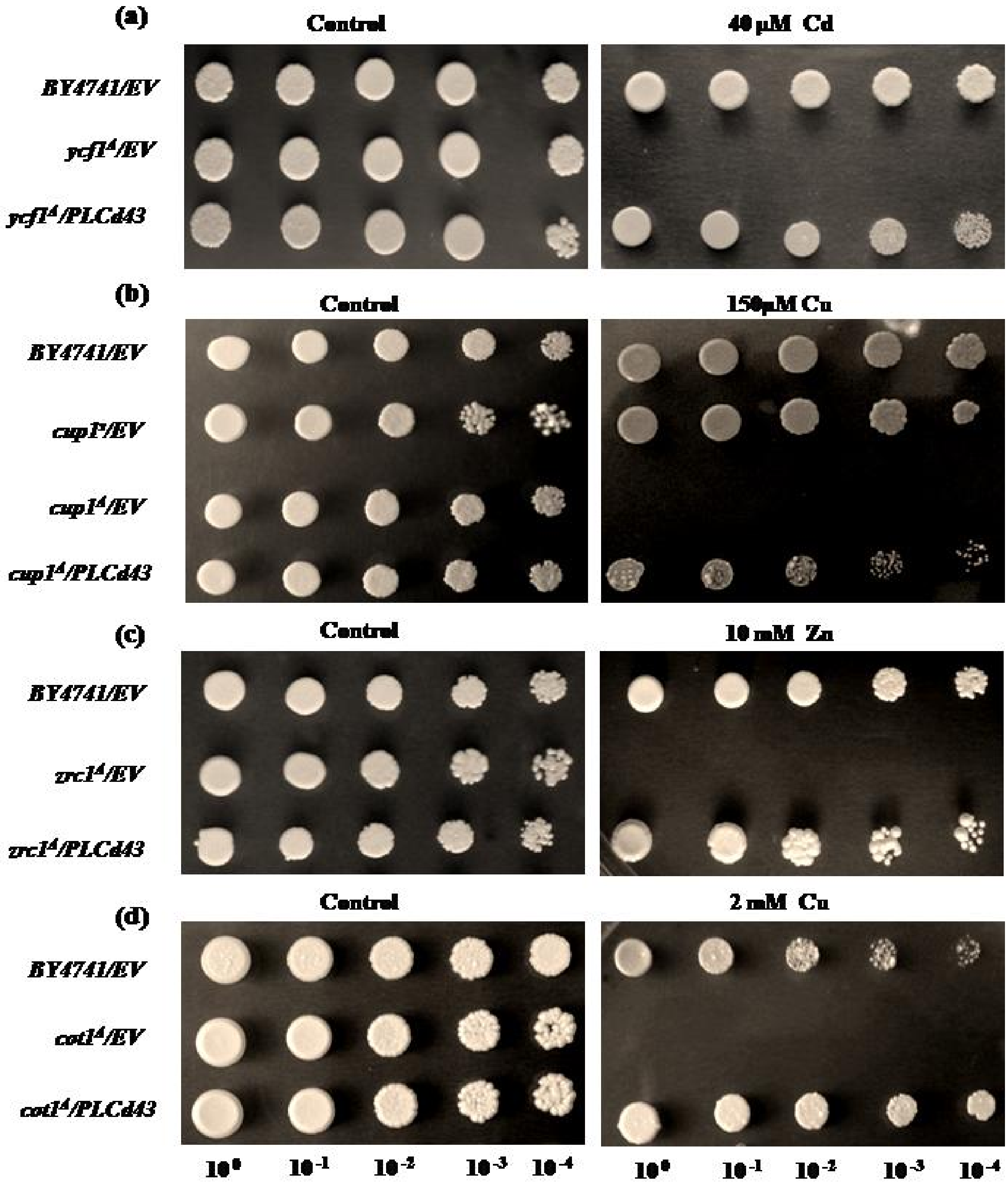


Figure 4.54: Drop assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLCd43 at various concentrations (a) $CuSO_4$ 150 µM (b) $ZnSO_4$ 10 mM, (c) $CoCl_2$ 2 mM. Wild strain BY4741/*EV*, *cup1*^{s}*/EV* serves as positive controls and *cup1*$^{\Delta}$*/EV*, *zrc1*$^{\Delta}$*/EV* and *cot1*$^{\Delta}$*/EV* as negative controls respectively. EV stands for empty vector pFL61

Prominent growth was observed on the metal amended media plate in case of *ycf1*$^{\Delta}$/PLCd43 in comparison to the mutant strains with empty vector while in case of Cu, moderate growth was observed for cDNA transformed yeast mutant strain in comparison to the mutant carrying empty vector. When PLCd43 was transformed into Zn sensitive strain (*zrc1*$^{\Delta}$/PLCd43) and Co sensitive mutants (*cot1*$^{\Delta}$/PLCd43), growth was observed on the media carrying respective metals. The dynamic role of cDNA PLCd43 in heavy metal tolerance had been further determined by monitoring the growth of yeast transformants in broth media amended with specific range of Cd, Cu, Zn and Co (Table 4.17). The transformant *ycf1*$^{\Delta}$/PLCd43 showed more tolerance at 40 µM and at 60 µM of $CdSO_4$ than at 80 µM of $CdSO_4$ in comparison to the mutant strain *ycf1*$^{\Delta}$pFL61 (Fig. 4.55a). Similarly, for Cu metal, the growth decreased with increase in Cu in the media (Fig. 4.55b). While for Zn metal, the growth was also decreased with increase in the Zn^{2+} concentration in yeast cells transformed with PLCd43 (Fig. 4.55c). Similarly, heterologous complementation of cDNA PLCd43 into *cot1*$^{\Delta}$, the tolerance level increased at 2 mM of $CoCl_2$ and decreased at higher Co concentration (Fig. 4.55d). Thus, cDNA PLCd43 conferred tolerance to different yeast mutants in an order Zn> Co> Cu> Cd and all these findings provided the evidence in support of cDNA PLCd43 as a prospective member of metal tolerant gene family.

Table 4.17: Effect of different metal concentrations on the growth of respective mutant strain carrying cDNA PLCd43 in comparison to both wild strain BY4741 (*cup1*^{s} in case of copper) and mutant strain with empty vector pFL61 (EV).

cDNA PLCd43			
	O.D. 600		
Cadmium (μM)	***ycf1***$^{\Delta}$***/EV***	***BY4741/EV***	***ycf1***$^{\Delta}$***/* PLCd43**
40	0.09 ± 0.01a	1.42 ± 0.116b	8.56 ± 0.21c
60	0.06 ± 0.01a	0.75 ± 0.29b	5.65 ± 0.10c
80	0.04 ± 0.006a	0.39 ± 0.18b	2.33 ± 0.12c
	O.D. 600		
Copper (μM)	***cup1***$^{\Delta}$***/EV***	***cup1***^{s}***/EV***	***cup1***$^{\Delta}$***/* PLCd43**
150	0.24 ± 0.032a	0.56 ± 0.012b	2.74 ± 0.16c
300	0.07 ± 0.007a	0.19 ± 0.03b	1.83 ± 0.083c
500	0.05 ± 0.001a	0.04 ± 0.001b	0.93 ± 0.094c
	O.D. 600		
Zinc (mM)	***zrc1***$^{\Delta}$***/EV***	***BY4741/EV***	***zrc1***$^{\Delta}$***/* PLCd43**
10	0.34 ± 0.07a	1.38 ± 0.19b	6.16 ± 0.203c
11	0.14 ± 0.03a	0.96 ± 0.03b	4.21 ± 0.05c
12	0.084 ± 0.014a	0.48 ± 0.09b	1.76 ± 0.19c
	O.D. 600		
Cobalt (mM)	***cot1***$^{\Delta}$***/EV***	***BY4741/EV***	***cot1***$^{\Delta}$***/* PLCd43**
2	0.26 ± 0.071a	1.49 ± 0.19b	5.49 ± 0.19c
3	0.14 ± 0.007a	0.97 ± 0.11b	3.26 ± 0.14c
4	0.05 ± 0.009a	0.14 ± 0.02b	1.66 ± 0.13c

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

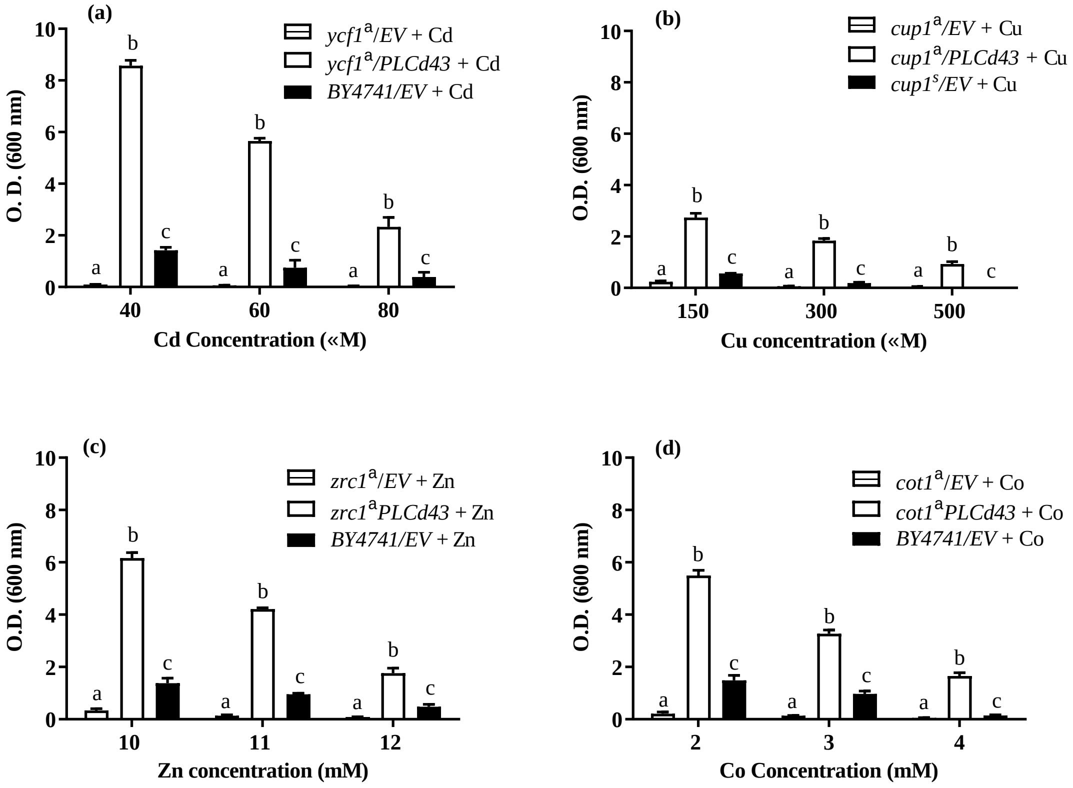


Figure 4.55: Growth assay of metal sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$expressing cDNA PLCd43 in SD-Ura medium supplemented with (a) $CdSO_4$ 40-80 µM, (b) $CuSO_4$ 150-500 µM, (c) $ZnSO_4$ 10-12 mM, (d) $CoCl_2$ 2-4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at $P<0.05$ (n=3).

4.3.4.4. Yeast metal uptake by cDNA PLCd43

Accumulation of metal by respective yeast transformant carrying PLCd43 was further determined by performing ICP-MS analysis for acid digested products of one gram of cell pellet of cDNA PLCd43 transformed yeast cells. It was observed that PLCd43 transformed cells helped in accumulated a good amount of metal (Table 4.18). For Cd metal, level of metal accumulation increased in *ycf1*$^{\Delta}$/PLCd43 up to 60 µM of $CdSO_4$

followed by a decrease at 80 µM of $CdSO_4$ as compared to wild strain BY4741/EV as well as mutant strain *ycf1*$^{\Delta}$*/EV* in all Cd concentrations (Fig. 4.56a).

Table 4.18: Accumulation of different metals in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLBe1 in comparison to both wild strain BY4741 and mutant strain with empty vector pFL61 (EV).

cDNA PLCd43			
Cadmium uptake (mg^{-1}g dry weight)			
Cd (µM)	***ycf1*$^{\Delta}$/ PLCd43**	***BY4741/EV***	***ycf1*$^{\Delta}$*/EV***
40	1.22 ± 0.026a	0.15 ± 0.012b	0.015 ± 0.001c
60	1.04 ± 0.035a	0.25 ± 0.002b	0.025 ± 0.010c
80	1.26 ± 0.015a	0.05 ± 0.0006b	0.009 ± 0.001c
Copper uptake (mg^{-1}g dry weight)			
Cu (µM)	***cup1*$^{\Delta}$/ PLCd43**	***BY4741/EV***	***cup1*$^{\Delta}$*/EV***
150	0.350 ± 0.010a	0.36 ± 0.013a	0.026 ± 0.001b
300	0.148 ± 0.007a	0.22 ± 0.005b	0.022 ± 0.001c
500	0.102 ± 0.003a	0.16 ± 0.009b	0.008 ± 0.001c
Zinc uptake (mg^{-1}g dry weight)			
Zn (mM)	***zrc1*$^{\Delta}$/ PLCd43**	***BY4741/EV***	***zrc1*$^{\Delta}$*/EV***
10	1.38 ± 0.026a	0.76 ± 0.066b	0.05 ± 0.005c
11	1.45 ± 0.027a	0.31 ± 0.006b	0.03 ± 0.0005c
12	0.77 ± 0.027a	0.21 ± 0.01b	0.02 ± 0.001c
Cobalt uptake (mg^{-1}g dry weight)			
Co (mM)	***cot1*$^{\Delta}$/ PLCd43**	***BY4741/EV***	***cot1*$^{\Delta}$*/EV***
2	1.23 ± 0.026a	0.04 ± 0.001b	BDL
3	0.92 ± 0.006a	0.05 ± 0.001b	BDL
4	0.47 ± 0.027a	0.013 ± 0.001b	BDL

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

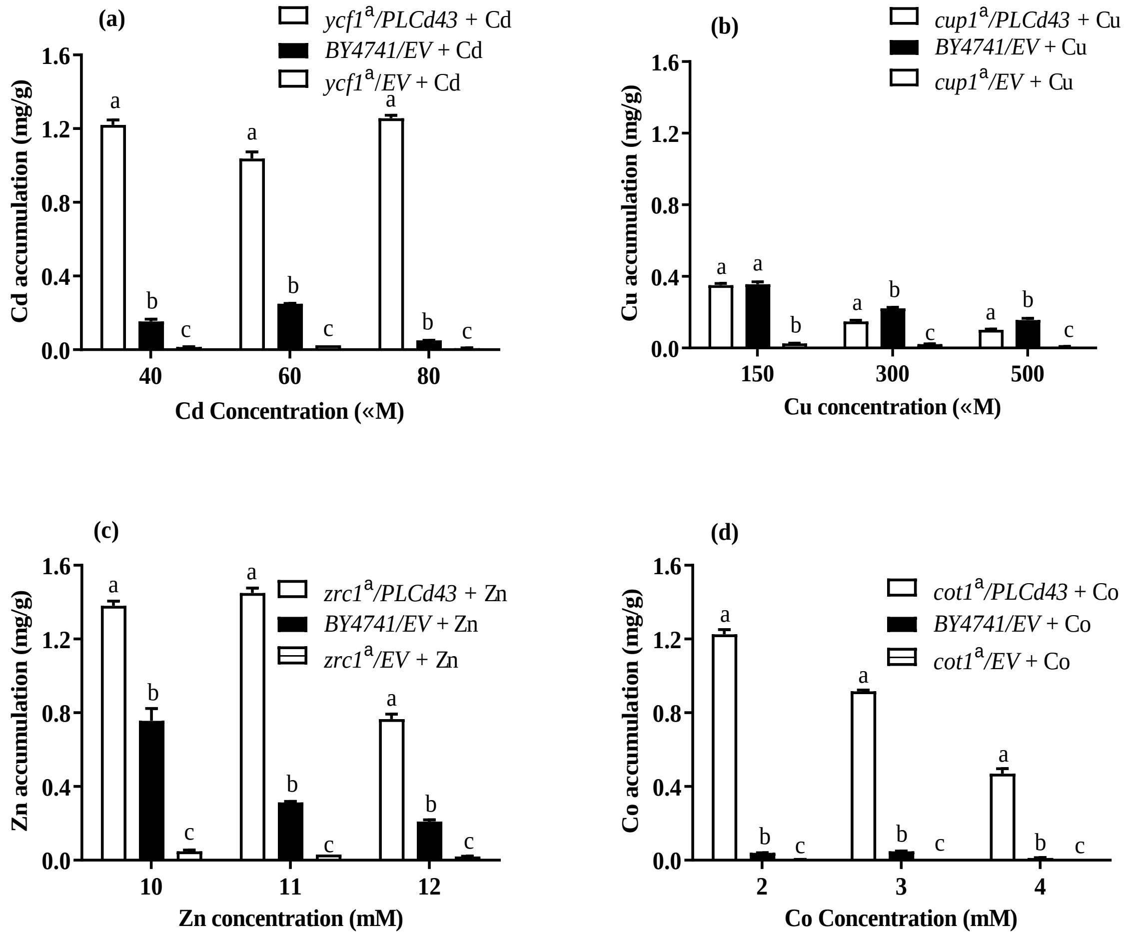


Figure 4.56: Accumulation of a) Cd^{2+}, b) Cu^{2+}, c) Zn^{2+} and d) Co^{2+} in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCd43. Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD. Bars sharing a common letter within the concentrations not significant at P<0.05 (n=3).

Concurrently, the yeast transformant *cup1*$^{\Delta}$/PLCd43 showed very less accumulation of Cu metal at all concentration used in this study while for Zn metal, the level of accumulation increased until 11 mM of $ZnSO_4$ with a decrease at 12 mM of $ZnSO_4$ concentration (4.56b & c). Similarly, for Co metal, decrease in metal accumulation with increase in metal concentration was observed (Fig. 4.56d). All above findings from accumulation studies further provided the evidence in support of cDNA PLCd43 as an efficient member of gene family with good metal accumulation efficiency.

Phytochelatin synthase (PC synthase) is an enzyme involved in the biosynthesis of phytochelatins (PCs) from glutathione and the biosynthetic pathway is induced by direct contact of PC synthase with the PTMs for detoxification of toxic elements (Liu *et al.* 2011; Hayashi *et al.* 2020). Various *in vitro* studies have verified the role of PTMs such as Cd^{2+}, Cu^{2+}, Ag^{+}, Hg^{2+}, Zn^{2+} and Pb^{2+} ions in activating the PC synthase enzyme and the top activator tested was Cd^{2+} followed by Zn^{2+}, Bi^{3+}, Ag^{+}, Cu^{2+}, Hg^{2+} and Au^{+} (Clemens *et al.* 1999; Ha *et al.* 1999, Vatamaniuk *et al.* 1999; Liu *et al.* 2011). Clemens *et al.* (1999) reported a wheat cDNA *TaPCS1* playing the catalytic role in mediating phytochelatin synthesis and providing tolerance against Cd metal. Ha *et al.* (1999) reported that these phytochelatin synthase (PC synthase) not only are activated by presence of PTMs but also prevent unnecessary interactions between the phytochelatins (PCs) and essential elements due to specificity of phytochelatin synthase towards every element. Phytochelatins (PCs) are member of family of metal complexing peptides, synthesized by transpeptidation of Glu-Cys moiety of glutathione molecule and possessing a general structure (c-Glu Cys)$_n$-Gly where n ranges from 2 to 11 (Rauser, 1995; Hall, 2002). Phytochelatins are known to play significant role in maintaining the metal ion homeostasis through chelation hence regulates the metal ion availability among plant cells (Guo *et al.* 2008). Chelation of metal ions by phytochelatins involves complex mechanisms where metal ions activates enzyme PC synthase and results into PC biosynthesis followed by formation of phytochelatin metal complex and sequestration in vacuole hence detoxification (Anjum *et al.* 2015). Although most of the studies have reported the PC synthase enzyme activity among plants such as pea, tomato and *Arabidopsis* only (Howden *et al.* 1995; Klapheck *et al.* 1995; Chen *et al.* 1997), few studies have also reported in other living organisms such as fungi (Gekeler *et al.* 1988; Piechalak *et al.* 2002). In present study, cDNA PLCd43 encoding for PC synthase like protein was observed to confer metal tolerance phenotype to all the metals used. Similar results were observed during over expression of Zn/Cd PC synthase from Zn/Cd hyperaccumulator *Thlaspi caerulescens* in tobacco plant for heavy metal tolerance and accumulation. Gasic and Korban (2007) observed that heterologous expression of PC synthase gene from *A. thaliana* results into elevated level of tolerance against Zn and Cd metal in *Brassica juncea*. Over-expression of PC synthase genes in some lower organisms like bacteria and fungi significantly increases their potential towards metal tolerance. A gene from *Caenorhabditis elegans CePCS1* similar to cDNA PLCd43 had been reported to confer tolerance to Cd hypersensitive *ycf1* mutant of *S. cerevisiae* and a PC synthase-lacking mutant of *S. pombe* towards Cd metal. Henceforth, various experimental findings in this study have classified cDNA PLCd43 encoding PC synthase like enzymatic protein as a potential member of metal tolerant gene family playing critical role in combating heavy metal toxicity in living organisms.

4.4. Functional eukaryotic diversity from metal contaminated soil

Soil sample collected from PL site in France was determined for the functional diversity of the site by performing NGS analysis of amplicons derived from gene pool of the site. This part of study illustrated about the total functional eukaryotic diversity of the soil that was active part of various phenomena or mechanisms occurring in the metal polluted environment.

4.4.1. Amplification of V4 region of 18S rRNA/DNA

The diversity of soil samples of PL region was determined by NGS analysis of V4 region of 18S rRNA/DNA by using high throughput sequencing technique i.e., Illumina-Miseq. The V4 region of 18S rRNA/DNA was amplified to yield good amount of cDNA and gDNA products (Figure 4.57).

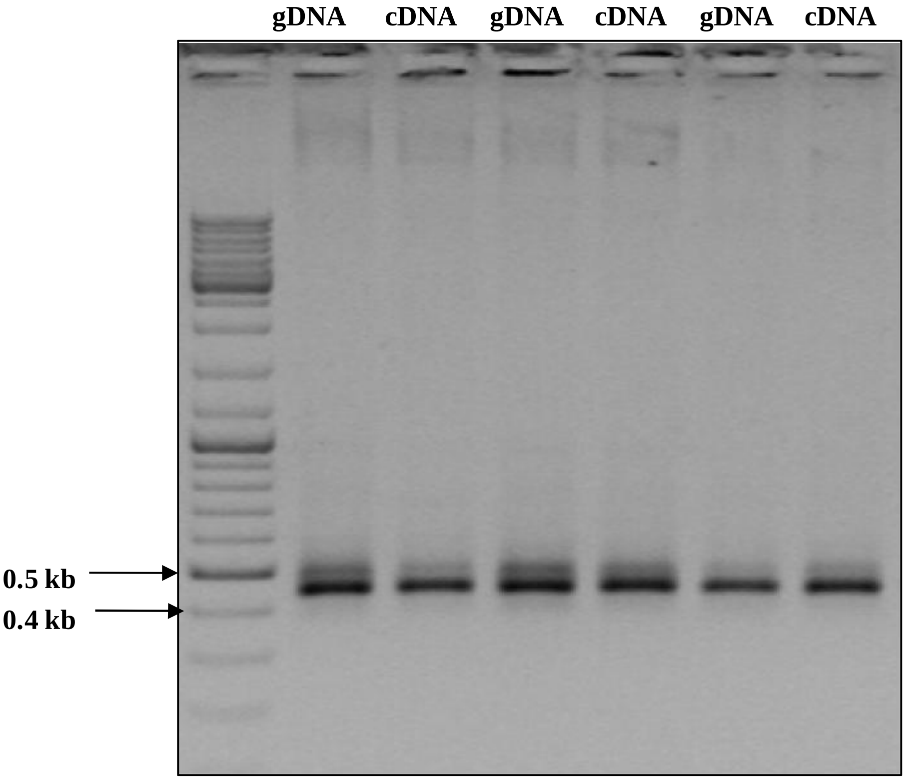


Figure 4.57: Agarose gel electrophoresis of amplified V4 region of 18S cDNA and 18s rDNA. Each well loaded with genomic DNA (gDNA) and cDNAs product of three replicates of soil samples. DNA marker of size 1 kb was run along with cDNA and gDNA amplified products.

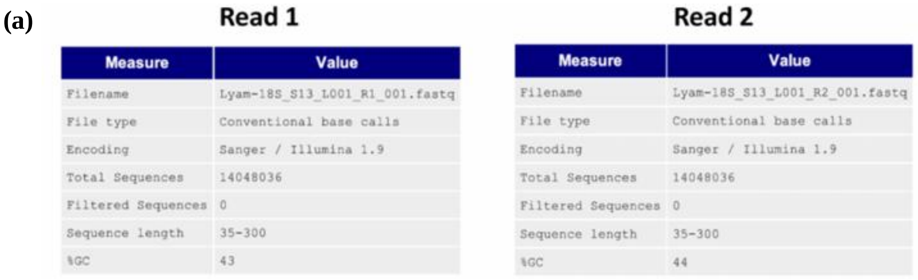
Each forward and reverse degenerate primer was designed with specific 8 nucleotide long adapter-tag at their 5’ends. Total 14 specific primer sets were used to amplify the genomic DNA templates and cDNA templates in duplicates. Three genomic DNA templates and three cDNA templates represented three biological replicates of each soil sub samples. Therefore, each sample was represented by three biological replicates and two technical replicates. PCR was performed in 100 µl reaction volume with 30 ng of template DNA and amplification with each primer set gave similar results as shown in figure 4.57 where the amplified DNA bands were ~400 bp in size in the agarose gel. All 14 PCR products were mixed in equimolar concentration and sequenced. To generate NGS data, a paired end (PE) sequencing chemistry was used in Illumina Miseq (2x250bp) sequencing platform which further yielded raw reads of equal length.

4.4.2. Data analysis

4.4.2.1. Quality analysis of reads: FastQC

First and foremost step in analysis was to check the quality of raw reads using FastQC which includes GC bias, K-mer quality; duplication levels, adaptor content etc. FastQC report revealed that Read1 (R1) file was of good quality in comparison to Read2 (R2) file. The reason behind this could be as the sequences also contained RNA derived sequences (Conesa *et al.* 2016). The first analyses result of FastQC is basic statistics which gives simple information about the various properties of FastQ file such as name, total sequences, filtered sequences, read length and GC content etc (Figure 4.58a). In the present data, R1 and R2 contained 1,40,48,036 total sequences, in which filtered sequences were nil with length of each sequence is between 35-300 bp. GC content for both the files were 43% and 44% respectively. Second part of FastQC analysis is the most important analysis module depicted the "per base sequence quality" plot of the NGS data in which a box and whisker plot showed the distribution of quality score at each

(a)



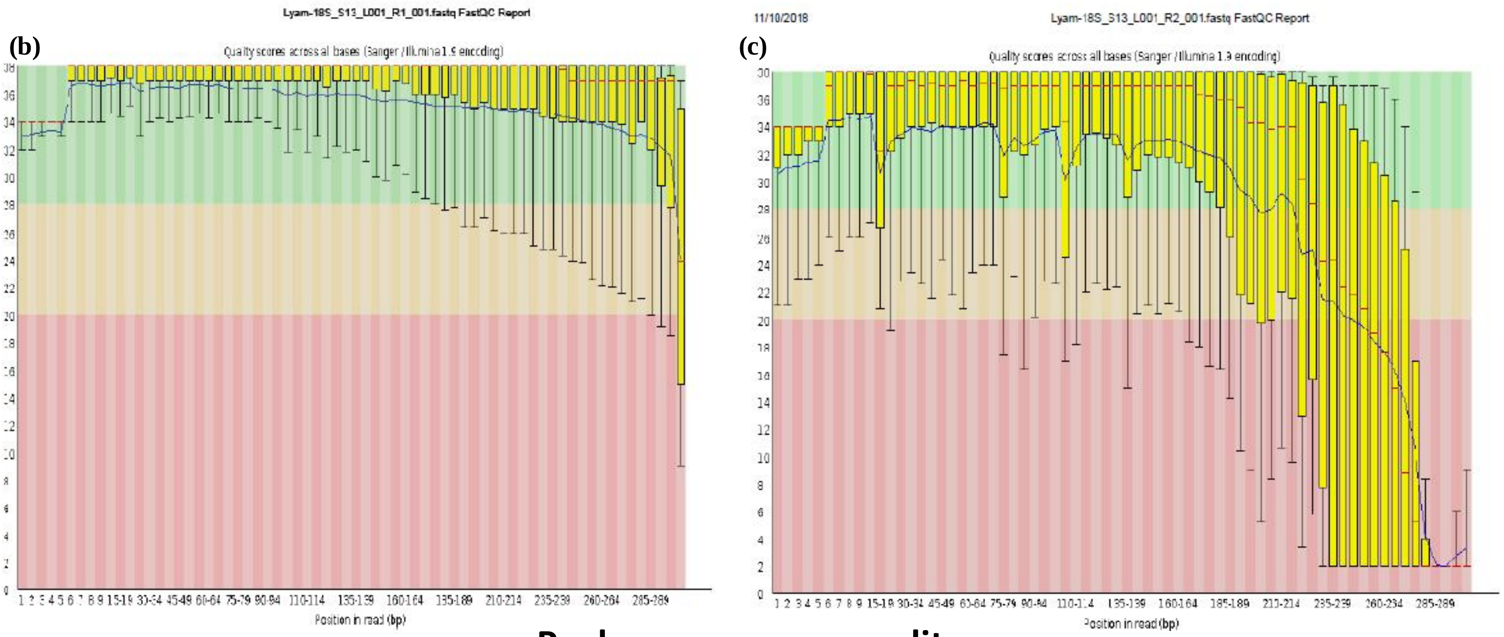
Read 1

Measure	Value
Filename	Lyam-18S_S13_L001_R1_001.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	14048036
Filtered Sequences	0
Sequence length	35-300
%GC	43

Read 2

Measure	Value
Filename	Lyam-18S_S13_L001_R2_001.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	14048036
Filtered Sequences	0
Sequence length	35-300
%GC	44

Basic statistics



Per base sequence quality

Figure 4.58: FastQC report of Read 1 and Read 2. (a) Basic statistics of the run, (b) Per base sequence quality of Read 1 and (c) per base sequence quality of Read 2. Yellow box represents the median quality score at each base window; blue line is mean quality score at each base window and red line within each yellow box is median quality score at that base window.

position along the reads in both file R1 and R2. In figure 4.58 b & c, a fall in the quality towards the end of the read length of both the reads was observed (after 285 bp in R1 and after 210 bp in R2) which was not unexpected and could occur due to signal decay or phasing (Ledergerber and Dessimoz, 2011). In the box plot, yellow box represents 25th and 75th percentiles where as the red lines within this box represented median and the whiskers represent 10th and 90th percentiles. The blue line denotes the average quality score for the nucleotide. This module of FastQC was the most crucial step as it also provided information about any error occurred during sequencing. For a better interpretation of FastQC report other factors of the Illumina data were also analyzed such as per sequence quality scores, per base sequence content etc (Fig 4.59, 4.60 & 4.61). Per base sequence quality score provided the average quality score among the total number of reads which mean it depict the quality of subset of the sequences. It is a plot of total number of sequences at y-axis and the average quality score for these sequences at x-axis. Read1 has a good per base sequence score as no bump is there for any range of sequences while in R2 a small bump in the end of the sequences though does not represent a large proportion hence can be considered and checked for other modules in FastQC for further interpretations about R2 (Figure 4.60a & b).

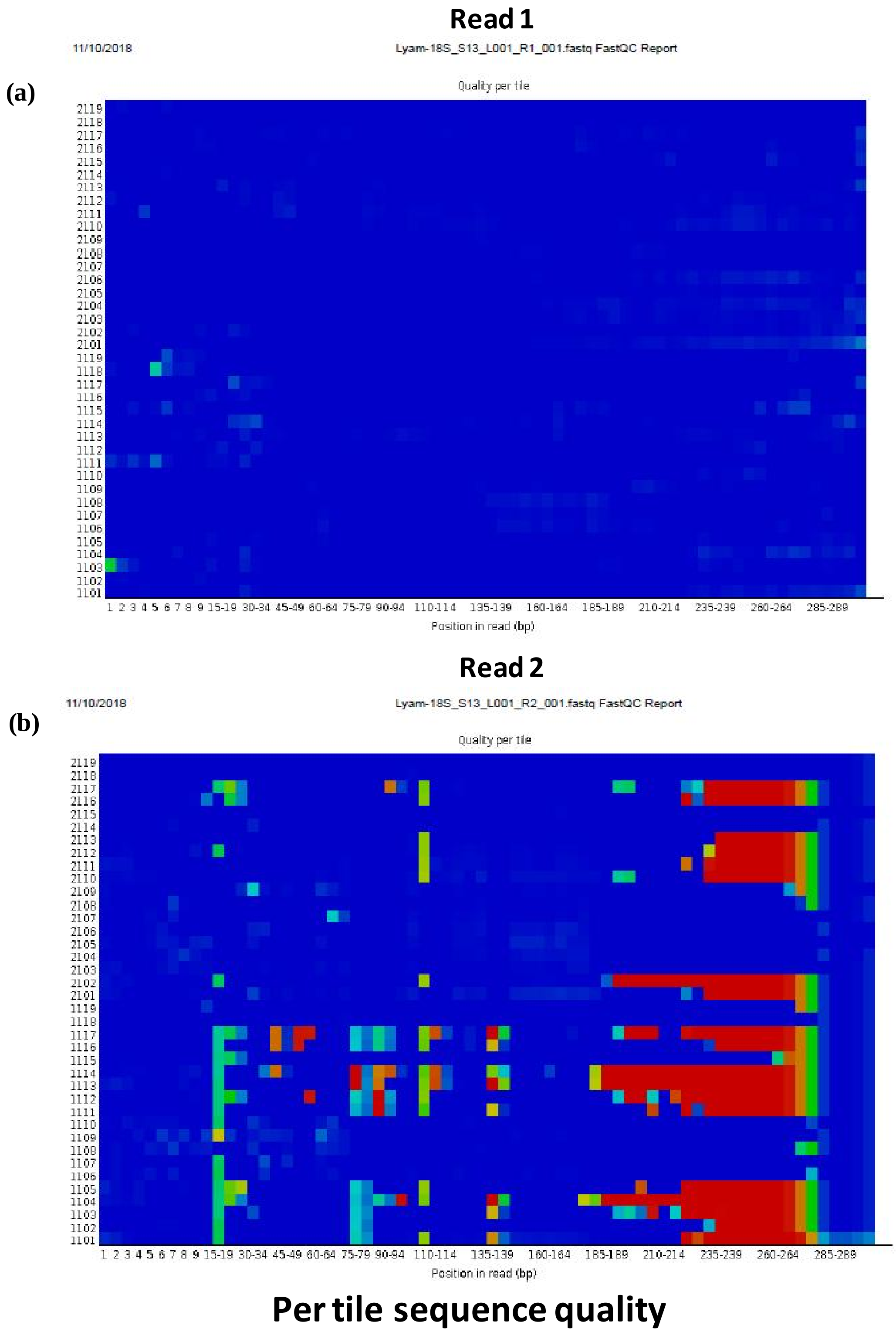
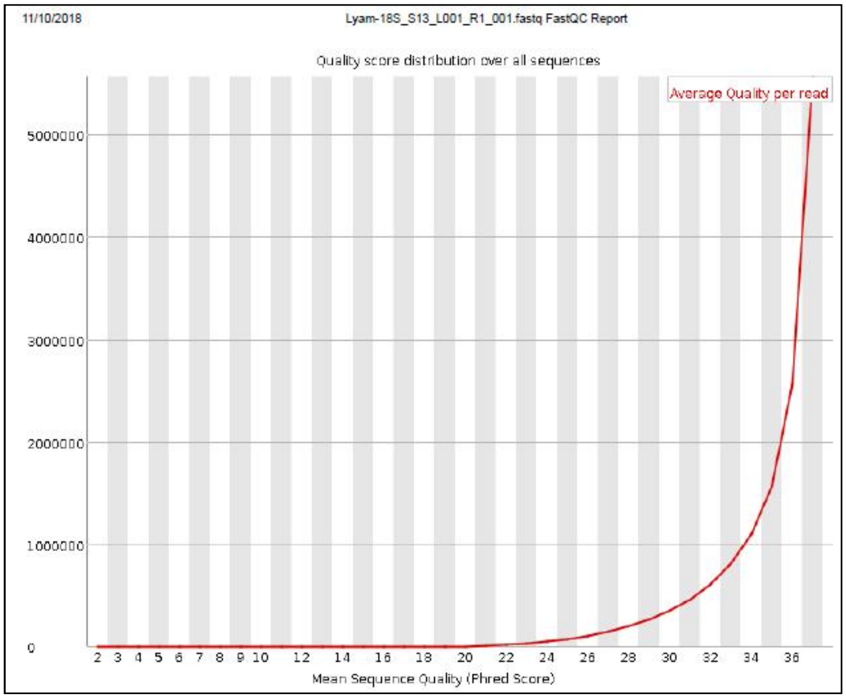
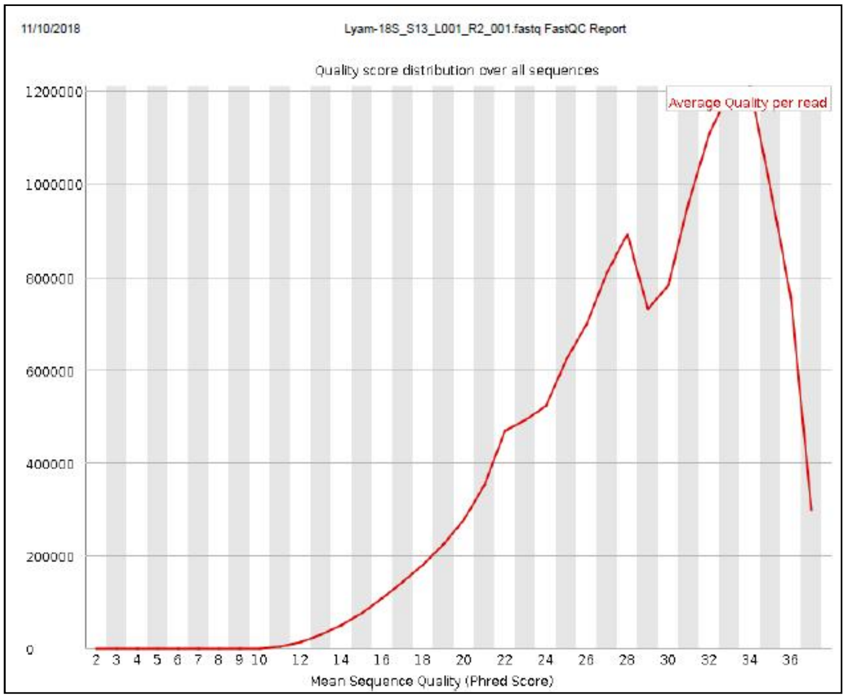


Figure 4.59: FastQC report of Read 1 and Read 2. (a) Per tile sequence quality for Read 1 (b) for Read 2

(a) **Read 1**



(b) **Read 2**

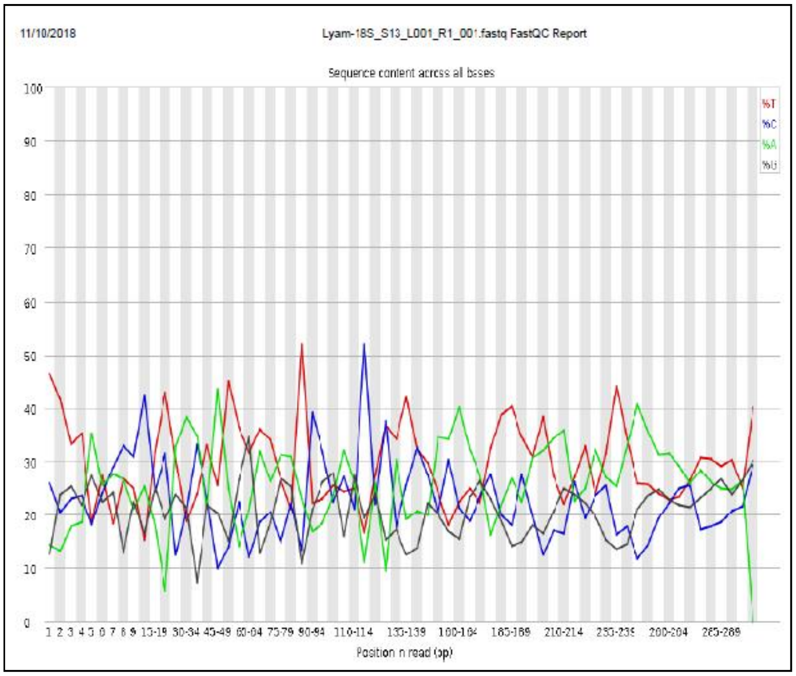


Per sequence quality score

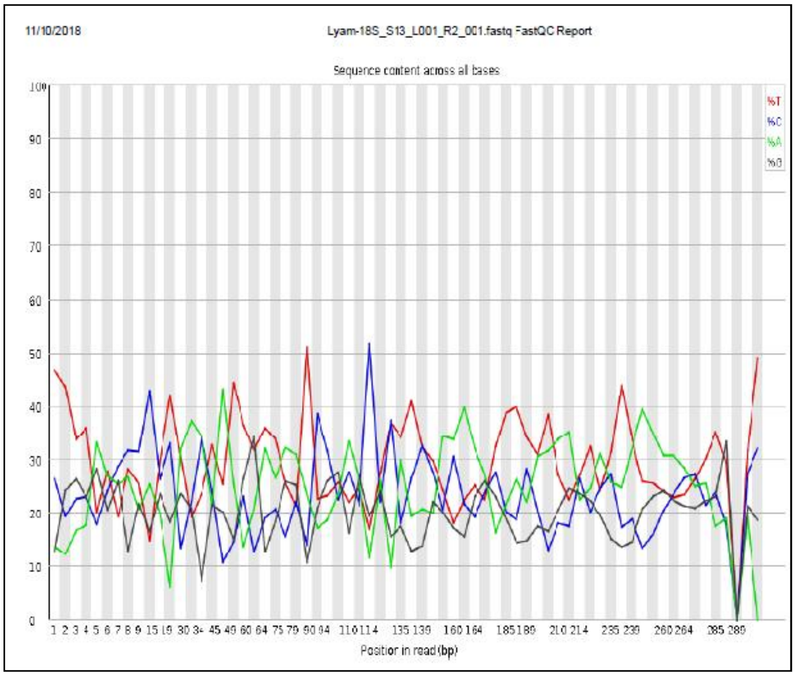
Figure 4.60: FastQC report of Read 1 and Read 2. (a) Per sequence quality score for Read 1 (b) for Read 2

Similarly, “per base sequence content” of the data reported the percent of bases for each of the four normal nucleotides at each position across all the reads in the file. The “per base sequence content” for NGS data was “failed” for both the reads R1 and R2 (Figure 4.61a & b). It was usually expected in case of RNA-seq where the first 10 or 15 bases belong to random hexamer priming occur during the library preparation and data showing this kind of non-uniformity in base composition is always classified as “Failed” by FastQC, even if the sequence are of perfectly good quality. The relative amount of each base reflected the overall quantity of these bases in NGS datasets and results of this module also interpret any error occurred either during library preparation or sequencing. Any imbalance between the relative abundance of each base indicate the contamination of the library with the over-represented sequences and if a bias is consistent among all the bases, it indicates that the original library was either sequence biased or there is some error while sequencing. Per sequence GC content is also an important module of FastQC report which interpreted the percent of GC content present along the sequence read. In this NGS sequence data, plots for both the reads showed a good amount of GC content i.e., 43% and 44% for R1 and R2 respectively and plots were observed to be a little wider for R1 and little narrow for R2 to the threshold distribution (Figure 4.62a & b). In RNA sequencing, there may be a wider or narrower distribution of the mean GC content which results into such variations from the idealized normal distribution.

(a) **Read 1**



(b) **Read 2**

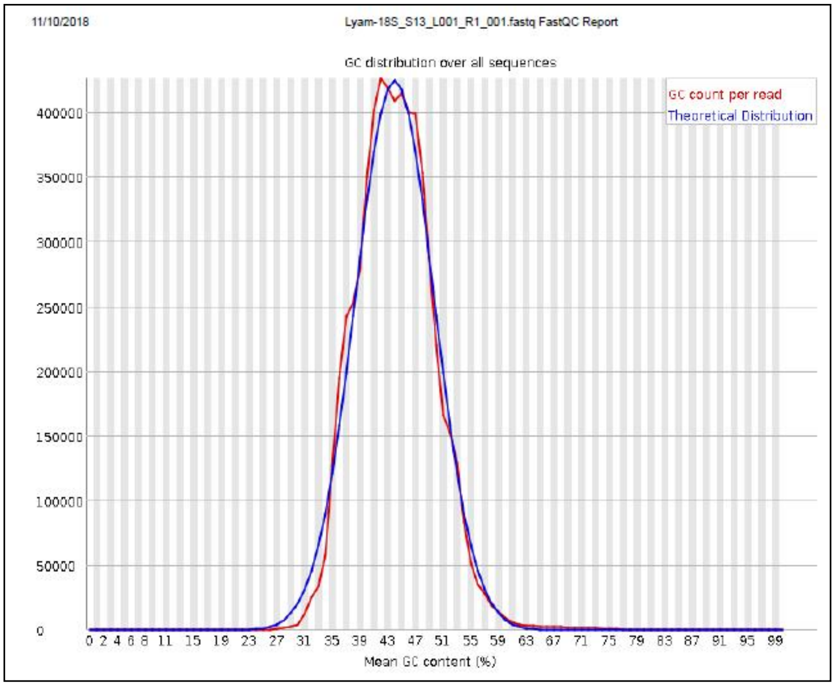


Per base sequence content

Figure 4.61: FastQC report of Read 1 and Read 2. (a) Per base sequence content for Read 1 (b) for Read 2

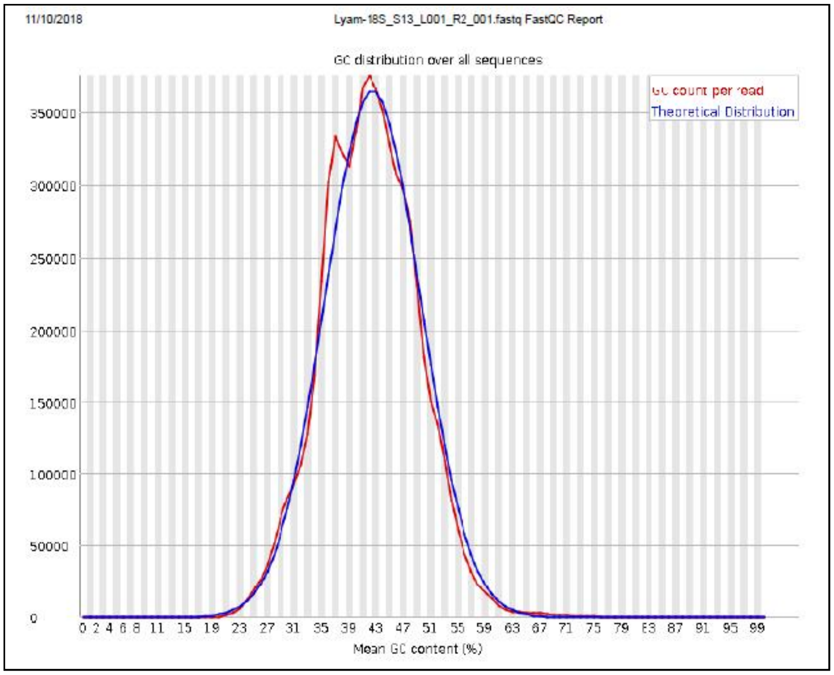
Read 1

(a)



(b)

Read 2



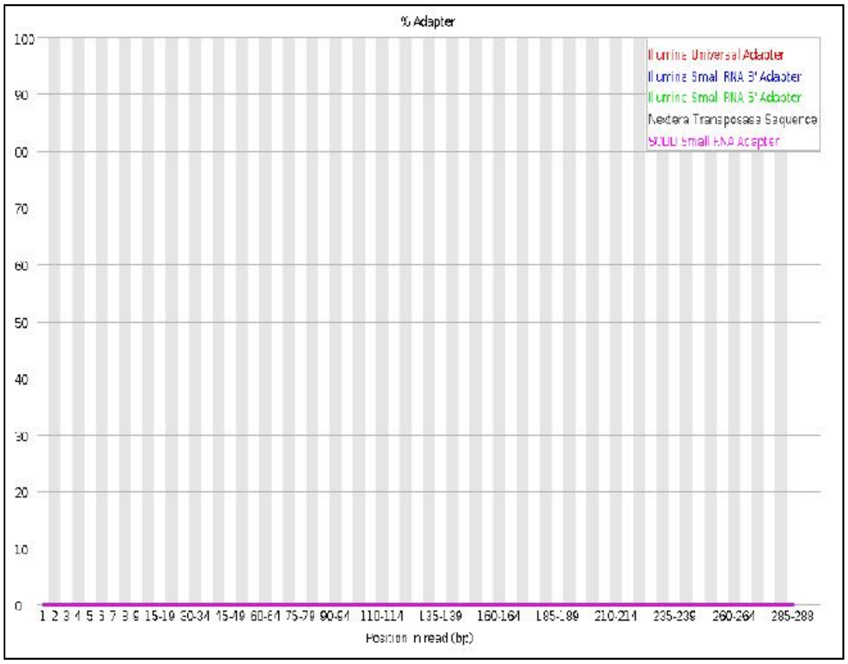
Per sequence GC content

Figure 4.62: FastQC report of Read 1 and Read 2. (a) Per sequence GC content for Read 1 (b) for Read 2

Another module of FastQC was adaptor content which comprised of cumulative plot of fraction of reads where no adaptor sequence specific to the library read R1 and R2 were identified at the specific base position (Figure 4.63). In this module, only library specific adaptors are searched. This interpretation reflected both the NGS reads R1 and R2 as ideal Illumina sequence data. Kmer content module of FastQC interpreted the count of short nucleotide of length k (k= 7) at each position along the length of each read. This module count the enrichment of every 7-mer within the sequence data. It reported the list of k-mers appeared at specific positions with greater than expected frequency. It first calculated the expected level of appearance of each k-mer based upon the base content of the whole library and then used the actual count to calculate the observed/expected ratio for that k-mer. In this work, the positions for the six most biased k-mers were plotted for each library read i.e., R1 and R2 (figure 4.64a & b). Furthermore, when checked for any over-represented sequences, it was found that none of the reads had any over-represented sequences. From all above interpretations of different modules of FastQC, it was concluded the quality of R1 was more in comparison to the R2 which further supported NGS sequence data as an ideal data for other analysis.

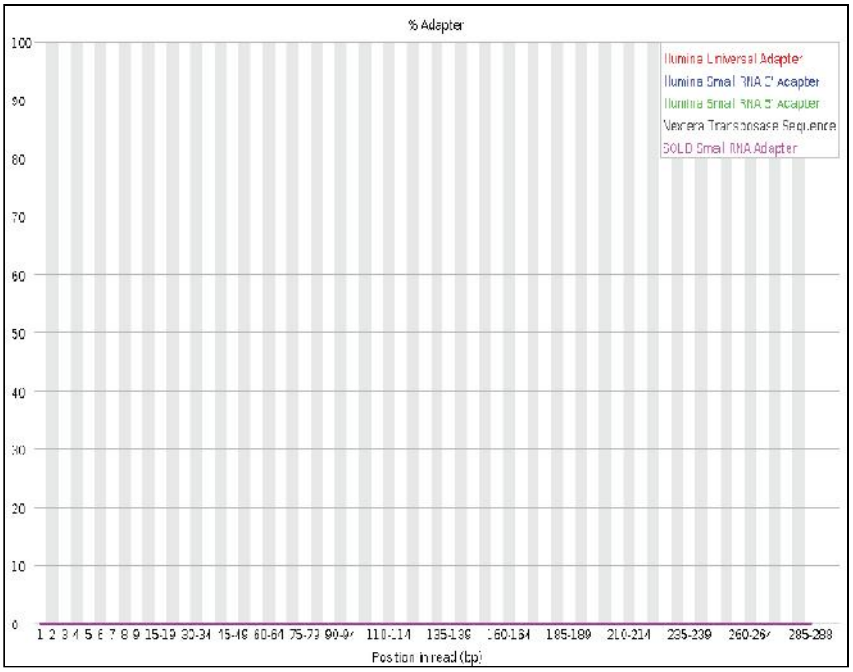
Read 1

(a)



Read 2

(b)

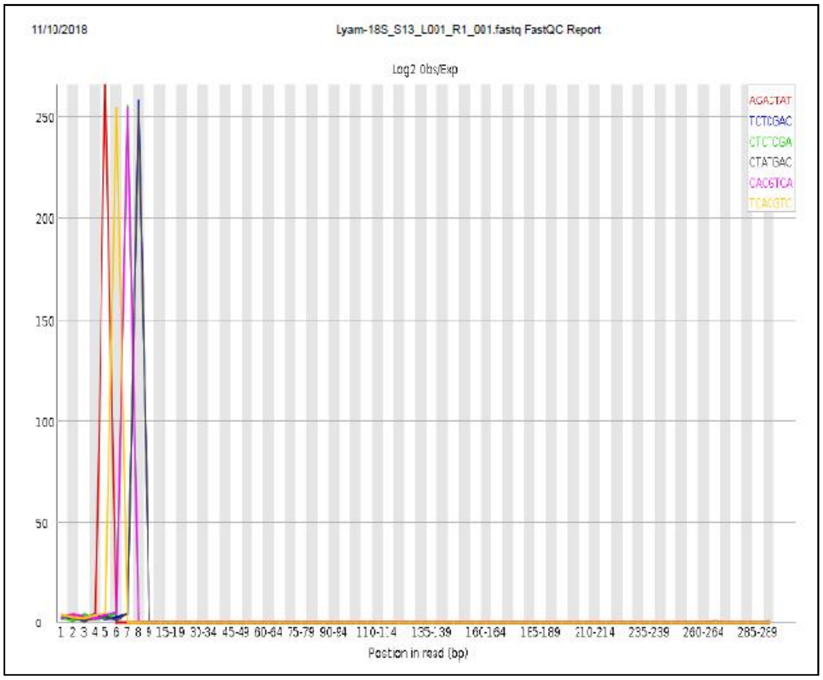


Adaptor content

Figure 4.63: FastQC report of Read 1 and Read 2. (a) Adaptor content for Read 1 (b) for Read 2

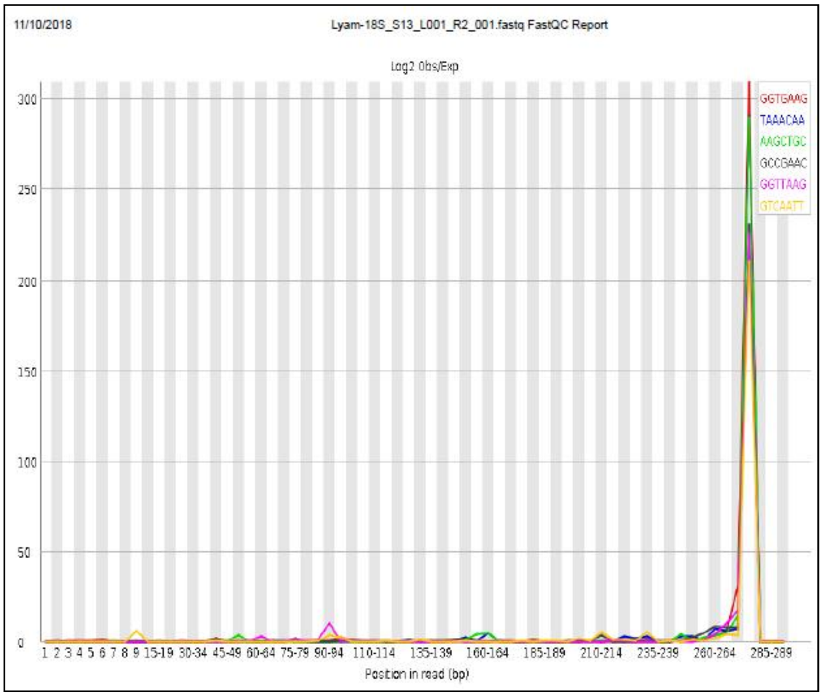
Read 1

(a)



Read 2

(b)



Kmer content

Figure 4.64: FastQC report of Read 1 and Read 2. (a) K-mer content for Read 1 (b) for Read 2

4.4.2.2. Data analysis via Mothur

After the quality check of the both the reads, the paired-end sequencing on Illumina MiSeq of samples showed that all cases had presence of eukaryotic DNA which yielded data of raw sequences with unique sequences produced from Mothur analysis. The final data of sequence reads (>300 base pairs) was obtained after excluding low-quality sequence reads, preclustering and removing chimeras, if any. The final cluster was generated with operational taxonomic units (OTUs) of 97% similarity which were assigned to phyla, families and genera. The findings in figure 4.65 suggested the presence of all the major eukaryotic phyla such as Metazoa, Fungi, Cercozoa, Conosa, Ciliophora, Streptophyta, Ochrophyta, Stremenopiles, Lobosa, Apicomplexa and Chlorophyta in the metal polluted soil PL. It was observed that the majority of the eukaryotes in this metal polluted soil are dominated by members of Metazoa, Fungi, Cercozoa and Ciliophora. Various studies observed that eukaryotic microbial communities from various environments such as soils, aquifers, brackish water, alpine lakes and marine sediments embrace the representative species of almost all the major eukaryotic lineages though each of them is characterized by the presence of specific eukaryotic phyla (Grossmann *et al.* 2016; Groussman *et al.* 2020). For instance, Tedersoo et al. (2014) reported the dominance of filamentous fungi and Amoebozoa in a soil environment in comparison to the environments like lakes which are mainly characterized by the abundance of Ciliates and Chytrids. Further findings in figure 4.65, also reported higher expression level of genes corresponding to eukaryotic microbial communities active at the time of sampling, which further suggested that these eukaryotic microbial communities, through various

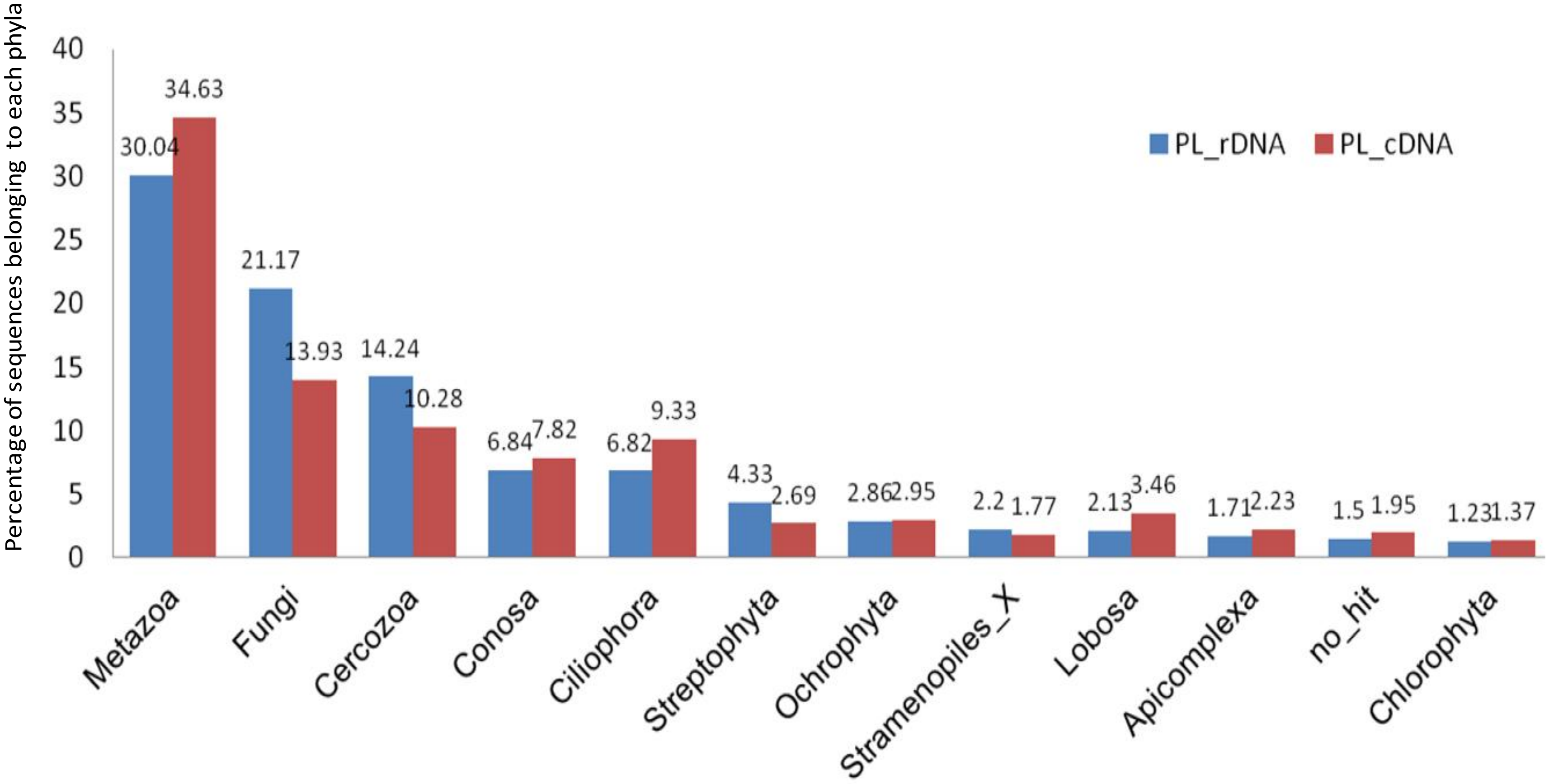


Figure 4.65: Taxonomic annotation of both 18S rDNA and cDNA sequences obtained from soil sample from PierreLaye (PL), France. All of the major eukaryotic phyla are represented in the analyzed datasets where Fungi and Metazoa phyla are predominant

novel elements, species or mechanisms, are combating various adverse conditions due to metal toxic environment. Furthermore, sequences of 14 amplified sequences samples were divided into two groups *viz.* Group A and Group B. It was observed that the relative abundance was equal in both groups A and B, with the classification of sequences by alignment to reference database SILVA at 97% similarity revealing sequences that were not aligned to reference at different taxonomic levels. Further, we also compared the microbial community across the different sub-samples of PL soil. A heat map was plotted to estimate the similarities of the samples (Figure 4.66). The blue part denotes low relative abundance across a genus, and the red denotes high relative abundance. The above findings are in agreement with the PCA plot wherein two components represent about 99% of the difference between the samples (Figure 4.67). In conclusion, all these findings also shown that various soil abiotic and biotic factors had more impact on eukaryotic diversity flourishing in the metal polluted soil site. Various global studies have confirmed that majority of the eukaryotic microbes are not ubiquitous and variations in distribution of species illustrated a strong bio-geographical outline at large geographical scales (deVargas *et al.* 2015, Talbot *et al.* 2014). Bates et al. (2013) observed that environmental factors such as availability of water, nutrients and soils can also affect the taxonomic distribution among the eukaryotic microbial communities. Thus, all these data and factors helps in understanding the environmental diversity and are essential to wisely collect the environmental sample with maximum coverage of different species and microbial communities flourishing in an environment.

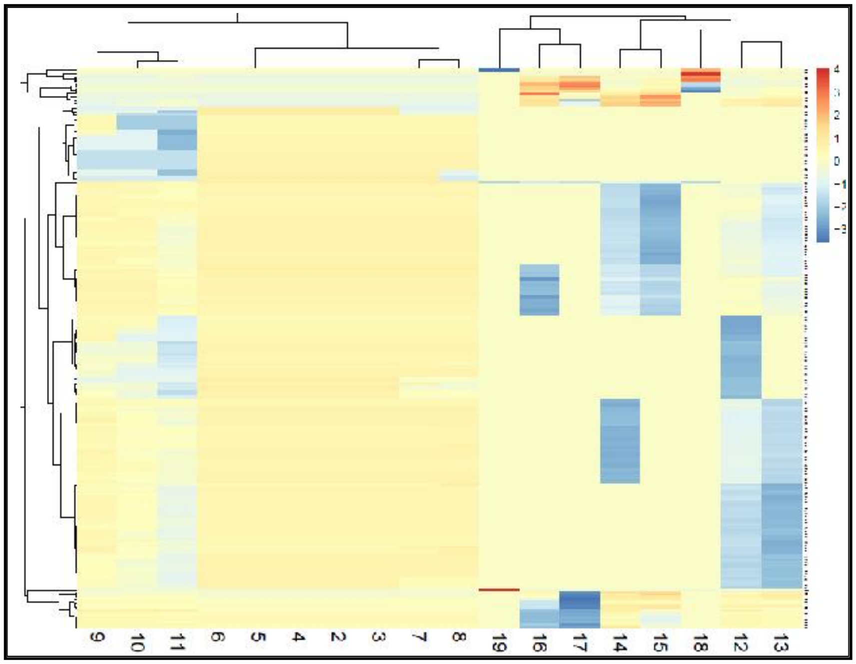


Figure 4.66: Heatmap showing eukaryotic diversity of agro-forestry soil contaminated with potentially toxic metals

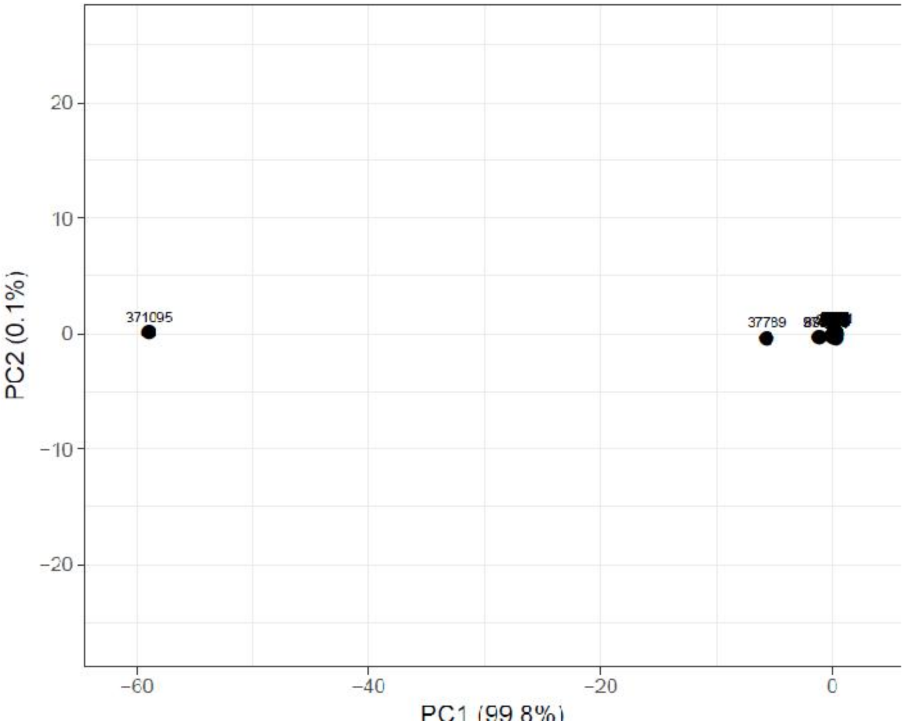


Figure 4.67: Principal component analysis of soil sample from agro-forestry soil contaminated with potentially toxic metals

Determining the functional and taxonomic diversity of microbial community flourishing in an environment is essential to understand almost all aspects of microbial ecology. In recent years, the advents of NGS technology accelerated this process and made it a common practice to target and amplify functional marker genes or phylogenetic genes to characterize their diversity across multiple samples. Among various type of NGS or amplicon sequencing platforms such as 454 (Sogin *et al.* 2006), Ion torrent (Junemann *et al.* 2012), PacBio (Fichot and Norman, 2013) and Illumina (Gloor *et al.* 2010), latter one is the most promising platform generating good quality of data with long reads to characterize microbial communities (Kozich *et al.* 2013). NGS technologies provided a revolutionary technique for countless applications with the competence to produce gigabases of sequence data in a single experimental run. These amplicon approaches are being progressively used for various genome and transcriptome analysis of various samples such as soils (Dumbrell *et al.* 2011), sediments (McKew *et al.* 2013) or aquatic systems (Papaspyrou *et al.* 2014) due to their speed, cost efficiency and high throughput nature (Mardis *et al.* 2008; Wang *et al.* 2009). Although, NGS sequencing has explored various samples for various microflora, yet eukaryotic microflora is still lagged behind due to variations in target gene copy number among the species e.g. copy number for 18S rRNA gene varies from 1 to >25000 (Prokopowich *et al.* 2003) or due to long multiple hypervariable regions for a gene, which were difficult to sequence for an older DNA sequencing platform (Amaral-Zettler *et al.* 2009). In case of eukaryotes, amplicons sequencing approaches are of great significance as adoption of various earlier sequencing methods in case of eukaryotes has been quite slow due to huge variations among the target genes, even for 18S rRNA gene, there was no known consensus where hypervariable region provides the most suitable representation of eukaryotic diversity (Bradley *et al.* 2016). Earlier studies have exploited various regions of other genes such as *rrn* operon (Lundholm *et al.* 2006, Guo *et al.* 2015), 18S and 23S regions, mitochondrial gene cytochrome c oxidase (Hebert *et al.* 2003), and rbcL gene from chloroplast, but all of these studies confined themselves to only a limited type of microbes in an environment (Vaulot *et al.* 2008; Eiler *et al.* 2013). Thus, 18S rRNA gene is most commonly preferred gene to study eukaryotic communities because it possess numerous alternative hypervariable regions i.e., V1-V9 and conserved regions (Hugerth *et al.* 2014; Hadziavdic *et al.* 2014). Numerous studies explored different variable regions within the 18S rRNA gene and have identified various primer combinations with the help of *in silico* sequence database coverage as well as taxonomic resolution and validate their feasibility with environmental surveys. Hence, recent development in the field of amplicon sequencing results into development of a broad category of eukaryotic primers designed for targeting hypervariable regions of 18S rRNA genes and concluded that V4 and V9 are the most efficient and frequently used regions (Hugerth *et al.* 2014; de Vargas *et al.* 2015, Stoeck *et al.* 2010; Pagenkopp *et al.* 2016). In our study, we used set of degenerate primers specific for V4 region of 18S rRNA gene with high coverage among the reference sequences to evaluate and quantify the eukaryotic microflora flourishing in the metal polluted soil. These degenerate primers were best suited for identifying the eukaryotic taxa through 18S rDNA using read lengths of approximately 300 bp obtained by Illumina Miseq. Sequencing of these reads provided huge amount of data which was further analyzed for taxonomic origin of these sequences.

Though NGS approaches producing high quality data, however with some inherent biases like erroneous sequences and noise, thus required sophisticated bioinformatics tools to check and quality filter the sequences to avoid any misinterpretation from the artifact results. Various sequence artifacts generated due to base calling errors, insertions/deletions, poor quality reads and contamination due to reads and primers or adaptors are among the most common errors in the NGS data which can significantly impact the downstream sequence analysis (Schmieder and Edwards, 2011). Thus, the first and foremost step in analyzing the NGS data is to quality check the data before any other downstream analyses like sequence assembly and alignments. FastQC is the most commonly used tool to check and provide an outline about basic quality control metrics of raw reads in NGS data. It is comprised of different modules that performed different analyses for data sets and highlighted the errors originated either during initial lab experimental work or in the sequencer (Andrews, 2010). Correspondingly, downstream statistical analysis also required a similar consideration as any wrong approach for analysis can generate false results and could be perceived as a hurdle in completing the research (Dumbrell *et al.* 2016). Recent advances in computational tools have addressed the numerous ecologically relevant questions. Development of tools such as ARB (Ludwig *et al.* 2004), DOTUR (Schloss, and Handelsman. 2005.), SONS (Schloss, and Handelsman. 2006) LIBSHUFF (Schloss, *et al.* 2004; Singleton *et al.* 2001), UniFrac (Lozupone and Knight, 2005; Lozupone *et al.* 2006), rRNA specific databases (Cole *et al.* 2009; DeSantis *et al.* 2006), microbial ecology had advanced from a descriptive field to an experimental endeavor. Though, these tools have been widely accepted and successful, but a number of limitations had limited their use. For instance, rRNA specific database provided various online tools such as aligners, classifiers and pipelines, which usually perform a general analysis which is further a limitation for an analysis focusing on a specific set of criteria in a dataset and transferring this huge amount of data set across the internet is itself a challenge for a researcher. Secondly, with the advancement in NGS techniques and expanding data size, it became a necessity to use more efficient algorithms by refactoring the existing software. Thus, to overcome these hurdles while analyzing a huge dataset corresponding to metatranscriptome of soil sample, we used the more sophisticated single analytical platform i.e., mothur which is more flexible and easily maintained software. This tool has the algorithms implemented in earlier tools along with additional features such as calculators for quantify main ecological parameters for measuring alpha and beta diversity, NAST based sequence aligner, visualization tools such as dendograms, venn diagrams and heat maps etc in conjunction with the facility to give individual commands either from within mothur using batch files or directly from command line providing greater flexibility in setting up pipelines (Schloss *et al.* 2009). Mothur has been used successfully in previous sequence analysis studies from various samples such as permafrost soil (Dan *et al.* 2013), Diatoms (Rivera *et al.* 2020) and fungal genes (Smith *et al.* 2011). In our study, it was found that majority of the significant eukaryotic phylum contributed actively in maintaining the homeostasis in metal polluted soil PL mainly dominated by metazoans and fungi. Fungi mainly constituted major proportion of microflora in the soil with significant involvement in material cycle of the soil ecosystem and similar observations were reported in our study thus depicted the active role of this phylum in adapting the adverse condition due to metal contamination in the sampling site. Soil microflora is known to develop various mechanisms to tolerate or resist the metal toxicity in a metal contaminated environment (Rathanayake *et al.* 2011). The mechanisms generally contributed towards the microbial resilience to metals such as Cd, Ni, Zn, Cu, Co, Pb and As involve the transfer of heavy metal tolerance genes by substitution across the metal sensitive strains and percentage of heavy metal bioavailability (Sobecky and Coombs, 2009; Griffiths and Philippot, 2013). Development of such metal tolerant species of fungi and other microbes indicated the deteriorated condition of the soil ecosystem. Studies exploring either 18S rRNA sequences amplified from soil DNA (O'Brien *et al.* 2005; Lesaulnier *et al.* 2008; Damon *et al.* 2012) or soil cDNAs (Urich *et al.* 2008; Damon *et al.* 2012) showed that soil hosts an incredible diversity of eukaryotes, all consistently show that the phyla Metazoa and Fungi dominate the soil eukaryotic communities. In one of these studies, the fungi contributed almost up to 40 percent of 18S soil sequences and 70 percent of the cDNA sequences (Damon *et al.* 2012). Thus, our findings showed that microbial diversity and community structure of sampling site was significantly influenced due to metal contamination of soil site which further significantly changed ecological environment of soil.

4.5. Conclusion

This study presented an attractive approach for construction of full length sized cDNA libraries with three different size ranges of cDNA molecules. In this study, the RNA content (gene expression) of the soil sample was explored through metatranscriptomic approach which gave an insight into microbial activity and their response towards environmental cues and the ratio of rRNA to DNA indicated the relative microbial activity in metal polluted soil. This work strategy involved the use of total RNA to synthesize mRNA derived cDNA libraries thus sidestep the need of various RNA enrichment protocols and limited the chances of RNA degradation and quantity loss. Next experimental strategy involved in this study was size fractionation of metatranscriptomic cDNA, which resulted into three different sized cDNA fractions and significantly bypassed various size based PCR biases as well as cDNA cloning and screening biases generally observed in recombinant DNA technology experiments.

This approach of metatranscriptomics reported four metal tolerant genes (PLBe1, PLCe10, PLCc43 and PLCd43) from two different sized mRNA derived cDNA libraries. First cDNA reported in this work, cDNA PLBe1 belongs to the Lib B (0.1 - 0.5 Kb) and encoded polypeptide marked presence of N-terminal ubiquitin domain and C-terminal extension of AN1 type Zn finger. Another cDNA screened for Cd tolerance belong to the largest library *i.e.,* Lib C which contained cDNA fractions ranged between 1-4 Kb size and showed homology with von Willebrand factor type D (VWD) of Vg protein. Further, cDNA PLCc43 from Lib C shared sequence homology with Hsp40 Type I proteins of *Acanthameoba castellanii* and the putative polypeptide displayed the fundamental structural features of Hsp40 type I protein as well. Fourth cDNA PLCd43 screened for Cd tolerance belongs to largest Lib C and shown sequence similarity with γ-Glutamylcysteinyl transferase that catalyses the synthesis of phytochelatins from glutathione molecules. Further studies revealed that all of the four cDNAs showed tolerance towards range of metals i.e., Cd^{2+}, Cu^{2+}, Zn^{2+} and Co^{2+} and their role in metal tolerance was further confirmed by their expression and metal accumulation studies. Amplicons sequencing of V4 region of 18S rRNA/DNA revealed high degree of eukaryotic diversity especially metazoan and fungi flourishing in soil site at the time of sampling.

Further, as in this work we have analyzed gene expression at a range of metal concentrations, which may be able to demonstrate exposure-related changes in the expression of particular genes within the range of different metals used in the study. The eukaryotic genes recovered from polluted soil through metatranscriptomic approach thus can be used as molecular probes to study unknown taxonomic groups and their contribution towards metal tolerance. These genes will have great biotechnological as well as ecological significance either as biomarkers or organism based biosensors. For example, measurement of vitellogenin levels among males has been used as biomarker to highlight the presence of estrogenic disruptors such as metals present in an aquatic environment (Matozzo *et al.* 2008; Jubeaux *et al.* 2008). Thus, all the above findings in this study supported the efficiency and potential of experimental design and methodology (size fractionation, heterologous complementation, 18S rRNA/rDNA amplicons sequencing) followed in this approach to explore various other environments for numerous unknown microbes, mechanisms and novel elements to remediating the metal polluted sites.

Introduction of genes with a capacity to detoxify the potentially toxic elements present in a polluted environment is considered as an eco-friendly, nonintrusive and aesthetically pleasing remediation technology that removes metal pollutants from the contaminated sites. It will also help us to 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 but they do not come hurdle-free. Finding suitable host for expression of environmental gene other than yeast can be a problem but there are ongoing studies carefully designed to express these genes in other living systems. For instance 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 biocatalyst would thus open up numerous profitable paths in the area of biotechnology. Thus, in this study by exploiting genomic advances, we are able to quantify potential molecular indices that may aid environmental risk assessment with the development of assay technologies delivering quantitative and specific analyses that allow differentiation between very similar gene products.

SUMMARY

Chapter 5 SUMMARY

Metatranscriptomics is a modern “omic” approach developed to elucidate the response of microbial communities in adverse environmental conditions by studying gene expression profiling of functional and taxonomic marker genes through community mRNA (Handelsman *et al.* 2007). Recent researches in the field of microbial ecology advocates presence of numerous varieties of novel micro organisms in the environment in comparison to the few lab cultivated microbes (Ward *et al.* 1990; Giovannoni *et al.* 1990). Though traditional culturing methods and other PCR based techniques such as genetic finger printing and metagenomics were exploited in past studies to find novel genes and species in various studies, however, none of these studies absorbed eukaryotic microflora especially the uncultivable one and largely preferred the unveiling of the archea and bacteria only (Daniel, 2005). Thus, to explore the true functional diversity of an environment, we explored advanced omic approach i.e. metatranscriptomics which unlike other approaches targeted mainly eukaryotic microbial communities which had been unnoticed and unexplored from a long time in metal polluted soil environment.

In the present study, functional approach of metatranscriptomics was explored to appreciate the true functional diversity of the polluted soil site from agro-forestry land in Pierrelaye (PL) (49°1′45”N, 2°10′32″E) in northwest of Paris, France. The experimental strategy adapted to synthesize cDNA library was different from the traditional methods as it involved total RNA as template material and synthesized cDNA by template switching mechanism of MMLV RT using 3’-end CDS adaptor and 5’-end plugOligo-3M adaptors. This technology circumvents the problems related to low amount or half life of mRNA, presence of introns and dilution of eukaryotic DNA due to large amount of prokaryotic DNA. Next significant strategy used in this work was size fractionation of cDNAs performed before amplification. Size fractionation of cDNA resulted into three different sized cDNA fractions by using bi-dimensional gel electrophoresis thus bypass the size based biases generally observed during cloning and screening of varied size cDNAs from a consortium of short and long cDNAs. Full length cDNAs synthesized from reverse transcription are further amplified by PCR and cloned downstream of PGK1 promoter region in a pFL61 shuttle vector. PGK promoter helped in transcription of the cDNA inserts and their successful expression in the form of proteins inside the bacterial as well as eukaryotic host and also circumvents the problems related to presence of introns and dilution of eukaryotic DNA, generally observed in most of the metagenome studies. Three different sized eukaryotic cDNA libraries *i.e.*, Lib A (<0.5 kb), Lib B (0.5-1.0 kb), Lib C (>4.0 kb) were prepared from extracted soil RNA.

A good metatranscriptomic cDNA library constructed from highly purified mRNA encompasses more than 10^6 independent inserts (Bailly *et al.* 2007). Thus, each of the libraries was screened for the cDNAs providing tolerance to cadmium (Cd) toxicity through a heterologous complementation assay using Cd sensitive baker's yeast mutant *ycf1*$^{\Delta}$. Screening of libraries for Cd tolerance using uracil auxotrophic yeast mutant *ycf1*$^{\Delta}$ is a simple and rapid process of filtering out the huge number of metal sensitive cDNA inserts. The level of Cd tolerance among these clones varied from higher to medium to lower in secondary screening i.e. drop assay. In the present study, a total of 5x 10^6 clones were screened from Lib C, 1,11,949 clones were screened from Lib B and 1,74,020 clones were screened from Cd tolerance out of which 112 clones were sequenced after passing primary (yeast transformation), secondary (Drop test), tertiary (5-FOA test) and quaternary (colony PCR) mode of screening. Further, BLASTX analysis of these retrieved sequences cDNA inserts provided us four full length ORF carrying cDNAs which were further characterized for providing tolerance phenotype to the metal sensitive yeast mutants used in this study.

The first cDNA with full length ORF reported in this work belongs to the Lib B in which all the cDNAs range between 0.1 - 0.5 Kb. Characterization of 147 amino acid long polypeptide encoding cDNA PLBe1 marked presence of N-terminal 76 amino acid ubiquitin domain followed by a 66 amino acid long extension of AN1 type Zn finger located at C-terminal. Presence of diglycine residues at the N-terminal of ubiquitin domain and lysine residues at 48 and 63 positions in deduced polypeptide UFP signified the proteolytic processing of ubiquitin moiety and hence its active role in destruction of aberrant proteins accumulated inside the cell due to metal stress, *via* ubiquitin proteosomal pathway. Phylogeny of cDNA PLBe1 showed strong connection with the affiliates of phylum Amoebozoa which supported the fact that this gene might have its origin from the kingdom Protista. Induction of cDNA PLBe1 on exposure to toxic concentrations of Cd also supported the role of cDNA in combating Cd toxicity. Furthermore, cDNA PLBe1 efficiently supported the growth of all the yeast sensitive mutants on different metal amended media while performing heterologous complementation assays. Role of cDNA PLBe1 in favoring the accumulation of metals inside the yeast cells was highlighted by performing ICP-MS for respective yeast transformants in the varied metal concentrations with maximum accumulation observed at 40 µM of $CdSO_4$, 300 µM of $CuSO_4$, 10 mM of $ZnSO_4$ and 2 mM of $CoCl_2$. Tolerance profile and the level of accumulation inside yeast transformants accommodating cDNA PLBe1 highlighted the active role of cDNA PLBe1 in combating metal toxicity in metal polluted soil.

Another cDNA screened for Cd tolerance belong to the largest library *i.e.*, Lib C which contained cDNA fractions ranged between 1-4 Kb size. The cDNA PLCe10 is one of the four cDNAs, encoding a 338 amino acid long polypeptide showing homology with von Willebrand factor type D (VWD) of Vg protein. Presence of high degree of conservation in deduced polypeptide such as two C-terminal RXXR consensus cleavage sites and N-terminal GL/ICG motif determined through various bioinformatic tools also advocated that the deduced polypeptide showed homology with VWD domain of Vg protein. Phylogenetic analysis of cDNA PLCe10 clearly clustered cDNA PLCe10 with the members of phylum Nematoda. The cDNA PLCe10 further confirmed its role in attributing metal tolerance to the mutant yeast strains (*ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$) used in the study by showing significant growth on metal amended media in both the heterologous complementation assays *i.e.*, drop assay as well as growth assay. It has been observed that cDNA PLCe10 is capable of accumulating a higher range of metals (Cd, Cu, Zn and Co) in comparison to wild strain BY4741 transformed with empty vector pFL61. Relative expression of cDNA PLCe10 also reported an increase in Cd sensitive yeast mutant *ycf1*$^{\Delta}$ on Cd exposure with maximum level of gene expression at 60 µM of $CdSO_4$ thus supported the fact that metal toxicity induced the expression of cDNA PLCe10 which may be actively involved in the some known or unknown metal toxicity combating mechanism.

cDNA PLCc43 from Lib C was third cDNA screened for Cd tolerance, sharing sequence homology with Hsp40 Type I proteins of *Acanthameoba castellanii* and the putative polypeptide displayed the fundamental structural features of Hsp40 type I proteins *i.e.*, a partial J domain with the conserved tri-peptide HPD motif, a stretch of G/F rich region and four zinc finger domains. cDNA PLCc43 had huddled with the member of kingdom Animalia which indicated the origin of this gene from higher animals. Increase in the relative expression of Cd tolerant cDNA PLCc43 on Cd exposure highlighted the ability of cDNA to get induced when metal is present above the threshold level. Functional complementation of cDNA PLCc43 in various yeast mutants developed the capability of these mutants to grow and accumulate a good amount of respective metal on exposure to toxic range of different metal concentrations.

In a similar way, fourth cDNA PLCd43 screened for Cd tolerance at 40 µM of $CdSO_4$ belongs to largest Lib C and shown sequence similarity with γ-Glutamylcysteinyl transferase or phytochelatin synthase (PCS) which catalyses the synthesis of phytochelatins from glutathione molecules. cDNA PLCd43 marked presence of highly conserved N- terminal domain with presence of C^{120}, H^{237} and D^{256} residues essential for catalysis and five highly conserved cysteine residues at position C^{120}, C^{154}, C^{155}, C^{173} and C^{188} which are known to play important role in recognition of metal ions in all the eukaryotes. Screened cDNA PLCd43 shared its phylogeny with affiliates of family *Brassicaceae* of kingdom Plantae. Relative expression of mRNA transcripts coded by cDNA PLCd43 with a maximum level of expression at 80 µM of $CdSO_4$, confirmed the stimulation of cDNA PLCd43 in presence of lethal concentrations of Cd. Transformation of various metal sensitive yeast mutants with cDNA PLCd43 has shown tremendous growth at lethal concentration of Cd, Cu, Zn and Co and also observed to accumulate a good amount of metal inside respective yeast cell.

All these characteristic features of putative polypeptides, their metal tolerance and accumulation ability characterized these full length ORF carrying metatranscripts as potential affiliates of an environment that may play some significant role in conferring tolerance towards metal toxicity. This study provided a highly efficient experimental strategy to isolate and characterize various eukaryotic genes whose over-expression restored tolerance in hypersensitive yeast mutants towards all the four potentially toxic metals thus strongly supported probability of existence of various novel and unknown mechanisms and biomolecules. Hence, function based approach of metatranscriptomics has been found suitable to characterize novel genes belonging to unknown and uncultivable members of eukaryotic community playing vital role in adaptation of biodiversity of an ecosystem against metal toxicity.

Functional diversity of the soil samples was determined by targeting the hypervariable V4 region of 18S rRNA/rDNA using pair end chemistry on Illumina MiSeq platform (2x250 bp). This high throughput sequencing platform is more advanced platform of NGS involving pair end sequence chemistry with read capacity of 2x250 bp thus capable of reading a long sequence and provide more consolidated data unlike previous NGS techniques. V4 region specific amplicons were sequenced using Illumina Miseq platform, were analyzed through Mothur pipeline and using SILVA reference database. Taxonomic classification of the sequences through Naive Bayesian classification revealed that majority of the eukaryotic diversity in metal polluted soils was dominated by metazoan and fungi followed by other eukaryotic phylum such as Cercozoa, Ciliphora and Conosa. Hence, analysis of V4 specific amplicons revealed that sequences retrieved from total RNA pool of the soil were majorly of eukaryotic origin and advocated the this strategy to study the functional eukaryotic diversity flourishing in various environments. Efficiency of both the approaches used in the study *i.e.,* functional as well as sequence based approaches supported these experimental strategies and protocols as highly efficient to mine numerous novel and known genes from various environmental gene pools.

Recent years in the field of microbial ecology have evolved as a revolutionary era as it involved a profound shift from traditional phylogenetic analyses of single cultivable microbes to the taxonomic shifts in different communities by exploiting the complex experimental designs (Gevers *et al.* 2012). Most of the earlier work was focused on prokaryotes and sequence based approach of metatranscriptomics while this is the first reported experimental analysis of an environmental sample where both function based as well as sequence based approach of metatranscriptomics has been used simultaneously. In past studies, most of the metal tolerant genes were related only to individual organisms, while in present study these metal tolerant metatranscripts screened from metatranscriptomic cDNA library from metal polluted soil environment is novel finding of this work. As future perspective, we may consider these cDNAs or encoded polypeptides as molecular probes in gene related processes and to assess their involvement in the overall capacity of organisms to survive in potentially toxic metal enriched environment. Furthermore, these genes and their encoded peptides could be exploited to develop as biomarkers in order to determine potentially toxic metals present in the environment principally in the soil and the levels of tolerance or resistance among the species of concern for the bioremediation of polluted lands.

REFERENCES

REFERENCES

Abdel Sabour, M. F., Ismail, F. S. & Abou Naga, H. (1995). Environmental impact of Cairo sewage effluent in El-Gabel El Asfer farm. Eqypt. *J. Soil Sci.* 35: 225-237.

Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., & White, O. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* 8, p.e1002358.

Achiba, W. B., Lakhdar, A., Gabteni, N., Du Laing, G., Verloo, M., Boeckx, P., Van Cleemput, O., Jedidi, N. & Gallali, T. (2010). Accumulation and fractionation of trace metals in a Tunisian calcareous soil amended with farmyard manure and municipal solid waste compost. *J. Hazard. Mater*. 176, 99-108.

Adamis, P. D., Gomes, D. S., Pinto, M. L. C., Panek, A. D. & Eleutherio, E. C. (2004). The role of glutathione transferases in cadmium stress. *Toxicol. Lett.* 154, 81-88.

Adler, G., Konrad, Z., Zamir, L., Mishra, A. K., Raveh, D., & Bar-Zvi, D. (2017). The Arabidopsis paralogs, PUB46 and PUB48, encoding U-box E3 ubiquitin ligases, are essential for plant response to drought stress. *BMC Plant Biol. 17*, 8.

Adriano, D. C. (1986). Trace Elements in the Terrestrial Environment, Springer-Vedag, New York.

Adriano, D. C. (2003). *Trace Elements in Terrestrial Environments: Biogeochemistry, Bioavailability and Risks of Metals*, Springer, New York, NY, U.S.A.

Agency for Toxic Substances and Disease Registry (ATSDR). Toxicological Profile for Copper. (2002). Atlanta, GA: Centers for Disease Control.

Agger, J. W., Isaksen, T., Várnai, A., Vidal-Melgosa, S., Willats, W. G., Ludwig, R., Horn, S.J., Eijsink, V.G. & Westereng, B. (2014). Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6287-6292.

Aislabie, J., & Deslippe, J. R. (2013). Soil microbes and their contribution to soil services. *In* Dymond JR ed. Ecosystem services in New Zealand– conditions and trends. Manaaki Whenua Press, Lincoln, New Zealand.

Algreen, M., Rein, A., Legind, C. N., Amundsen, C. E., Karlson, U. G., & Trapp, S. (2012). Test of tree core sampling for screening of toxic elements in soils from a Norwegian site. *I. J. Phytoremediat.* 14, 305-319.

Allison, S. D. & Martiny, J. B. (2008). Resistance, resilience and redundancy in microbial communities. *Proc. Natl. Acad. Sci. USA.* 105: 11512–11519.

Alloway, B. J. (2008). Zinc in soils and crop nutrition. Brussels/Paris: International Zinc Association/International Fertilizer Industry Association.

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143-169.

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., Huse, S. M. (2009). A method for studying rotistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS One 4:e6372.

Amdam, G. V., Norberg, K., Fondrk, M. K., Page, R. E. (2004). Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11350-11355.

Anderson, R. C. (2006). Evolution and origin of the Central Grassland of North America: climate, fire, and mammalian grazers1. *J. Torrey Bot. Soc.* 133, 626-647.

Ando, S., & Yanagida, K. (1999). Susceptibility to oxidation of copper-induced plasma lipoproteins from Japanese eel: protective effect of vitellogenin on the oxidation of very low density lipoprotein. *Comp. Biochem. Phys. C. 123*, 1-7.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Anirudhan, T.S. & Sreekumari, S.S., 2011. Adsorptive removal of heavy metal ions from industrial effluents using activated carbon derived from waste coconut buttons. *J. Environ. Sci.* 23, 1989-1998.

Anjum, N. A., Hasanuzzaman, M., Hossein, M. A., Thangavel, P., Roychoudhury, A., & Gill, S. S. (2015). Jacks of metal/metalloid chelation trade in plants-an overview. *Fr ont. Plant. Sci* .6,192.

Annabi, A., Kessabi, K., Kerkeni, A., Said, K., & Messaoudi, I. (2012). Influence of cadmium exposure on growth and fecundity of freshwater mosquitofish *Gambusia affinis*: *in situ* and *in vivo* studies. *Biol. Trace. Elem. Res.* 148, 345-355.

Appenroth, K. J. (2010). Definition of "heavy metals" and their role in biological systems. In *Soil heavy metals*. Springer, Berlin, Heidelberg. 19-29.

Arbeli, Z., & Fuentes, C. L. (2007). Improved purification and PCR amplification of DNA from environmental samples. *FEMS Microbiol. Lett. 272*, 269-275.

Armon, R. H., & Starosvetsky, J. (2015). Point source pollution indicators. In Environmental Indicators Springer, Dordrecht. 495-499.

Arruti A, Fernandez-Olmo I & Irabien A. (2010). Evaluation of the contribution of local sources to trace metals levels in urban PM2.5 and PM10 in the Cantabria region (Northern Spain). *J. Environ. Monit.* 12,1451–1458.

Artz, R., Anastasiou, D., Arrouays, D., Bastos, A. C., Bendetti, A., Bispo, A., Brandmayr, P., Broll, G., Bunning, S., Castracani, C. & Campbell, C. (2010). *European atlas of soil biodiversity*. Office des Publications Officielles des Communautés Européennes. 128-p

Auffret, M. D., Kharu, K., Kachane, A., Dungait, J. A. J., Fraser, F., Hopkins, D. W., Wookey, P. A., Singh, B. K., Freitag, T. E., Hartley, I. P. and Prosser, J. I. (2016). The role of microbial community composition in controlling soil respiration responses to temperature. *PLoS ONE*, 11: e0165448.

Bailly, J., Fraissinet-Tachet, L., Verner, M. C., Debaud, J. C., Lemaire, M., Wésolowski-Louvel, M., &Marmeisse, R. (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISME J.* 1, 632-642.

Banecki. B., Liberek, K., Wall, D., Wawrzynow, A., Georgopoulos, C., Bertoli, E., Tanfani, F., &Zylicz, M. (1996). Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone. *J. Biol. Chem.* 271, 14840-14848.

Barr, D. P., & Aust, S. D. (1994). Mechanisms white rot fungi use to degrade pollutants. *Environ. Sci. Technol.* 28, 78A-87A.

Basta, N. T., Ryan, J. A., & Chaney, R. L. (2005). Trace element chemistry in residual-treated soil: Key concepts and metal bioavailability. J. Environ. Qual. 34, 49-63.

Bates, S. T., Clemente, J. C., Flores, G. E., Walters, W. A., Parfrey, L. W., Knight, R., &Fierer, N. (2013). Global biogeography of highly diverse protistan communities in soil. *ISME J*. 7, 652-659.

Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J.M., Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res*. 10, 1001-1010.

Belasco, J. G. (2010). All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat. Rev. Mol. Cell. Biol.* 11, 467-478.

Bellion, M., Courbot, M., Jacob, C., Blaudez, D., & Chalot, M. (2006). Extracellular and cellular mechanisms sustaining metal tolerance in ectomycorrhizal fungi. *FEMS Microbiol. Lett.* 254, 173-181.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R. & Boutell, J. M. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 456, 53-59.

Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S., & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9697-9702.

Berryman, C. (1968). Residual values of slurries. In *Paper presented at the NAAS Soil Scientists Open Conference*.

Bertin, G., & Averbeck, D. (2006). Cadmium: cellular effects, modifications of biomolecules, modulation of DNA repair and genotoxic consequences (a review). *Biochimie*. 88, 1549-1559.

Beyersmann, D. & Hartwig, A. (2008). Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. *Arch. Toxicol.* 82, 493.

Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E., & 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.

Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberon, 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.

Blanco, L., Bernad, A., Lázaro, J. M., Martín, G., Garmendia, C., & Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.*, 264, 8935-8940.

Boeke, J. D., Lacroute, F., & Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* 197, 345–346

Bohlen, P. J., Scheu, S., Hale, C. M., McLean, M. A., Migge, S., Groffman, P. M., & Parkinson, D. (2004). Non-native invasive earthworms as agents of change in northern temperate forests. *Front. Ecol Environ. 2*, 427-435.

Bolan, N., Adriano, D., & Mahimairaja, S. (2004). Distribution and bioavailability of trace elements in livestock and poultry manure by-products. *Crit. Rev. Environ. Sci. Technol.* 34, 291-338.

Bolan, N. S., Adriano, D. C., Duraisamy, P., Mani, A., & Arulmozhiselvan, K. (2003). Immobilization and phytoavailability of cadmium in variable charge soils: I. Effect of phosphate addition. *Plant Soil.* 250, 83-94.

Bomar, L., Maltz, M., Colston, S., & Graf, J. (2011). Directed culturing of microorganisms using metatranscriptomics. *MBio*, 2, 2.

Bradley, I. M., Pinto, A. J., & Guest, J. S. (2016). Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. *Appl. Environ. Microbiol.* 82, 5878-58

Brady, N. C., & Weil, R. R. (2002). *The nature and properties of soils*. Upper Saddle River, NJ: Prentice Hall. 13, 662-710.

Bringezu, K., Lichtenberger, O., Leopold, I., & Neumann, D. (1999). Heavy metal tolerance of Silene vulgaris. *J. Plant Physiol*. 154, 536-546.

Brodsky, J. L., & Schekman, R. (1993). A Sec63p-BiP complex from yeast is required for protein translocation in a reconstituted proteoliposome. *J. Cell Biol.* 123, 1355-1363.

Bryant, S., & Manning, D. L. (1998). Formaldehyde gel electrophoresis of total RNA. In *RNA Isolation and Characterization Protocols,* Humana Press. 69-72.

Bundt, M., Widmer, F., Pesaro, M., Zeyer, J., & Blaser, P. (2001). Preferential flow paths:biological 'hot spots' in soils. *Soil Biol. Biochem*. 33, 729-738

Burges, A., Epelde, L., &Garbisu, C. (2015). Impact of repeated single-metal and multi-metal pollution events on soil quality. *Chemosphere*. 120, 8-15.

Callis, J., Raasch, J. A., & Vierstra, R. D. (1990). Ubiquitin extension proteins of Arabidopsis thaliana. Structure, localization, and expression of their promoters in transgenic tobacco. *J. Biol. Chem.* 265, 12486-12493.

Cao, M., Wei, C., Zhao, L., Wang, J., Jia, Q., Wang, X., Jin, Q. & Deng, T. (2014) DnaJA1/Hsp40 is co-opted by influenza A virus to enhance its viral RNA polymerase activity. *J. Virol.* 88:14078-14089.

Caplan, A. J. & Douglas, M. G. (1991). Characterization of YDJ1: a yeast homologue of the bacterial dnaJ protein. *J. Cell. Biol.* 114: 609-621.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N. Owens, S.M., Betley, J., Fraser, L., Bauer, M., & Gormley, N., (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621-1624.

Carninci, P. & Hayashizaki, Y. (1999) High-efficiency full-length cDNA cloning. *Meth. Enzymol.* Academic Press. *303*, 19-44.

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., & Muramatsu, M. (1996). High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*. 37, 327-336.

Carvalhais, L. C., Dennis, P. G., Tyson, G. W., & Schenk, P. M. (2013). Rhizosphere metatranscriptomics: challenges and opportunities. *Molecular microbial ecology of the rhizosphere*. 2, 1137-1144.

Carvalhais, L. C., Dennis, P. G., Tyson, G. W., & Schenk, P. M. (2012). Application of metatranscriptomics to soil environments. *J. microbial. methods*, 91, 246-251.

Cervera, A., Maymo, A. C., Martinez-Pardo, R. & Garcera, M. D. (2005). Vitellogenesis inhibition in *Oncopeltusfasciatus* females (Heteroptera: Lygaeidae) exposed to cadmium. *J. Insect Physiol.* 51, 895-911.

Cervera, A., Maymo, A. C., Martınez-Pardo, R., & Garcera, M. D. (2006). Vitellogenin polypeptide levels in one susceptible and one cadmium-resistant strain of *Oncopeltusfasciatus* (Heteroptera: Lygaeidae) and its role in cadmium resistance. *J. Insect Physiol.* 52, 158-168.

Chang, L. W., Magos, L. & Suzuki, T. (1996). Toxicology of Metals. Boca Raton. FL, USA: CRC, Press.

Chaudhary, K., Agarwal, S. & Khan, S. (2018). Role of phytochelatins (PCs), metallothioneins (MTs) and heavy metal ATPase (HMA) genes in heavy metal tolerance. In *Mycoremediation and Environmental Sustainability*. Springer, Cham. 39-60.

Cheetham, M. E., & Caplan, A. J. (1998). Structure, function and evolution of DnaJ: Conservation and adaptation of chaperone function. *Cell Stress Chaperone.* 3, 28-36.

Chen, J., Zhou, J., & Goldsbrough, P. B. (1997). Characterization of phytochelatin synthase from tomato. *Physiol. Plant*. 101, 165-172.

Chen, W., Chang, A. C., & Wu, L. (2007). Assessing long-term environmental risks of trace elements in phosphate fertilizers. *Ecotoxicol. Environ. Safe.* 67, 48-58.

Chistoserdova, L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol. Lett.*32, 1351-1359.

Chomczynski, P. & Sacchi, N. (1987). Single step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. *Anal. Biochem.* 162, 156-159.

Chudaeva, V. A., Shesterkin, V. P. & Chudaev, O. V. (2011). Trace elements in surface water in Amur River basin. *Water Res.* 38(5), 650.

Clemens, S. (2001). Molecular mechanism of plant metal tolerance and homeostasis. Planta. 212, 475-486.

Clemens, S., Kim, E. J., Neumann, D. & Schroeder, J. I. (1999). Tolerance to toxic metals by a gene family of phytochelatin synthases from plants and yeast. *EMBO J.* 18, 3325-3333.

Cobbett, C. S. (2000). Phytochelatin biosynthesis and function in heavy-metal detoxification. *Curr. Opin. Plant Biol.* 3, 211-216.

Cocquet, J., Chong, A., Zhang, G.L. & Veitia, R. A. (2006). Reverse transcriptase template switching and false alternative transcripts. *Genomics.* 88, 127-131.

Codling, E. E., Chaney, R. L., & Green, C. E. (2015). Accumulation of lead and arsenic by carrots grown on lead-arsenate contaminated orchard soils. *J. Plant Nutr*. 38, 509-525.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B. Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M. &Tiedje, J. M., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37, D141–D145.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R. &Tiedje, J.M. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 42, 633-642.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast 2 GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 21, 3674-3676.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X. & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13.

Conklin, D. S., McMaster, J. A., Culbertson, M. R. & Kung, C. (1992). COT1, a gene involved in cobalt accumulation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 3678-3688.

Corona, M., Velarde, R. A., Remolina, S., Moran-Lauter, A., Wang, Y., Hughes, K. A. & Robinson, G. E. (2007). Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Pro. Natl. Acad. Sci. U.S.A.* 104, 7128-7133.

Costa, R., Gotz, M., Mrotzek, N., Lottmann, J., Berg, G. & Smalla, K. (2006). Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. *Fems Microbiol. Ecol.* 56, 236-249.

Courtois, S., Cappellano, C. M., Ball, M., Francou, F. X., Normand, P. & Helynck, G. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* 69, 49-55.

Courty, P. E., Franc, A., Pierrat, J. C., & Garbaye, J. (2008). Temporal changes in the ectomycorrhizal community in two soil horizons of a temperate oak forest. *Appl. Environ. Microbiol.* 74, 5792-5801.

Curtis, T. P., Sloan, W. T. & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. Proc. *Nat. Acad. Sci. U.S.A*, 99, 10494-10499.

Cyr, D. M. & Neupert, W. (1996). Roles for hsp70 in protein translocation across membranes of organelles. *Exs.*77, 25-40

Cyr, D. M., Lu, X., & Douglas, M. G. (1992). Regulation of Hsp70 function by a eukaryotic DnaJ homolog. *J. Biol. Chem.* 267, 20927-20931.

Cyr, D. M., Langer, T., & Douglas, M. G. (1994). DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem. Sci.* 19, 176-181.

D'Amore, J. J., Al-Abed, S. R., Scheckel, K. G., & Ryan, J. A. (2005). Methods for speciation of metals in soils: a review. *J. Environ. Qual.* 34, 1707-1745.

Da Rocha, U. N., Andreote, F. D., de Azevedo, J. L., van Elsas, J. D. & van Overbeek, L. S. (2010). Cultivation of hitherto-uncultured bacteria belonging to the Verrucomicrobia subdivision 1 from the potato (*Solanum tuberosum* L.) rhizosphere. *J. Soil Sediment.* 10, 326-339.

da Silva Oliveira, A., Bocio, A., Trevilato, T. M. B., Takayanagui, A. M. M., Domingo, J. L., & Segura-Muñoz, S. I. (2007). Heavy metals in untreated/treated urban effluent and sludge from a biological wastewater treatment plant. *Environ. Sci. Poll. Res. Int.* 14, 483.

Da Silva, N. A., & Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. 12, 197-214.

Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L. & Marmeisse, R. (2012). Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. *PloS one*, 7, pe2896.

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.

Daniel, R. (2005). The metagenomics of soil. *Nat. Rev. Microbiol.*3, 470-478.

Davinic, M., Fultz, L. M., Acosta-Martinez, V., Calderón, F. J., Cox, S. B., Dowd, S. E., Allen, V. G., Zak, J. C., & Moore-Kucera, J. (2012). Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. *Soil Biol. Biochem.* 46, 63-72.

Davis, K. E., Sangwan, P. & Janssen, P. H. (2011). *Acidobacteria*, *Rubrobacteridae* and *Chloroflexi*are abundant among very slow-growing and mini-colony-forming soil bacteria. *Environ. Microbiol.* 13, 798-805.

Degens, B. P., Schipper, L. A., Sparling, G. P. & 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.

De Bruyn, L. & Conacher, A. J. (1990). The role of termites and ants in soil modification-a review. *Soil Res.* 28, 55-93.

de Menezes, A., Clipson, N., & Doyle, E. (2012). Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. *Environ Microbiol.* 14, 2577-2588.

De Miranda, J. R., Thomas, M. A., Thurman, D. A., & Tomsett, A. B. (1989). Metalliothionein genes from the flowering plant *Mimulusguttatus*. *FEBS Lett.* 260, 277-280.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I. & Carmichael, M. (2015). Eukaryotic plankton diversity in the sunlight ocean. *Science.* 348, 1261605.

De Vos, C. H. R., Schat, H., De Waal, M. A. M., Vooijs, R., & Ernst, W. H. O. (1991). Increased resistance to copper-induced damage of the root cell plasmalemma in copper tolerant *Silene cucubalus*. *Physiol*. *Plant*, 82, 523-528.

Del Campo, J., Sieracki, M. E., Molestina, R., Keeling, P., Massana, R., Ruiz-Trillo, I. (2014). The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* 29, 252-259.

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., Martinez, A., Sullivan, M. B., Edwards, R., Brito, B. R. & Chisholm, S. W. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, 311, 496-503.

Denslow, N. D., Chow, M. C., Kroll, K. J. & Green, L. (1999). Vitellogenin as a biomarker of exposure for estrogen or estrogen mimics. *Ecotoxicol.* 8, 385-398.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol. 72*(7), 5069-5072.

Deslippe, J. R., Hartmann, M., Mohn, B. W. & Simard, S. W. (2012). Long term warming alters arctic soil microbial communities. *FEMS Microbiol. Ecol.* 81, 594-607.

Deutscher, M.P. (2006). Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res.* 34, 659-666.

Devêvre, O., Garbaye, J., & Botton, B. (1996) Release of complexing organic acids by rhizosphere fungi as a factor in Norway spruce yellowing in acidic soils. *Mycol. Res.* 100, 1367-1374.

Diaconu, M., Pavel, L. V., Hlihor, R. M., Rosca, M., Fertu, D. I., Lenz, M., Corvini, P.X. & Gavrilescu, M. (2020). Characterization of heavy metal toxicity in some plants and microorganisms- A preliminary approach for environmental bioremediation. *N. Biotechnol*, 56, 130-139.

Díaz-Cruz, M. S., García-Galán, M. J., Guerra, P., Jelic, A., Postigo, C., Eljarrat, E., Farré, M., De Alda, M. L., Petrovic, M. & Barceló, D. (2009). Analysis of selected emerging contaminants in sewage sludge. *Trends Anal. Chem.* 28, 1263-1275.

Dietz, K. J., Baier, M., & Kramer, U. (1999). Free radicals and reactive oxygen species as mediators of heavy metal toxicity in plants. In *Heavy metal stress in plants.* Springer, Berlin, Heidelberg, 73-97.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., Furlan, M., Desnues, C., Haynes, M., Li, L. & McDaniel, L. (2008). Functional metagenomic profiling of nine biomes. *Nature.* 452, 629-32.

Donato, J. J., Moe, L. A., Converse, B. J., Smart, K. D., Berklein, F. C. & McManus, P. S. (2010). Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. *Appl. Environ. Microbiol.*76, 4396-4401.

Dong, X., Kleiner, M., Sharp, C. E., Thorson, E., Li, C. & Liu, D. (2017). Fast and simple analysis of MiSeq amplicon sequencing data with MetaAmp. *Front. Microbiol*. 8, 1461.

D'Onofrio, A., Crawford, J. M., Stewart, E. J., Witt, K., Gavrish, E. & Epstein, S. (2010). Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* 17, 254-264.

Dreyfus, M. & Regnier, P. (2002). The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell.* 111, 611-613.

Duan, C. J., Xian, L., Zhao, G. C., Feng, Y., Pang, H., Bai, X. L., Tang, J. L., Ma, Q.S. and Feng, J. X. (2009). Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* 107, 245-256.

Dudal, R. (2005). The sixth factor of soil formation. *Eurasian Soil Sci. C/C of Pochvovedenie*, 38, 60.

Duffus, J. H. (2002). "Heavy metals" a meaningless term. (IUPAC Technical Report). *Pure. Appl. Chem.* 74, 793-807.

Dumbrell, A. J., Ashton, P. D., Aziz, N., Feng, G., Nelson, M., Dytham, C., Fitter, A. H. & Helgason, T. (2011). Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. *New Phytol.*190, 794-804.

Edery, I., Chu, L. L., Sonenberg, N., & 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.

Eilers, K. G, Lauber, C. L., Knight, R. & Fierer, N. (2010). Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. *Soil Biol. Biochem.* 42, 896-903.

Falk, N., Reid, T., Skoyles, A., Grgicak-Mannion, A., Drouillard, K. & Weisener, C. G. (2019). Microbial metatranscriptomic investigations across contaminant gradients of the Detroit River. *Sci.Total Environ.* 690, 121-131.

Fan, W., Guo, Q., Liu, C., Liu, X., Zhang, M., Long, D., Xiang, Z. & Zhao, A. (2018). Two mulberry phytochelatin synthase genes confer zinc/cadmium tolerance and accumulation in transgenic *Arabidopsis* and tobacco. *Gene.* 645, 95-104.

Fantke, P., Friedrich, R. & Jolliet, O. (2012). Health impact and damage cost assessment of pesticides in Europe. *Environ. Int.* 49, 9-17.

Fenik, J., Tankiewicz, M., & Biziuk, M. (2011). Properties and determination of pesticides in fruits and vegetables. *Trend Anal. Chem.* 30, 814-826.

Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J. & Ordóñez, C. (2018). Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: application to rough mountain areas in Northwest Spain. *Geoderma.* 311, 120-129.

Fernández-Luqueño, F., Valenzuela-Encinas, C., Marsch, R., Martínez-Suárez, C., Vázquez-Núñez, E., & Dendooven, L. (2011). Microbial communities to mitigate contamination of PAHs in soil-possibilities and challenges: a review. *Environ. Sci. Poll. Res.* 18, 12-30.

Ferrer, M., Beloqui, A., Timmis, K. N., & Golyshin, P. N. (2009). Metagenomics for mining new genetic resources of microbial communities. *J. Mol. Microbiol. Biotechnol.* 16, 109-123.

Festa, R. A., & Thiele, D. J. (2011). Copper: an essential metal in biology. *Curr. Biol.* 21: R877-R883.

Fichot, E. B., & Norman, R. S. (2013). Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome.* 1, 10.

Fierer, N., Bradford, M. A. & Jackson, R. B. (2007). Towards an ecological classification of soil bacteria. *Ecology.* 88, 1354-1364.

Fierer, N., Bradford, M. A., & Jackson, R. B. (2007). Towards an ecological classification of soil bacteria. *Ecol*. 88, 1354-1364.

Finley, D., Ozkaynak, E. & Varshavsky, A. (1987). The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. *Cell.* 48, 1035-1046.

Finn, R. N. & Fyhn, H. J. (2010). Requirement for amino acids in ontogeny of fish. *Aquacult. Res*. 41, 684-716.

Fomina, M. A., Alexander, I. J., Colpaert, J. V. & Gadd, G. M. (2005). Solubilization of toxic metal minerals and metal tolerance of mycorrhizal fungi. *Soil Biol. Biochem.* 37, 851-866.

Forzani, C., Lobreaux, S., Mari, S., Briat, J.F. & Lebrun, M. (2002). Metal resistance in yeast mediated by the expression of a maize 20S proteasome α subunit. *Gene.* 293, 199-204.

Freemont, P. S. (2000). Ubiquitination: RING for destruction. *Curr. Biol.* 10, 84-87.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W. & DeLong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3805-3810.

Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.* 70, 603-647.

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature.*459, 193-199.

Gabe, U. & Rodella, A. A. (1999). Trace elements in Brazilian agricultural limestones and mineral fertilizers. *Commun. Soil Sci. Plant Anal.* 30, 605-620.

Galal, T. M. & Shehata, H. S. (2015). Impact of nutrients and heavy metals capture by weeds on the growth and production of rice (*Oryza sativa* L.) irrigated with different water sources. *Ecol. Indic.* 54, 108-115.

Garcia-Martinez, J., Acinas, S. G., Anton, A. I. & Rodriguez-Valera, F. (1999). Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J. Microbiol. Meth.* 36, 55-64.

Gasic, K. & Korban, S. S. (2007). Expression of Arabidopsis phytochelatin synthase in Indian mustard (*Brassica juncea*) plants enhances tolerance for Cd and Zn. *Planta*, 225, 1277-1285.

Gekeler,W., Grill, E., Winnacker, E. L. & Zenk, M. H. (1988). Algae sequester heavy metals via synthesis of phytochelatin complexes. *Arch. Microbiol.* 150, 197-202.

George, I. F., Hartmann, M., Liles, M. R. & Agathos, S. N. (2011). Recovery of as-yet-uncultured soil *Acidobacteria* on dilute solid media. *Appl. Environ. Microbiol.*, 77, 8184-8188.

Giannoukos, G., Ciulla, D.M., Huang, K., Haas, B.J., Izard, J., Levin, J.Z., Livny, J., Earl, A.M., Gevers, D., Ward, D.V. & Nusbaum, C. (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* 13, 23.

Gietz, R. D. & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/ss carrier DNA/PEG method. *Nat. Protoc.*2:31-34

Gilbert, J. A., Field, D., Huang, Y., Edwards, R., Li, W. Z., Gilna, P. & Joint, I. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One 3, 8.

Gilluly, J., Waters, A. C. & Woodford, A. O. (1975). Principles of geology (4th Ed.). San Francisco, California: W.H. Freeman. ISBN: 978-0716702696.

Giovannoni, S. J., Britschgi, T. B., Meyer, C. L. & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature.* 345, 60-63.

Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D. & Meyer, F. (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc*. 1, 5368.

Gloor, G. B., Hummelen, R., Macklaim, J. M., Dickson, R. J., Fernandes, A. D., MacPhee, R. & Reid, G., (2010). Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One* 5:e15406.

Goldfarb, K. C., Karaoz, U., Hanson, C. A., Santee, C. A., Bradford, M. A., Treseder, K. K., Wallenstein M. D. & Brodie, E. L. (2011). Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Front. Microbiol.* 2, 94.

Goldsbrough, P. (2000). Metal tolerance in plants: the role of phytochelatins and metallothioneins. *Norman Terry*.

Gonzalez, E., Pitre, F. E., Pagé, A. P., Marleau, J., Nissim, W. G., St-Arnaud, M., Labrecque, M., Joly, S., Yergeau, E. & Brereton, N. J. (2018). Trees, fungi and bacteria: tripartite metatranscriptomics of a root microbiome responding to soil contamination. *Microbiome*, 6, 53.

Goodman, R. M. (1997). A dirty look: soil microflora and rhizosphere biology. *In* Radical biology: advances and perspectives on the function of plant roots. 219-231.

Gosalbes, M. J., Durban, A., Pignatelli, M., Abellan, J. J., Jimenez-Hernandez, N., Perez-Cobas, A. E., Latorre, A. & Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PloS one,* 6, e17447.

Gottesman, S. & Storz, G. (2011). Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring, *Harb. Perspect. Biol*. 3.

Goyer, R. A. (2001). Toxic effects of metals. *Cassarett and Doull's Toxicology: The Basic Science of Poisons*. 811-867.

Grant, S., Grant, W. D., Cowan, D. A., Jones, B. E., Ma, Y. & Ventosa, A. (2006). Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. *Appl. Environ. Microbiol.72*, 135-143.

Gregory, T. R. (2004). Macroevolution, hierarchy theory, and the C-value enigma. *Paleobiology*. 30, 179-202.

Griffiths, B. S. & Philippot, L. (2013). Insights into the resistance and resilience of the soil microbial community. *FEMS Microbiol. Rev.* 37, 112-129.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F. & Smirnova, T. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, 699-704.

Grossmann, L., Jensen, M., Heider, D., Jost, S., Glucksman, E., Hartikainen, H., Mahamdallie, S. S., Gardner, M., Hoffmann, D., Bass, D. & Boenigk, J. (2016). Protistan community analysis: key findings of a large-scale molecular sampling. *ISME J.* 10, 2269-2279.

Groussman, R. D., Coesel, S., Durham, B. P., Morales, R. & Armbrust, V. (2020). Diel Gene Expression of Eukaryotic Microplankton in the North Pacific Subtropical

Gyre: Metatranscriptomic Windows into Temporal Function, Ecology, and Evolution of Uncultured Microbes. In *Ocean Sciences Meeting 2020*. AGU.

Gubler, U. & Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene*. 25, 263-269.

Guo, L., Sui, Z., Zhang, S., Ren, Y. & Liu, Y. (2015). Comparison of potential diatom "barcode" genes (18S and ITS rDNA, COI, *rbcL*) and their effectiveness in discriminating and determining species taxonomy in Bacillariophyta. *Int. J. Syst. Evol. Microbiol.* 65, 1369-1380.

Guo, W. J., Meetam, M., & Goldsbrough, P. (2008). Examining the specific contributions of individual *Arabidopsis* metallothioneins to copper distribution and metal tolerance. *Plant Physiol.* 164, 1697-1706.

Gyssels, G., Poesen, J., Bochet, E., & Li, Y. (2005). Impact of plant roots on the resistance of soils to erosion by water: a review. *Prog. Phys. Geog.* 29, 189-217.

Ha, S. B., Smith, A. P., Howden, R., Dietrich, W. M., Bugg, S., O'Connell, M. J., Goldsbrough, P.B. & Cobbett, C. S. (1999). Phytochelatin synthase genes from *Arabidopsis* and the yeast *Schizosaccharomycespombe*. *Plant Cell,* 11, 1153-1163.

Haag-Kerwer, A., Schafer, H. J., Heiss, S., Walter, C., &Rausch,T. (1999). Cadmium exposure in *Brassica juncea*causes a decline in transpiration rate and leaf expansion without effect on photosynthesis. *J. Exp. Bot.* 50, 1827-1835.

Hadziavdic, K., Lekang, K., & Lanzen, A. (2014). Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PLoS One9, e87624.

Hageman, J., Rujano, M. A., Van Waarde, M. A., Kakkar, V., Dirks, R. P., Govorukhina, N., Oosterveld-Hut, H. M., Lubsen, N. H. & Kampinga, H. H. (2010). A DNAJB chaperone subfamily with HDAC-dependent activities suppresses toxic protein aggregation. Mol Cell 37:355-369.

Haichar, F. E., Roncato, M. A., &Achouak, W. (2012). Stable isotope probing of bacterial community structure and gene expression in the rhizosphere of *Arabidopsisthaliana*. *FEMS Microbiol. Ecol.* 81, 291-302.

Hall, J. Á. (2002). Cellular mechanisms for heavy metal detoxification and tolerance. *J. Exp Bot.* 53, 1-11

Hambraeus, G., von Wachenfeldt, C. & Hederstedt, L. (2003). Genome-wide survey of mRNA half-lives in Bacillus subtilis identifies extremely stable mRNAs. *Mol. Genet. Genomics*, 269, 706-714.

Hamelink, J., Landrum, P. F., Bergman, H. & Benson, W. H. (1994). *Bioavailability: physical, chemical, and biological interactions*. CRC press.

Hamer, D. H. (1986). Metallothionein. *Annu. Rev. Biochem.* 55, 913-951.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.*68, 669-685.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol*. 5, R245-R249.

Handelsman, J., Tiedjem., Alvarez-Cohen, L., Ashburner, M., Cann, I. K. O., Doolittle, W. F., Fraser-Liggett, C. M., Godzik, A., Gordon, J. I. & Riley, M. (2007) The New Science of metagenomics: revealing the secrets of our microbial planet, Washington, DC: The National Academies Press.

Haon, M., Grisel, S., Navarro, D., Gruet, A., Berrin, J. G. & Bignon, C. (2015). Recombinant protein production facility for fungal biomass-degrading enzymes using the yeast *Pichia pastoris*. *Front. Microbiol.* 6, 100

Hårdeman, F. & Sjoling. S. (2007). Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol. Ecol.*59, 524-534.

Harvey, L. J., & McArdle, H. J. (2008). Biomarkers of copper status: a brief update. *Br. J. Nutr.* 99, 10-13.

Hayashi, S., Tanikawa, H., Kuramata, M., Abe, T. & Ishikawa, S. (2020). Domain exchange between Oryza sativa phytochelatin synthases reveals a region that determines responsiveness to arsenic and heavy metals. *Biochem. Biophy. Res. Commun.* 523, 548-553.

He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R. H., Warnecke, F., Tringe, S.G. & Hugenholtz, P. (2013). Comparative metagenomic and Metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One*. 8:e61126.

He, S., Wurtzel, O., Singh, K., Froula, J. L., Yilmaz, S., Tringe, S. G., Wang, Z., Chen, F., Lindquist, E. A., Sorek, R. & Hugenholtz, P. (2010). Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nature Meth. 7,* 807-812.

He, W., Megharaj, M., Wu, C. Y., Subashchandrabose, S. R., & Dai, C. C. (2020). Endophyte-assisted phytoremediation: mechanisms and current application strategies for soil mixed pollutants. *Crit. Review Biotechnol.* 40. 31-45.

Healy, F. G., Ray, R. M., Aldrich, H. C., Wilkie, A. C., Ingram, L. O. & Shanmugam, K. T. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl. Microbiol. Biotechnol.* 43, 667-674.

Heath, C., Hu, X. P., Cary, S. C. & Cowan, D. (2009). Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. *Appl. Environ. Microbiol.* 75, 4657-4659.

Hebert, P. D., Ratnasingham, S. & De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B.* 270, S96-S99.

Heck, P., Luthi, D. & Schar, C. (1999). The influence of vegetation on the summertime evolution of European soil moisture. *Phys. Chem. Earth, Part B*, 24, 609-614.

Herawati, N., Suzuki, S., Hayashi, K., Rivai, I. F. & Koyoma, H. (2000). Cadmium, copper and zinc levels in rice and soil of Japan, Indonesia and China by soil type. *Bull. Env. Contam. Toxicol.* 64, 33-39.

Hernández-Pellón, A. & Fernández-Olmo, I. (2019). Airborne concentration and deposition of trace metals and metalloids in an urban area downwind of a manganese alloy plant. *Atmos. Poll. Res.* 10, 712-721.

Hershko, A., Ciechanover, A. & Rose, I. A. (1981). Identification of the active amino acid residue of the polypeptide of ATP-dependent protein breakdown. *J. Biol. Chem.* 256, 1525-1528.

Heyninck, K. & Beyaert, R. (2005). A20 inhibits NF-_B activation by dual ubiquitin-editing functions. *Trends Biochem. Sci.* 30, 1-4.

Hishiya, A., Iemura, S., Natsume, T., Takayama, S., Ikeda, K. & Watanabe, K. (2006). A novel ubiquitin-binding protein ZNF216 functioning in muscle atrophy. EMBO J. 25, 554-564.

Hofrichter, M. & Esser, K. (2010). Mycota X–Industrial Applications (2nd edn), Springer

Hofrichter, M. & Ullrich, R. (2014). Oxidations catalyzed by fungal peroxygenases. *Curr. Opin. Chem. Biol.* 19, 116-125.

Hori, C., Gaskell, J., Cullen, D., Sabat, G., Stewart, P. E., Lail, K., Peng, Y., Barry, K., Grigoriev, I.V., Kohler, A. & Fauchery, L. (2018). Multi-omic analyses of extensively decayed Pinus contorta reveal expression of a diverse array of lignocellulose-degrading enzymes. *Appl. Environ. Microbiol.* 84,20.

Horner-Devine, M. C., Leibold, M. A., Smith, V. H. & Bohannan, B. J. M. (2003). Bacterial diversity patterns along a gradient of primary productivity. *Ecol. Lett.,* 6: 613-622.

Housecroft, C. E., & Alan G. S. (2008). "Chapter 15: The group 15 elements." *Inorg. Chem.*, 449.

Howden, R., Andersen, C. R., Goldsbrough, P. B. & Cobbett, C. S. (1995b). A cadmium-sensitive, glutathione-deficient mutant of *Arabidopsis thaliana*. *Plant Physiol*. 107, 1067-1073.

Howden, R., Goldsbrough, P. B., Andersen, C. R. & Cobbett, C. S. (1995). Cadmium-sensitive, *cad1*, mutants of *Arabidopsis thaliana* are phytochelatin deficient. *PlantPhysiol.*107, 1059-1066.

Hseu, Z. Y. (2004). Evaluating heavy metal contents in nine composts using four digestion methods. *Bioresour. Technol.* 95, 53-59.

Huang, G. Y., Ying, G. G., Liang, Y. Q., Liu, S. S. & Liu, Y. S. (2014). Expression patterns of metallothionein, cytochrome P450 1A and vitellogenin genes in western mosquitofish (*Gambusiaaffinis*) in response to heavy metals. *Ecotoxicol. Environ. Saf.* 105, 97-102.

Hugenholtz, P. & Pace, N. R. (1996). Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends Biotechnol.* 14, 190-197.

Hugerth, L. W., Muller, E. E., Hu, Y. O., Lebrun, L. A., Roume, H., Lundin, D., Wilmes, P. & Andersson, A. F. (2014). Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PloS one,* 9, e95567.

Hurt, R. A., Qiu, X., Wu, L., Roh, Y., Palumbo, A. V., Tiedje, J. M. & Zhou, J. (2001). Simultaneous recovery of RNA and DNA from soils and sediments. *Appl. Environ. Microbiol.* 67, 4495-4503.

Hutchison III, C. A. & Venter, J. C. (2006). Single-cell genomics. *Nat. Biotechnol.* 24, 657.

Hwang, D. S., Lee, K. W, Han, J., Park, H. G., Lee, J., Lee,Y. M. Lee, J. S. (2010). Molecular characterization and expression of vitellogenin (Vg) genes from the cyclopoid copepod, *Paracyclopina nana* exposed to heavy metals. *Comp. Biochem. Physiol.*151, 360-368.

Ireland, H. E., Harding, S. J., Bonwick, G. A., Jones, M., Smith, C. J. & Williams, J. H. H. (2004). Evaluation of heat shock protein 70 as a biomarker of environmental stress in *Fucus serratus* and *Lemna minor*. *Biomarkers.* 9, 139-155

Ivanova, A. A., Wegner, C. E., Kim, Y., Liesack, W. & Dedysh, S. N. (2016). Identification of microbial populations driving biopolymer degradation in acidic peatlands by metatranscriptomic analysis. *Mol. Ecol.* 25, 4818-4835.

Jiang, C., Ma, G., Li, S., Hu, T., Che, Z., Shen, P., Yan, B. & Wu, B.(2009). Characterization of a novel beta-glucosidase-like activity from a soil metagenome. *J. Microbiol.* 47, 542-548.

Jiao, W., Chen, W., Chang, A. C. & Page, A. L. (2012). Environmental risks of trace elements associated with long-term phosphate fertilizers applications: a review. *Environ. Poll.* 168, 44-53.

Jones, E. W. & Fink, G. R. (1982). Regulation of amino acid and nucleotide biosynthesis in yeast. In *The Molecular Biology of the Yeast* Saccharomyces. *Metabolism and Gene Expression*, Strathern JN, Jones EW, Broach JR (eds). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY; 181-299.

Jouquet, P., Dauber, J., Lagerlöf, J., Lavelle, P. & Lepage, M. (2006). Soil invertebrates as ecosystem engineers: intended and accidental effects on soil and feedback loops. *Appl. Soil Ecol.* 32, 153-164.

Jubeaux, G., Audourd-Combe, F., Simon, R., Tutundijn, R., Salvador, A., Geffard, O. & Chaumot A. (2012). Vitellogenin- like proteins among invertebrate species diversity: Potential of proteomic mass spectrometry for biomarker development. *Environ. Sci. Technol.* 46, 6315- 6323.

Junemann, S., Prior, K., Szczepanowski, R., Harks, I., Ehmke, B., Goesmann, A., Stoye, J. & Harmsen, D.(2012). Bacterial community shift in treated periodontitis patients revealed by Ion Torrent 16S rRNA gene amplicon sequencing. *PLoS One* 7, e41606

Kabata-Pendias, A. & Pendias, H. (2001). Trace Elements in Soils and Plants. CRC Press, Washington, D.C.

Kaeberlein, T., Lewis, K. & Epstein, S. S. (2002). Isolating “uncultivable” microorganismsin pure culture in a simulated natural environment. *Science.* 296, 1127-1129.

Kalsotra, T., Khullar, S., Agnihotri, R. & Reddy, M. S. (2018). Metal induction of two metallothionein genes in the ectomycorrhizal fungus *Suillus himalayensis* and their role in metal tolerance. *Microbiol.* 164, 868-876.

Kammenga, J., Dallinger, R., Donker, M. H., Kohler, H. R., Simonsen, V. R. T. & Weeks, J. M. (2000). Biomarkers in terrestrial invertebrates: potential and limitations for ecotoxicological soil risk assessment. *Rev. Environ. Contam. T.* 164: 93-147.

Kaplan, O., Yildirim, N. C., Yildirim, N. & Cimen, M. (2011). Toxic elements in animal products and environmental health.

Karpinets, T. V., Greenwood, D. J., Sams, C. E. & Ammons, J. T. (2006). RNA: protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biol.* 4, 30.

Kasprzak, K. S., Sunderman Jr, F. W. & Salnikow, K. (2003). Nickel carcinogenesis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 533, 67-97.

Kato, S., Sekine, S., Oh, S. W., Kim, N. S., Umezawa, Y. & Abe, N. (1994). Construction of a human full-length cDNA bank. *Gene*. 150, 243-250.

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., & Amaral-Zettler, L. A. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*. 12, e1001889.

Kellner, H. & Vandenbol, M. (2010). Fungi unearthed: transcripts encoding lignocellulolytic and chitinolytic enzymes in forest soil. *PLoS One*, 5, e10971.

Kellner, H., Luis, P., Portetelle, D. & Vandenbol, M. (2011). Screening of a soil metatranscriptomic library by functional complementation of *Saccharomycescerevisiae* 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. Bio.* 18,355–364.

Khatri, I., Tomar, R., Ganesan, K., Prasad, G. S. & Subramanian, S. (2017). Complete genome sequence and comparative genomics of the probiotic yeast Saccharomyces boulardii. *Sci. Rep.* 7, 1-12.

Kim, Y. & Liesack, W. (2015). Differential assemblage of functional units in paddy soil microbiomes. *PLoS One*, 10, e0122221.

Kinlaw, A. E. (2006). Burrows of semi-fossorial vertebrates in upland communities of Central Florida: their architecture, dispersion, and ecological consequences (Doctoral dissertation, University of Florida).

Kirpichtchikova, T. A., Manceau, A., Spadini, L., Panfili, F., Marcus, M. A., & Jacquet, T. (2006). Speciation and solubility of heavy metals in contaminated soil using X-ray microfluorescence, EXAFS spectroscopy, chemical extraction, and thermodynamic modeling. *Geochimt. Cosmochim. Acta*. 70, 2163-2190.

Kitson, R. E. & Mellon, M. G. (1944). Colorimetric determination of phosphorus as molybdovanadophosphoric acid. *Ind. Eng. Chem.* 16: 379-383.

Klaassen, C. D. & Lehman-Mckeeman, L. D. (1989). Induction of metallothionein. *J the Am. Coll. Toxicol.* 8, 1315-1321.

Klapheck, S., Schlunz, S. & Bergmann, L. (1995) Synthesis of phytochelatins and homo-phytochelatins in *Pisumsativum* L. *Plant Physiol.*107, 515-521.

Klug, A. & Rhodes, D. (1987). ‘Zinc fingers’: a novel protein motif for nucleic acid recognition. *Trends Biochem. Sci.* 12, 464-469.

Ko, H. J., Kim, K. Y., Kim, H. T., Kim, C. N. & Umeda, M. (2008). Evaluation of maturity parameters and heavy metal contents in composts made from animal manure. *Waste Manage.* 28, 813-820.

Kodrík, D., Bednářová, A., Zemanová, M. & Krishnan, N. (2015). Hormonal regulation of response to oxidative stress in insects-an update. *Int. J. Mol. Sci.* 16, 25788-25816.

Köhler, H. R. & Eckwert, H. (1997). The induction of stress proteins (hsp) in *Oniscusasellus* (Isopoda) as a molecular marker of multiple heavy metal exposure. II. Joint toxicity and transfer to field situations. *Ecotoxicol*. 6, 263-274.

Köhler, H. R., Triebskorn, R., Stocker, W., Kloetzel, P. & Alberti, G. (1992). The 70 kD heat shock protein (hsp70) in soil invertebrates: a possible tool for monitoring environmental toxicants. *Arch. Environ. Contam. Toxicol*. 22, 334-338.

Konstantinidis, K. T. & Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* 10, 504-509.

Korf, I. (2013). Genomics: the state of the art in RNA-seq analysis. *Nat. Methods*. 10, 1165-6.

Kovacs, E., Nyitrai, P., Czovek, P., Ovari, M. & Keresztes, A. (2009). Investigation into the mechanism of stimulation by low-concentration stressors in barley seedlings. *J. Plant Physiol.* 166, 72-79.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112-5120.

Kumar, R., Mishra, R.K., Mishra, V., Qidwai, A., Pandey, A., Shukla, S.K., Pandey, M., Pathak, A. and Dikshit, A. (2016). Detoxification and tolerance of heavy metals in plants. In *Plant Metal Interaction,* 335-359.

Kuramae, E. E., Hillekens, R. H., Hollander, M. D., van der Heijden, M. G., van den Berg, M., van Straalen, N. M. & Kowalchuk, G. A. (2013). Structural and functional variation in soil fungal communities associated with litter bags containing maize leaf. *FEMS Microbiol. Ecology*, 84, 519-531.

Lafond, M., Navarro, D., Haon, M., Couturier, M., & Berrin, J. G. (2012). Characterization of a broad-specificity β-glucanase acting on β-(1, 3)-, β-(1, 4)- and β-(1, 6)-glucans that defines a new glycoside hydrolase family. *Appl. Environ. Microbiol.* 78, 8540-8546.

Lakay, F. M., Botha, A. & Prior, B. A. (2007). Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J. Appl. Microbiol.* 102, 265-273.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K. & Hartl, F. U. (1992). Successive action of DnaK DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature.* 356, 683-689.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. 9, 357-9

Laughton, S., Laycock, A., von der Kammer, F., Hofmann, T., Casman, E. A., Rodrigues, S. M. & Lowry, G. V. (2019). Persistence of copper-based nanoparticle-containing foliar sprays in *Lactucasativa* (lettuce) characterized by spICP-MS. *J. Nanopart Res.* 21, 174.

Lavelle, P. (2012). Soil as a habitat. *Soil ecology and ecosystem services*, 1-27.

Lazor, P., Tomáš, J., Tóth, T., Tóth, J. & Čéryová, S. (2020). Monitoring of air pollution and atmospheric deposition of heavy metals by analysis of honey. *J. Microbiol. Biotechnol Food Sci.* 9, 522-533.

Leach, M. D., Stead, D. A., Argo, E., MacCallum, D. M. & Brown, A. J. (2011). Molecular and proteomic analyses highlight the importance of ubiquitination for the stress resistance, metabolic adaptation, morphogenetic regulation and virulence of *Candidaalbicans*. *Mol. Microbiol.* 79, 1574-1593.

Lee, J. M., Nishimori, Y., Hatakeyama, M., Bae, T. W., Oishi, K. (2000). Vitellogenin of the cicada *Graptopsaltrianigrofuscata* (Homoptera), Analysis of its primary structure. *Insect Biochem. Mol. Biol.* 30, 1-7.

Lee, K. E. & Foster, R. C. (1991). Soil fauna and soil structure. *Soil Res.* 29, 745-775.

Lee, K. W., Hwang, D. S., Rhee, J. S., Ki, J. S., Park, H. G., Ryu, J. C., Raisuddin, S. & Lee, J. S. (2008). Molecular cloning, phylogenetic analysis and developmental expression of vitellogenin (Vg) gene from the intertidal copepod *Tigriopus japonicus. Comp. Biochem. Physiol. B.* 150, 395-402.

Lee-Pow, C. S., Yost, E. E., Aday, D. D., Kullman, S. W. (2016). Sharing the Roles: An Assessment of Japanese Medaka Estrogen Receptors in Vitellogenin Induction. *Environ. Sci. Technol.* 50, 8886-8895.

Lehembre, F., Doillon, D., David, E., Perrotto, S., Baude, J., Foulon, J., Harfouche, L., Vallon, L., Poulain, J., Da Silva, C. &Wincker, P. (2013). Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. *Environ. Microbiol*. 15, 2829-2840.

Lemire, J. A., Harrison, J. J. & Turner, R. J. (2013). Antimicrobial activity of metals: Mechanisms, molecular targets and applications. *Nat. Rev. Microbiol.*11, 371-384.

Lesaulnier, C., Papamichail, D., McCorkle, S., Ollivier, B., Skiena, S., Taghavi, S., Zak, D. & Van Der Lelie, D. (2008). Elevated atmospheric CO2 affects soil microbial diversity associated with trembling aspen. *Environ. Microbiol.* 10. 926-941.

Levesque, V., Fayad, T., Ndiaye, K., Nahe, Diouf, M. & Lussier, J. G. (2003). Size selection of cDNA libraries for the cloning of cDNAs after suppression subtractive hybridization. *Biotechniques*.35, 72-78.

Levins, I. & Gosk, E. (2008). Trace elements in groundwater as indicators of anthropogenic impact. *Environ. Geol*.55, 285-290.

Lewis, S., Donkin, M. E. & Depledge, M. H. (2001). Hsp70 expression in *Enteromorphaintestinalis* (Chlorophyta) exposed to environmental stressors. *Aquat. Toxicol*. 51, 277-291.

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$^{\Delta}$-catalyzed transport of bis(glutathionato) cadmium. *Proc. Natl. Acad. Sci. U.S.A.* 94, 42-47.

Li,L. & Kaplan, J. (1998). Defects in the yeast high affinity iron transport system result in increased metal sensitivity because of the increased expression of transporters with a broad transition metal specificity. *J. Biol. Chem. 273*, 22181-22187.

Liberek, K., Marszalek, J., Ang, D., Georgopoulos, C. & Zylicz, M. (1991). *Escherichiacoli*DnaJ and GrpE heat shock proteins jointly stimulateATPase activity of DnaK. *Proc. Natl. Acad. Sci. U S A.* 88, 2874-2878

Lin, Y. F., Severing, E., te Lintel Hekkert, B., Schijlen, E. & Aarts, M. G. M. (2014). A comprehensive set of transcript sequences of the heavy metal hyperaccumulator *Noccaeacaerulescens*. *Front. Plant Sci.* 5, 261.

Lin, Y., Vogt, R. & Larssen, T. (2012). Environmental mercury in China: a review. *Environ. Toxicol. Chem.* 31, 2431-2444.

Ling, W., Shen, Q., Gao, Y., Gu, X. & Yang, Z. (2008). Use of bentonite to control the release of copper from contaminated soils. *Soil Res.* 45, 618-623.

Linnen, J., Bailey, M., & Week, C. P. (1993). Two related localized mRNAs from *Xenopuszaevis* encode ubiquitin-like fusion proteins. *Gene,* 128, 181-188.

Lisetskii, F. N., & Pichura, V. I. (2016). Assessment and forecast of soil formation under irrigation in the steppe zone of Ukraine. *Russi. Agri. Sci.* 42, 155-159.

Liu, G. Y., Zhang, Y. X. & Chai, T. Y. (2011). Phytochelatin synthase of *Thlaspi caerulescens* enhanced tolerance and accumulation of heavy metals when expressed in yeast and tobacco. *Plant Cell Rep.* 30, 1067-1076.

Liu, W. T., Marsh, T. L., Cheng, H. & 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. J. A., Finley, B. K., Mau, R. L., Schwartz, E., Dijkstra, P., Bowker, M. A. & Hungate, B. A. (2020). The soil priming effect: Consistent across ecosystems, elusive mechanisms. *Soil Biol. Biochem.* 140, 107617.

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ method. *Methods*. 25, 402-408.

Loeffler, S., Hochberger, A., Grill, E., Winnacker, E. L. & Zenk, M. H. (1989). Termination of the phytochelatin synthase reaction through sequestration of heavy metals by the reaction product. *FEBS Lett.* 258, 42-46.

Longo, V. D., Gralla, E. B. & Valentine, J. S., 1996. Superoxide dismutase activity is essential for stationary phase survival in Saccharomyces cerevisiae. *J. Biol. Chem.* 271, 12275-12280.

López, S. L., Aiassa, D., Benítez-Leite, S., Lajmanovich, R., Manas, F., Poletta, G., Sánchez, N., Simoniello, M. F. & Carrasco, A. E. (2012). Pesticides used in South American GMO-based agriculture: A review of their effects on humans and animal models. In *Advances in Molecular Toxicology*, Elsevier, 6, 41-75.

Loppi, S., Putorti, E., Pirintsos, S. A. & De Dominicis, V. (2000). Accumulation of heavy metals in epiphytic lichens near a municipal solid waste incinerator (Central Italy). *Environ. Monit. Assess.* 61, 361-371.

Lozupone, C. & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228-8235.

Lozupone, C., Hamady, M. & Knight, V. (2006). UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinfo.* 7, 371.

Lu, Z. & Cyr, D. M. (1998). The conserved carboxyl terminus and zinc finger like domain of the co-chaperone Ydj1 assist Hsp70 in protein folding. *J. Biol. Chem.* 273:5970-5978.

Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Buchner, A., Lai, T., Steppi, S. & Jobb, G. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363-1371.

Lueders, T., Wagner, B., Claus, P. & Friedrich, M. W. (2004). Stable isotope probing of rRNA and DNA reveals a dynamic methylotroph community and trophic interactions with fungi and protozoa in oxic rice field soil. *Environ. Microbiol.* 6, 60-72.

Lundell, T. K., Mäkelä, M. R., de Vries, R. P. & Hilden, K. S. (2014). Genomics, lifestyles and future prospects of wood-decay and litter-decomposing basidiomycota. *Adv. Bot. Res.* 70, 329-370.

Lundholm, N., Moestrup. O., Kotaki, Y., Hoef-Emden, K., Scholin. C. & Miller, P. (2006). Inter-and intraspecific variation of the *Pseudo-nitzschiadelicatissima* complex (Bacillariophyceae) llustrated by rRNA probes, morphological data and phylogenetic analyses. *J. Phycol.*42, 464-481.

Lynch, J. (1995). Root architecture and plant productivity. *Plant Physiol.* 109, 7.

Ma, J. F., Hiradate, S. & Matsumoto, H. (1998). High aluminum resistance in buckwheat: II. Oxalic acid detoxifies aluminum internally. *Plant Physiol.* 117, 753-759.

Maimaiti, J., Zhang, Y., Yang, J., Cen, Y.P., Layzell, D. B., Peoples, M. & Dong, Z. (2007). Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environ. Microbiol.* 9, 435-444.

Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods.* 9, 1185-8.

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

Marmeisse, R., Kellner, H., Fraissinet-Tachet, L. & Luis, P. (2017). Discovering Protein-Coding Genes from the Environment: Time for the Eukaryotes? *Trends Biotechnol.* 35, 824-835.

Marrugo-Negrete, J., Pinedo-Hernández, J. & Díez, S. (2017). Assessment of heavy metal pollution, spatial distribution and origin in agricultural soils along the Sinú River Basin, Colombia. *Environ. Res.* 154, 380-388.

Marschner, H. (1995). Mineral nutrition of higher plants, 2nd edn. London: Academic Press.

Martín-Díaz, M. L., Sales, D. & DelValls, Á. (2008). Toxicokinetic approach for the assessment of endocrine disruption effects of contaminated dredged material using female *Carcinusmaenas*. *Ecotoxicol*.17, 495-503.

Martinez, A., Kolvek, S. J., Yip, C. L. T., Hopke, J., Brown, K. A., MacNeil, I. A. & Osburne, M. S. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* 70, 2452-2463.

Maslin, P. & Maier, R. M. (2000). Rhamnolipid-enhanced mineralization of phenanthrene in organic-metal co-contaminated soils. *Bioremediat. J*. 4, 295-308.

Matozzo, V., Gagné, F., Marin, M. G., Ricciardi, F. & Blaise, C. (2008). Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review. *Environ. Int.* 34, 531-545.

McBride, M. B. & Spiers, G. (2001). Trace element content of selected fertilizers and dairy manures as determined by ICP–MS *Commun. Soil Sci. Plant Anal.* 32, 139-156.

McCarren, J., Becker, J.W., Repeta, D.J., Shi, Y., Young, C.R., Malmstrom, R.R., Chisholm, S.W. & DeLong, E. F. (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16420–16427.

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Gary L Andersen G. L., Knight, R. & Hugenholtz, P. (2012). Improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*. 6, 610-618.

McGrath, K.C., Thomas-Hall, S.R., Cheng, C.T., Leo, L., Alexa, A., Schmidt, S. & Schenk, P.M. (2008). Isolation and analysis of mRNA from environmental microbial communities. *J. Microbiol. Methods*. 75, 172-176.

McGrath, S. P., Shen, Z. G. & Zhao, F. J. (1997). Heavy metal uptake and chemical changes in the rhizosphere of *Thlaspi caerulescens* and *Thlaspi ochroleucum* grown in contaminated soils. *Plant Soil*. 188, 153-159.

McKew, B. A., Dumbrell, A. J., Taylor, J. D., McGenity, T. J. & Underwood, G. J. (2013). Differences between aerobic and anaerobic degradation of microphytobenthic biofilm-derived organic matter within intertidal sediments. *FEMS Microbiol. Ecol.* 84, 495-509.

McLean, J. S., Lombardo, M. J., Ziegler, M. G., Novotny, M., Yee-Greenbaum, J., Badger, J. H., Tesler, G., Nurk, S., Lesin, V., Brami, D. & Hall, A. P. (2013). Genome of the pathogen Porphyromonasgingivalis recovered from a biofilm in a hospital sink using a high-throughput single-cell genomics platform. *Genome Res. 23*, 867-877.

McMurray, C. T. & Tainer, J. A. (2003). Cancer, cadmium and genome integrity. *Nat. Genet. 34*, 239-241.

Meharg, A. A. (2003). The mechanistic basis of interactions between mycorrhizal associations and toxic metal cations. *Mycol. Res.* 107, 1253-1265.

Meysman, F. J., Middelburg, J. J. & Heip, C. H. (2006). Bioturbation: a fresh look at Darwin's last idea. *Trends Ecol. Evol.* 21, 688-695.

Mitchell, R. L. & Burridge, S. C. (1979) Trace Element in soils and crops. *Phil.Trans. Royal. Soc.* London B 288: 15-24.

Molina, M., Aburto, F., Calderón, R., Cazanga, M. & Escudey, M. (2009). Trace element composition of selected fertilizers used in Chile: phosphorus fertilizers as a source of long-term soil contamination. *Soil Sediment Contam.*18, 497-511.

Montorzi, M., Falchuk, K. H. & Vallee, B. L. (1994). *Xenopuslaevis*vitellogenin is a zinc protein. *Biochem. Biophy. Res. Commun.* 200, 1407-1413.

Moor, C., Lymberopoulou, T. & Dietrich V. J. (2001). Determination of heavy metals in soils, sediments and geological materials by ICP-AES and ICP-MS. *Microchim. Acta.* 136, 123-128.

Morel, J. L. (1996). Bioavailability of Trace Elements. *Soil. Ecotoxicol.* 141.

Morgan, X. C. & Huttenhower, C. (2014). Meta'omic analytic techniques for studying the intestinal microbiome. *Gastroenterology*. 146, 1437-48.

Morris, C. A., Nicolaus, B., Sampson, V., Harwood, J. L. & Kille, P. (1999). Identification and characterization of a recombinant metallothionein protein from a marine alga, Fucus vesiculosus. *Biochem. J.* 338, 553-560.

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.

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.

Mukhopadhyay, A., Vij, S. & 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. U. S. A.* 101, 6309-6314.

Müller-Taubenberger, A., Westphal, M., Noegel, A. & Gerisch, G. (1989). A developmentally regulated gene product from Dictyosteliumdiscoideum shows high homology to human α-L-fucosidase. *FEBS lett.* 246, 185-192.

Murase, J. & Frenzel, P. (2007). A methane-driven microbial food web in a wetland rice soil. *Environ. Microbiol.* 12, 3025-3034.

Naumann, B., Eberius, M. &Appenroth, K. J. (2007) Growth rate based dose–response relationships and EC-values of ten heavy metals using the duckweed growth inhibition test (ISO 20079) with *Lemna minor* L. clone St. *J. Plant. Physiol.* 164, 1656-1664

Neumann, D., Lichtenberger, O., Gunther, D., Tschiersch, K. &Nover, L. (1994). Heat-shock proteins induce heavy metal tolerance in higher plants. *Planta* 194, 360-367.

Neumann, D., ZurNieden, U., Lichtenberger, O. & Leopold, I. (1995). How does *Armeriamaritime* tolerate high heavy metal concentrations. *J. Plant Physiol.* 146, 704-717.

Nguyen, C. (2003). Rhizodeposition of organic C by plants: mechanisms and controls. *Agronomie*, 23, 375-396.

Nickens, K. P., Patierno, S. R. & Ceryak, S. (2010). Chromium genotoxicity: a double-edged sword. *Chem. Biol. Interact.* 188, 276-288.

Novillo, A., Won, S. J., Li, C. & Callard, I. P. (2005). Changes in nuclear receptor and vitellogenin gene expression in response to steroids and heavy metal in Caenorhabditis elegans. *Int. Comp.Biol.* 45, 61-71.

Nriagu, J. O. (1989). A global assessment of natural sources of atmospheric trace metals. *Nature*. 338, 47-49.

Nyitrai, P., Mayer, M., Ovari, M. & Keresztes, A. (2007). Involvement of the phosphoinositide signalling pathway in the anti-senescence effect of low-concentration stressors on detached barley leaves. *Plant Biol.* 9, 420-426.

O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J. M., & Vilgalys, R. (2005). Fungal community analysis by large-scale sequencing of environmental samples. *Appl. Environ. Microbiol.* 71: 5544-5550.

Olsen, S. R., Cole, C.V., Watanabe, F. S. & Dean, L. A. (1954). Estimation of available phosphorus in soils by extraction with sodium bicarbonate. U.S. Government printing office. Washington (DC): US Department of Agriculture. Circular No. 939.

Olsson, S., Puene-Sanchez, F., Gomez, M. J. & Aquilera, A. (2015). Transcriptional response to copper excess and identification of genes involved in heavy metal tolerance in the extremophilic microalga *Chlamydomonasacidophila*. *Extremophile.* 19, 657-672.

Opel, K. L., Chung, D. & McCord, B. R. (2010). A study of PCR inhibition mechanisms using real-time PCR. *J. Forensic Sci.* 55, 25-33.

Orsini, M. & Romano-Spica, V. (2001). A microwave-based method for nucleic acid isolation from environmental samples. *Lett. Appl. Microbiol.* 33, 17-20.

Oves, M., Sagir Khan M., Huda Qari, A., NadeenFelemban, M. & Alkeelbi, T. (2016). Heavy metals: Biological importance and detoxification strategies. *J. Bioremediat. Biodegrad.*7, 334-348.

Pace, N. R. (1985). Analyzing natural microbial populations by rRNA sequences. *ASM news*, 51, 4-12.

Pacyna, J. M. (1996). Monitoring and assessment of metal contaminants in the air. *Toxicology of metals*, 9-28.

Page, A. L. & Chang, A. C. (1979). Contamination of soil and vegetation by atmospheric deposition of trace elements. *Phytopathology*. 69, 1007-1011.

Pagenkopp Lohan, K. M., Fleischer, R. C., Carney, K. J., Holzer, K. K. & Ruiz, G. M. (2016). Amplicon-based pyrosequencing reveals high diversity of protistan parasites in ships' ballast water: implications for biogeography and infectious diseases. *Microb. Ecol.* 71, 530-542.

Pal, R. & Rai, J. P. N. (2009). Phytochelatins: peptides involved in heavy metal detoxification. *Appl. Biochem. Biotech.* 160, 945-963.

Papaspyrou, S., Smith, C. J., Dong, L. F., Whitby, C., Dumbrell, A. J. & Nedwell, D. B. (2014). Nitrate reduction functional genes and nitrate reduction potentials persist in deeper estuarine sediments. Why?. *PloS one*, 9, e94111.

Paradelo, R., Villada, A., Devesa-Rey, R., Moldes, A. B., Domínguez, M., Patiño, J. & Barral, M. T. (2011). Distribution and availability of trace elements in municipal solid waste composts. *J. Environ. Monitor.*13, 201-211.

Parat, C., Chaussod, R., Lévêque, J. & Andreux, F. (2005). Long-term effects of metal-containing farmyard manure and sewage sludge on soil organic matter in a fluvisol. *Soil Biol. Biochem.* 37, 673-679.

Pathak, G. P., Ehrenreich, A., Losi, A., Streit, W. R. & Gartner, W. (2009). Novel blue light-sensitive proteins from a metagenomic approach. *Environ. Microbiol.* 11, 2388-2399.

Patlolla, A., Barnes, C., Yedjou, C., Velma, V. & Tchounwou, P. B. (2009). Oxidative stress, DNA damage and antioxidant enzyme activity induced by hexavalent chromium in Sprague Dawley rats. *Environ*. *Toxicol*. 24, 66-73.

Pearson, R. G. (1968). Hard and soft acids HSAB. 1. Fundamental principles. *J. Chem. Educ.* 45, 581-587.

Pena, L. B., Pasquini, L. A., Tomaro, M. L., &Gallego, S. M. (2007). 20S proteasome and accumulation of oxidized and ubiquitinated proteins in maize leaves subjected to cadmium stress. *Phytochemistry.* 68, 1139-1146.

Peng, H., Yang, J., Li, G., You, Q., Han, W., Li, T., Gao, D., Xie, X., Lee, B.H., Du, J. & Hou, J. (2017). Ubiquitylation of p62/sequestosome1 activates its autophagy receptor function and controls selective autophagy upon ubiquitin stress. *Cell Res.* 27, 657-674.

Persans, M. W., Yan, X., Patnoe, J. M. M., Kramer, U. & Salt, D. E. (1999). Molecular dissection of the role of histidine in nickel hyperaccumulation in *Thlaspigoesingense* (Hálácsy). *Plant Physiol.* 121, 1117-1126.

Pickart, C. M. & Fushman, D. (2004). Polyubiquitin chains: polymeric protein signals. *Curr. Opin. Chem. Biol.* 8, 610–616.

Piechalak, A., Tomaszewska, B., Baralkiewicz, D. & Malecka, A. (2002). Accumulation and detoxification of lead ions in legumes. *Phytochemistry.* 60, 153-162.

Pierzynski, G. M., Baker, L. R., Hettiarachchi, G. M. & Scheckel, K. G. (2008). The use of soil amendments for the remediation of heavy metal-contaminated sites. In *Proceedings of the 14th International Conference on Heavy Metals in the Environment*, 16-23.

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. Bombay: Hans Publications.

Plassart, P., Terrat, S., Thomson, B., Griffiths, R., Dequiedt, S., Lelievre, M., Regnier, T., Nowak, V., Bailey, M., Lemanceau, P. & Bispo, A. (2012). Evaluation of the ISO standard 11063 DNA extraction procedure for assessing soil microbial abundance and community structure. *PloS one*, 7.

Podkaminski, D. & Vogel, J. (2010). Small RNAs promote mRNA stability to activate the synthesis of virulence factors. *Mol. Microbiol.* 78, 1327-1331.

Ponge, J. F. (2015). The soil as an ecosystem. Biology and fertility of soils, 51, 645-648.

Prasad, M. N. V. (1999). Metallothioneins and metal binding complexes in plants. Heavy metal stress in plants: from molecules to ecosystems. Berlin: Springer-Verlag, 51-72.

Prokopowich, C., Gregory, T. & Crease, T. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46, 48-50.

Prosser, J. I. (2007). Microorganisms cycling soil nutrients and their diversity. In *Modern soil microbiology*. CRC Press. 237-261.

Purves, D. (1977). Trace-Element Contamination of the Environment, Fundamental Aspects of Pollution Control and Environmental Science 1, Elsevier, Amsterdam.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., Selinger, L.B., Butler, G., Tsang, A., McAllister, T.A. & Forster, R.J. (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen—a metatranscriptomic approach. *PloS one*, 6(5), p.e20521.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S. & Manichanh, C. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 464, 59-65

Qiu, Y., Shi, H. C. & He, M. (2010). Nitrogen and phosphorous removal in municipal wastewater treatment plants in China: a review. *Int. J. Chem. Eng.* 10, 1155-1165.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590-596.

Qui, X. B., Shao, Y. M., Miao, S. & Wang, L. (2006). The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell. Mol. Life Sci.* 63, 2560-2570.

Quinlan, R. J., Sweeney, M. D., Leggio, L. L., Otten, H., Poulsen, J. C. N., Johansen, K. S., Krogh, K. B., Jørgensen, C. I., Tovborg, M., Anthonsen, A. & Tryfona, T. (2011). Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc.Natl. Acad. Sci.* 108, 15079-15084.

Rabinovich, M. L., Bolobova, A. V. & Vasil'chenko, L. G. (2004). Fungal decomposition of natural aromatic structures and xenobiotics: a review. *Appl. Biochem. Microbiol.* 40, 1-17.

Rai, V. K. (2002). Role of amino acids in plant responses to stress. *Biol. Plant.* 45, 481-487.

Ramesh, G., Podila, G. K., Gay, G., Marmeisse, R. & Reddy, M. S. (2009). Different patterns of regulation for the copper and cadmium metallothioneins of the ectomycorrhizal fungus *Hebeloma cylindrosporum*. *Appl. Environ. Microbiol.* 75, 2266-2274.

Ramesh, R., Kumar, K. S., Eswaramoorthi, S., & Purvaja, G. R. (1995). Migration and contamination of major and trace elements in groundwater of Madras City, India. *Environ Geol.* 25, 126-136.

Rani, A., Porwal, S., Sharma, R., Kapley, A., Purohit, H. J., & Kalia, V. C. (2008). Assessment of microbial diversity in effluent treatment plants by culture dependent and culture independent approaches. *Bioresour. Technol.* 99, 7098-7107.

Rashamuse, K., Magomani, V., Ronneburg, T. & Brady, D. (2009). A novel family VIII carboxylesterase derived from a leachate metagenome library exhibits promiscuous beta-lactamase activity on nitrocefin. *Appl. Microbiol. Biotechnol.*83, 491-500.

Rauch, S., Hemond, H. F., Barbante, C., Owari, M., Morrison, G. M., Peucker-Ehrenbrink, B. & Wass, U. (2005). Importance of automobile exhaust catalyst emissions for the deposition of platinum, palladium, and rhodium in the northern hemisphere. *Environ Sci. Technol.* 39, 8156-8162.

Rauser, W. E. (1995). Phytochelatins and related peptides. Structure, biosynthesis, and function. *Plant Physiol.* 109, 1141.

Ravid, T. & Hochstrasser, M. (2008). Diversity of degradation signals in the ubiquitin proteasome system. *Nat. Rev. Mol. Cell Biol.* 9, 679-690.

Redman, K. L. & Rechsteiner, M. (1989). Identification of the long ubiquitin extension as ribosomal protein S27a. *Nature.* 338, 438.

Redon, E., Loubiere, P. & Cocaign-Bousquet, M. (2005). Role of mRNA stability during genome-wide adaptation of *Lactococcuslactisto* carbon starvation. *J. Biol. Chem.*, 280, 36380-36385.

Reid, T., Chaganti, S. R., Droppo, I. G. & Weisener, C. G. (2018). Novel insights into freshwater hydrocarbon-rich sediments using metatranscriptomics: Opening the black box. *Water Res.* 136, 1-11.

Rhee, J. K., Ahn, D. G., Kim, Y. G., & Oh, J. W. (2005). New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl. Environ. Microbiol.* 71, 817-825.

Richards, M. P. (1989). Influence of egg production on zinc, copper and iron metabolism in the turkey hen (*Meleagrisgallopavo*). *Comp. Biochem. Physiol.* 93A, 811-817.

Richards, G. R. & Vanderpool, C. K. (2011). Molecular call and response: the physiology of bacterial small RNAs. *BBA-Gene Regul. Mech.* 1809, 525-531.

Richards, T. A. & Bass. D. (2005). Molecular screening of free-living microbial eukaryotes: Diversity and distribution using a meta-analysis. *Curr. Opin. Microbiol.* 8: 240–252.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in Drosophila. *Experientia*. 18, 571-573.

Ritz, K., McHugh, M. & Harris, J. (2003). Biological diversity and function in soils: contemporary perspectives and implications in relation to the formulation of effective indicators. OECD Expert Meeting on Soil Erosion and Soil Biodiversity Indicators, Rome, 1-11.

Rivera, S. F., Vasselon, V., Bouchez, A. & Rimet, F. (2020). Diatom metabarcoding applied to large scale monitoring networks: Optimization of bioinformatics strategies using Mothur software. *Ecol. Indic.* 109, 105775.

Robinson, N. J., Tommey, A. M., Kuske, C. & Jackson, P. J. (1993). Plant metallothioneins. *Biochem. J.* 295, 1-10.

Rodriguez-R, L. M. & Konstantinidis, K. T. (2014). Bypassing cultivation to identify bacterial species. *Microbe*, 9, 111-118.

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. & Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISMEJ.* 1,283-290.

Rondon, M. R., Raffel, S. J., Goodman, R. M. & Handelsman, J. (1999). Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacilluscereus*. *Proc. Natl. Acad. Sci. 96*, 6451-6455.

Rossman, A. Y. (1998). Protocols for an all taxa biodiversity inventory of fungi in a Costa Rican conservation area. Parkway Publishers, Inc., USA.

Rowley, E. R. & Mockler, T. C. (2011). Plant abiotic stress: insights from the genomics era. *Abiotic Stress Response in Plants-Physiological, biochemical and genetic perspectives. InTech, Rijeka: Croatia*, 221-68.

Saha, D., Chatterjee, D., Chakravarty, S. & Roychowdhury, T. (2019). Investigation of environmental-concern trace elements in coal and their combustion residues from thermal power plants in Eastern India. *Nat. Resource Research*. 28, 1505-1520.

Sahi, C., Kominek, J., Ziegelhoffer, T., Yu, H. Y., Baranowski, M., Marszalek, J. & Craig, E. A. (2013). Sequential duplications of an ancient member of the DnaJ-family expanded the functional chaperone network in the eukaryotic cytosol. *Mol. Biol. Evol.* 30:985-998.

Sahi, C., Lee, T., Inada, M., Pleiss, J. A. & Craig, E. A. (2010). Cwc23, an essential J protein critical for pre-mRNA splicing with a dispensable J domain. *Mol. Cell. Biol.* 30:33-42.

Salmela, H., Stark, T., Stucki, D., Fuchs, S., Freitak, D., Dey, A., Kent, C. F., Zayed, A., Dhayegude, K., Hokkanen, H. & Sundstorm, L. (2016). Ancient Duplications have led to functional divergence of Vitellogenin- like genes potentially involved in inflammation and oxidative stress in honey bees. *Genome Biol. Evol*. 8, 495-506.

Salt, D. E., Kato, N., Kramer, U., Smith, R. D. & Raskin, I. (1999). The role of root exudates in nickel hyperaccumulation and tolerance in accumulator and non-accumulator species of *Thlaspi*. In *Phytoremediation of contaminated soil and water*. CRC Press. 191-202.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular cloning: a laboratory anual* (No. Ed. 2). Cold spring harbor laboratory press.

Sa-Moura, B., Funakoshi, M., Tomko, R.J., Dohmen, R.J., Wu, Z., Peng, J. & Hochstrasser, M. (2013). A conserved protein with AN1 zinc finger and ubiquitin-like domains modulates Cdc48 (p97) function in the ubiquitin-proteasome pathway. *J. Biol. Chem.* 288, 33682-33696.

Scaps, P., Britel, M., Fabre, M. C., Gerard, S. & Descamps, M., (1998). Influence of heavy metals (cadmium and lead) on vitellogenin production in the centipede Lithobiusforficatus (Linne). *Bull. Soc. Zool.* 123, 43-51.

Scheu, S. (2003). Effects of earthworms on plant growth: patterns and perspectives: the 7th international symposium on earthworm ecology· Cardiff· Wales· 2002. *Pedobiologia*, 47, 846-856.

Schlesinger, M. J. & Bond, U. (1987). Ubiquitin genes. *Oxf. Surv Eukaryotic Genes*, 4, 77-91.

Schloss, P. D. & Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.*71, 1501-1506.

Schloss, P. D. & Handelsman, J. (2006). Introducing SONS, a tool that compares the membership of microbial communities. *Appl. Environ. Microbiol*. 72, 6773-6779.

Schloss, P. D., Larget, B. R. & Handelsman, J. (2004). Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl. Environ. Microbiol.*70:5485-5492.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J. & Sahi, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537-7541.

Schmidt, T. M., DeLong, E. F. & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* 173, 4371-4378.

Scott-Fordsmand, J. J. & Weeks, J. M. (2000). Biomarkers in earthworms: a review. *Rev. Environ. Contam. T.* 165: 117-159.

Schmieder, R. & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863-864.

Seehuus, S. C., Norberg, K., Gimsa, U., Krekling, T. & Amdam, G. V. (2006). Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proc. Natl. Acad. Sci. U. S. A.* 103, 962-967.

Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M. & Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* 13, 216-223.

Senesil, G. S., Baldassarre, G., Senesi, N. & Radina, B. (1999). Trace element inputs into soils by anthropogenic activities and implications for human health. *Chemosphere*. 39, 343-377.

Shallari, S., Schwartz, C., Hasko, A. & Morel, J. L. (1998). Heavy metals in soils and plants of serpentine and industrial sites of Albania. *Sci. Total Environ.* 19209, 133-142.

Sharma, R., Ranjan, R., Kapardar, R. K., & Grover, A. (2005). 'Unculturable' bacterial diversity: An untapped resource. *Curr. Sci.* 72-77.

Sharma, S. K., Goloubinoff, P. & Christen, P. (2008). Heavy metal ions are potent inhibitors of protein folding. *Biochem. Biophys. Res. Commun.*, 372, 341-345.

Sharma, S. K., Goloubinoff, P. & Christen, P. (2011). Non-Native Proteins as Newly Identified Targets of Heavy Metals and Metalloids. In *Cellular Effects of Heavy Metals*, Springer: Heidelberg, Germany, 263-274.

Sharma, S. S. & Dietz, K. J. (2006). The significance of amino acids and amino acid-derived molecules in plant responses and adaptation to heavy metal stress. *J. Exp. Bot.* 57, 711-726.

Sharma, V., Shukla, R.K., Saxena, N., Parmar, D., Das, M. & Dhawan, A. (2009). DNA damaging potential of zinc oxide nanoparticles in human epidermal cells. *Toxicol. Let.* 185, 211-218.

Shen, Y., Meunier, L. & Hendershot, L. M. (2002). Identification and characterization of a novel endoplasmic reticulum (ER) DnaJ homologue which stimulates ATPase activity of BiP*in vitro* and is induced by ER stress. *J. Biol. Chem.* 277, 15947-15956.

Shi, Y. M., Tyson, G. W. & DeLong, E. F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature*. 459, 266-269.

Shi, Y. M., Tyson, G. W., Eppley, J. M. & DeLong, E. F. (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J*. 5, 999-1013.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. U.S.A.* 89, 8794-8797.

Shrestha, P. M., Kube, M., Reinhardt, R. & Liesack, W. (2009). Transcriptional activity of paddy soil bacterial communities. *Environ. Microbiol.* 11: 960-970.

Szabo, A., Korszun, R., Hartl, F. U. & Flanagan, J. (1996). A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO J.* 15: 408-417.

Shu, Y., Zhou, J., Tang, W., Lu, K., Zhou, Q. & Zhang, G. (2009). Molecular characterization and expression pattern of Spodopteralitura (Lepidoptera, Noctuidae) vitellogenin, and its response to lead stress. *J. Insect Physiol.* 55, 608-616.

Silva, R. B. P. D., Campos, M. C. C., Silva, L. S., Filho, E. G. D. B., Lima, A. F. L. D., Pinheiro, E. N. & Cunha, J. M. (2020). Concentration of Heavy Metals in Soils under Cemetery Occupation in Amazonas, Brazil. *Soil Sediment Contam.* 29, 192-208.

Simon, C. & Daniel, R. (2010). Construction of small-insert and large insert metagenomic libraries. *Methods Mol. Biol.*668, 39-50.

Singh, D. P., Prabha, R., Gupta, V. K. & 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.

Singleton, D. R., Furlong, M. A., Rathbun, S. L. & Whitman, W. B. (2001). Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl. Environ. Microbiol.* 67, 4374-4376.

Sinha, S., Chattopadhyay, P., Pan, I., Chatterjee, S., Chanda, P., Bandyopadhyay, D., Das, K. & Sen, S.K. (2009). Microbial transformation of xenobiotics for environmental bioremediation. *Afr. J. Biotechnol.* 8, 6016–6027.

Siripornadulsil, S., Traina, S., Verma, D. P. S. & Sayre, R. T. (2002). Molecular mechanisms of proline-mediated tolerance to toxic heavy metal intransgenic microalgae. *Plant Cell.* 14, 2837-2847.

Sistani, K. R. & Novak, J. M. (2006). Trace metal accumulation, movement, and remediation in soils receiving animal manure. *Trace elements in the environment. Biogeochemistry, biotechnology, and bioremediation. Taylor & Francis, New York*, 689-706.

Sjöling, S. & Cowan, D. A. (2008). Metagenomics: microbial community genomes revealed. In *Psychrophiles: from biodiversity to biotechnology*. Springer, Berlin, Heidelberg. 313-332.

Smith, M. E., Henkel, T. W., Catherine Aime, M., Fremier, A. K. &Vilgalys, R. (2011). Ectomycorrhizal fungal diversity and community structure on three co-occurring leguminous canopy tree species in a Neotropical rainforest. *New Phytol.*192, 699-712.

Sobecky, P. A. & Coombs, J. M. (2009). Horizontal gene transfer in metal and radionuclide contaminated soils. *Methods Mol. Biol.* 532, 455-472.

Søfteland, L., Holen, E. & Olsvik, P.A. (2010).Toxicologicalapplicationofprimary hepatocyte cell cultures of Atlantic cod (*Gadusmorhua*) effects of BNF, PCDD and Cd. *Comp. Biochem. Physiol. C.* 151,401-411.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc. Natl. Acad. Sci. U. S. A.*103,12115–12120.

Solden, L., Lloyd, K. &Wrighton, K. (2016). The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr. Opin. Microbiol.* 31, 217-226.

Sorme, L. & Lagerkvist, R. (2002). Sources of heavy metals in urban wastewater in Stockholm. *Sci. Total Environ.* 298, 131-145

Spohner, S. C., Müller, H., Quitmann, H. & Czermak, P. (2015). Expression of enzymes for the usage in food and feed industry with *Pichiapastoris*. *J. Biotechnol.* 202, 118-134.

Sposito, G. & Page, A. L. (1984). Cycling of metal ions in the soil environment. *Metal ions in biological systems*, 18, 287-332.

Stacey, S. P., McLaughlin, M. J. & Hettiarachchi, G. M. (2010). Fertilizer-borne trace element contaminants in soils. *Trace Elem. Soil.* 135-154.

Staley, J. T. & Konopka, A. (1985). Measurement of *in situ* activities of non photosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.*39, 321-346.

Staley, J. T. & Konopka, A. (1985). Measurement of *in situ* activities of non photosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.*39, 321-346.

Steele, H. L., Jaeger, K. E., Daniel, R. & Streit, W. R., (2009). Advances in recovery of novel biocatalysts from metagenomes. *J. Mol. Microbiol. Biotechnol.*16, 25-37.

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* 178, 591-599.

Steinert, S. A. & Pickwell, G. V. (1993). Induction of Hsp70- proteins in mussels by ingestion of tributyltin. *Mar. Environ. Res.* 35, 89-93.

Stephan, U. W. & Scholz, G. (1993). Nicotianamine: mediator of transport of iron and heavy metals in the phloem. *Physiol. Plant.* 88, 522-529.

Stern, B. R. (2010). Essentiality and Toxicity in Copper Health Risk Assessment: Overview, Update and Regulatory Considerations. *J. Toxicol. Environ. Health. A.* 73, 114-127.

Stewart, E. J. (2012). Growing unculturable bacteria. *J. Bacteriol.* 194, 4151-4160.

Stewart, F. J., Ottesen, E. A. & DeLong, E. F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J.* 4, 896-907.

Stingl, U., Tripp, H. J. & Giovannoni, S. J. (2007). Improvements of high throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J.* 1, 361-371.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H-W. & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol*. 19, S21–S31.

Stofella, P. J. & Mateo Box, J. M. (2004). *Utilización de compost en los sistemas de cultivo hortícola* (No. 631.875 U8).

Sugahara, Y., Carninci, P., Itoh, M., Shibata, K., Konno, H., Endo, T., Muramatsu, M. & Hayashizaki, Y. (2001). Comparative evaluation of 5′-end-sequence quality of clones in CAP trapper and other full-length-cDNA libraries. *Gene*, 263. 93-102.

Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., Balzereit, D., Lehrach, H. & Yaspo, M. L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC genomics*, 15, 675.

Sutton, D., Tchounwou, P. B., Ninashvili, N. & Shen, E. (2002). Mercury induces cytotoxicity, and transcriptionally activates stress genes in human liver carcinoma cells. *Intl. J. Mol. Sci.* 3, 965-984.

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. & Sugano, S. (1997). Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene*, 200, 149-156.

Swarup, D. & Dwivedi, S. K. (2002). Environmental pollution and effects of lead and fluoride on animal health.

Szabolcs, I., 1994. The concept of soil resilience. In: Greenland, D., Szabolcs, I. (Eds.), Soil Resilience and Sustainable Land Use. CAB International, Wallingford, UK, 33-39.

Takasaki, K., Miura, T., Kanno, M., Tamaki, H., Hanada, S., Kamagata, Y. & Kimura, N. (2013). Discovery of glycoside hydrolase enzymes in an avicel-adapted forest soil fungal community by a metatranscriptomic approach. *PLoS One*, 8, e55485.

Talbot, J. M., Bruns, T. D., Taylor, J.W., Smith, D.P., Branco, S., Glassman, S.I., Erlandson, S., Vilgalys, R., Liao, H.L., Smith, M.E. & Peay, K.G., (2014). Endemism and functional convergence across the North American soil mycobiome. *Proc. Natl Acad. Sci. U.S.A.* 111, 6341-6346.

Tartar, A., Wheeler, M.M., Zhou, X., Coy, M.R., Boucias, D.G. & Scharf, M.E. (2009). Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol. Biofuels*. 2: 1-19.

Tchounwou, P. B., Centeno, J. A. & Patlolla, A. K. (2004). Arsenic toxicity, mutagenesis and carcinogenesis – a health risk assessment and management approach. *Mol. Cell. Biochem.* 255, 47-55.

Tchounwou, P. B., Ishaque, A. & Schneider, J. (2001). Cytotoxicity and transcriptional activation of stress genes in human liver carcinoma cells (HepG2) exposed to cadmium chloride. *Mol. Cell. Biochem.* 222, 21-28.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A. & Smith, M.E., (2014). Global diversity and geography of soil fungi. *Science*, 346, 1256688.

Teng, W., Kuang, J., Luo, Z. & Shu, W. (2017). Microbial diversity and community assembly across environmental gradients in acid mine drainage. *Minerals*. 7, 106.

Teste, M. A., Duquenne, M., François, J. M. & Parrou, J. L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomycescerevisiae*. *BMC Mol. Biol.* 10, 99.

Thavamani, P., Samkumar, R.A., Satheesh, V., Subashchandrabose, S.R., Kavitha, R., Naidu, R., Venkateswarlu, K. & Megharaj, M. (2017). Microbes from mined sites: harnessing their potential for reclamation of derelict mine sites. *Environ. Pollut.* 230, 495-505.

Thwaites, J. M., Farrell, R. L., Duncan, S. D., Lamar, R. T. & White, R. B. (2007). Fungal-based remediation: treatment of PCP contaminated soil in New Zealand. In *Environmental Bioremediation Technologies* Springer, Berlin, Heidelberg. 465-479.

Tirado, R., Englande, A. J., Promakasikorn, L. & Novotny, V. (2008). Use of agrochemicals in Thailand and its consequences for the environment. *Greenpeace Research Laboratories Technical. Bangkok, Thailand*.

Tkacz, A. (2013). *Plant genotype, immunity and soil composition control the rhizosphere microbiome* (Doctoral dissertation, University of East Anglia).

Todaka, N., Moriya, S., Saita, K., Hondo, T., Kiuchi, I., Takasu, H., Ohkuma, M., Piero, C., Hayashizaki, Y. & Kudo, T. (2007). Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *FEMS Microbiol. Ecol.* 59: 592-599.

Torsvik, V. & Ovreas, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol*. 5, 240-5.

Torsvik, V., Sørheim, R. & Goksøyr, J. (1996). Total bacterial diversity in soil and sediment communities-a review. *J. Indust. Microbiol.* 17, 170-178.

Treeby, M., Marschner, H. & Romheld, V. (1989). Mobilization of iron and other micronutrient cations from a calcareous soil by plant-borne, microbial, and synthetic metal chelators. *Plant Soil.* 114, 217-226.

Trench-Fiol, S. & Fink, P. (2020). Metatranscriptomics from a Small Aquatic System: Microeukaryotic Community Functions Through the Diurnal Cycle. *Front. Microbiol.* 11.

Tringe, S. G. & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* 6, 805-14.

Tringe, S. G., Von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C. & Bork, P. (2005). Comparative metagenomics of microbial communities. *Science*, 308: 554-557.

Tseng, T. S., Tzeng, S. S., Yeh, C. H., Chang, F. C., Chen, Y. M. & Lin, C. Y. (1993). The heat shock response in rice seedling-isolation and expression of cDNAs that encode class-I low molecular weight heat-shock proteins. *Plant Cell Physiol.* 34, 165-168.

Tufail, M., Naeemullah, M., Elmogy, M., Sharma, P.N., Takeda, M. & Nakamura, C. (2010). Molecular cloning, transcriptional regulation and differential expression profiling of vitellogenin in two wing morphs of the brown planthopper, *Nilaparvataligens*Stal (Hemiptera, Delphacidae). *Insect. Mol. Biol.* 19, 787-798.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R. & Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* 1, 6-14.

Turnbaugh, P.J., Quince, C., Faith, J.J., McHardy, A.C., Yatsunenko, T., Niazi, F., Affourtit, J., Egholm, M., Henrissat, B., Knight, R. & Gordon, J. I. (2010). Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7503-7508.

Tyson, G. W. & Banfield, J. F. (2005). Cultivating the uncultivated: a community genomics perspective. *Trends Microbiol.* 13, 411-415.

Underwood, E. J. (1971). Trace elements in human and animal nutrition, New York. Academic Press 461-477.

Uprety, D., Hejcman, M., Száková, J., Kunzová, E. & Tlustoš, P. (2009). Concentration of trace elements in arable soil after long-term application of organic and inorganic fertilizers. *Nutr. Cycling Agroecosyst.* 85, 241-252.

Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C. & Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PloS one*, 3.

USEPA (1986): Report to Congress on the Discharge of Hazardous Wastes to Publicly Owned Treatment Works (The Domestic Sewage Survey) (EPA/530-SW-86-004). Office for Water Regulation and Standards, Washington, DC

Valko, M. M., Morris, H., & Cronin, M. T. D. (2005). Metals, toxicity and oxidative stress. *Curr. Med. Chem.* 12, 1161-1208.

Van Assche, F. & Clijsters, H. (1990). Effects of metals on enzyme activity in plants. *Plant Cell Environ.* 13, 195-206.

Van Der Heijden, M. G., Bardgett, R. D. & Van Straalen, N. M. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* 11, 296-310.

Van Vliet, C., Anderson, C. R. & Cobbett, C. S. (1995). Copper-sensitive mutant of *Arabidopsisthaliana*. *PlantPhysiol*. 109, 871-878.

Vanwolterghem, I., Jensen, P. D, Ho, D. P., Batstone, D. & Tyson, G.W. (2014). Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr. Opin. Biotechnol.*27, 55–64.

Vatamaniuk, O. K., Mari, S., Lu, Y. P. & Rea, P. A. (1999). At PCS1, a phytochelatin synthase from *Arabidopsis*: isolation and *in vitro* reconstitution. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7110-7115.

Vaulot, D., Eikrem, W., Viprey, M. & Moreau, H. (2008). The diversity of small eukaryotic phytoplankton in marine ecosystems. *FEMS. Microbiol. Rev.* 32,795-820.

Verkleji, J. A. S. (1993). The effects of heavy metals stress on higher plants and their use as bio monitors. 1993. In Heavy metals, occurrence and toxicity for plants: A review. *Environ. Chem. Lett*. 8. 199-216.

Vierling, E. (1991). The roles of heat shock proteins in plants. *Ann. Rev. Plant Biol. 42*, 579-620.

Vogel, C., Radtke, M., Reinholz, U., Schäfers, F. & Adam, C. (2015). Chemical state of chromium, sulfur, and iron in sewage sludge ash based phosphorus fertilizers. *ACS Sustain. Chem. Eng.* 3, 2376-2380.

Vogiatzakis, I. N., Griffiths, G. H. & Mannion, A. M. (2003). Environmental factors and vegetation composition, Lefka Ori massif, Crete, S. Aegean. *Glob. Ecol. Biogeogr*, 12, 131-146.

Volossiouk, T., Robb, E. J. & Nazar, R. N. (1995). Direct DNA extraction for PCR mediated assays of soil organisms. *Appl. Environ. Microbiol.* 61, 3972-3976.

Vorobev, A., Dupouy, M., Carradec, Q., Delmont, T. O., Annamalé, A., Wincker, P. (2020). Transcriptome reconstruction and functional analysis of eukaryotic marine plankton communities via high-throughput metagenomics and metatranscriptomics. *Genome* Res. 30, 647-659.

Voroney, R. P. (2007). The soil habitat. In Soil microbiology, ecology and biochemistry. Academic Press. 25-49.

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.

Wallace, R. A. (1985). Vitellogenesis and oocyte growth in non-mammalian vertebrates. *Dev. Biol.* 1, 127-177.

Walters, K. J., Kleijnen, M. F., Goh, A. M., Wagner, G. & Howley, P. M. (2002). Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. *Biochemist.* 41, 1767-1777.

Wang, S. & Shi, X. (2001). Molecular mechanisms of metal toxicity and carcinogenesis. *Mol. Cell. Biochem.* 222, 3-9.

Wang, Y., Miao, X., Sun, J. & Cai, L. (2016). Oxidative stress in diabetes: Molecular basis for diet supplementation. In *Molecular Nutrition and Diabetes.* Academic Press. 65-72.

Wang, Y., Shimodaira, J., Miyasaka, T., Morimoto, S., Oomori, T., Ogawa, N., Fukuda, M. & Fujii, T. (2008). Detection of alpha gene expression of *Rhodococcus* sp. Strain RHA1 in soil using a new method of RNA preparation from soil. *Biosci. Biotechnol. Biochem.*72, 694-701.

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57-63.

Ward, D. M., Weller, R. & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature.* 345,63-65.

Warnecke, F. & Hess, M. (2009). A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J. Biotechnol.* 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. & Sorek, R. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*. 450, 560-565.

Welch, R. M., Norvell, W. A., Schaefer, S. C., Shaff, J. E. & Kochian, L. V. (1993). Induction of iron (III) and copper (II) reduction in pea (*Pisum sativum* L.) roots by Fe and Cu status: does the root-cell plasmalemma Fe(III)-chelate reductase perform a general role in regulating cation uptake? *Planta* 190, 555-561.

Wellenreuther, R., Schupp, I., Poustka, A., Wiemann, S. & German cDNA Consortium. (2004). SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. *BMC Genomics.* 5, 36.

Werner, I. & Nagel, R. (1997). Stress proteins HSP60 and HSP70 in three species of amphipodes exposed to cadmium, diazinon, dieldrin and flouranthene. *Environ. Toxicol. Chem*. 16, 2393-2403.

Whitmann, W. B., Coleman, D. C. & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A*, 95, 6578-6583.

WHO/FAO/IAEA. World Health Organization. Switzerland: Geneva. (1996). Trace Elements in Human Nutrition and Health.

Wiemann, S., Mehrle, A., Bechtel, S., Wellenreuther, R., Pepperkok, R. & Poustka, A. (2003). cDNAs for functional genomics and proteomics: the German Consortium. *C R. Biol*. 326, 1003-1009.

Williams, S. M. & Weil, R. R. (2004). Crop cover root channels may alleviate soil compaction effects on soybean crop. *Soil Sci. Soc. Am. J.* 68, 1403-1409.

Wollgiehn, R. & Neumann, D. (1999). Metal stress response and tolerance of cultured cells from Silene vulgaris and Lycopersicon peruvianum: role of heat stress proteins. *J. Plant Physiol.* 154, 547-553.

Wysocki, R. &Tamás, M. J. (2010). How *Saccharomyces cerevisiae* copes with toxic metals and metalloids. *FEMS Microbiol. Rev.* 34, 925–951.

Xalxo, R., Chandrakar, V., Kumar, M. & Keshavkant, S. (2020). Ecophysiological Responses of Plants Under Metal/Metalloid Toxicity. In *Plant Ecophysiology and Adaptation under Climate Change: Mechanisms and Perspectives I,* Springer, Singapore. 393-428.

Xiang, C. & Oliver, D. J. (1998). Glutathione metabolic genes coordinately respond to heavy metals and jasmonic acid in Arabidopsis. *Plant Cell.* 10, 1539-1550

Yadav, P., Maharjan, J., Korpole, S., Prasad, G. S., Sahni, G., Bhattarai, T. & Sreerama, L. (2018). Production, purification, and characterization of thermostable alkaline xylanase from *Anoxybacillus kamchatkensis* NASTPD13. *Front. Bioeng. Biotech.* 6, 65.

Yadav, R. K., Bragalini, C., Fraissinet-Tachet, L., Marmeisse, R. & Luis, P. (2016). Metatranscriptomics of soil eukaryotic communities. *Meth. Mol. Biol.* 1399, 273-287.

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.

Yang, X. & Romheld, V. (1999). Physiological and genetic aspects of micronutrient uptake by higher plants. In *Plant Nutrition—Molecular Biology and Genetics,* Springer, Dordrecht. 151-186.

Yang, Z. & Chu, C. (2011). Towards understanding plant response to heavy metal stress. *Abiotic stress in plants–Mechanisms and adaptations,* 10, 24204.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., & Heath, A. C. (2012). Human gut microbiome viewed across age and geography. *Nature.*486, 222-227.

Yedjou, G. C. & Tchounwou, P. B. (2008). N-acetyl-cysteine affords protection against lead-induced cytotoxicity and oxidative stress in human liver carcinoma (HepG2) cells. *Intl. J. Environ. Res. Public. Health.* 4, 132-137.

Yi, H., Cho, Y. J., Won, S., Lee, J. E., Jin Yu, H., Kim, S., Schroth, G. P., Luo, S., & Chun, J. (2011). Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Res*. 39.

Yilmaz, A., Allgaier, M. & Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods.* 7, 943-944.

Young, I. & Crawford, J. (2004). Interactions and self-organisation in the soil-microbe complex. Science. 304, 1634-1637.

Young, I. M. & Ritz, K. (2000). Tillage, habitat space and function of soil microbes. *Soil Till. Res.* 53, 201-213.

Yu, K. & Zhang, T. (2012). Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. *PloS one*. 7, e38183.

Yu, X. Z., Wu, S. C., Wu, F. & Wong, M. H. (2011). Enhanced dissipation of PAHs from soil using mycorrhizal ryegrass and PAH-degrading bacteria. *J. Hazard. Mater.* 186, 1206-1217.

Yuvaraj, A., Karmegam, N., Tripathi, S., Kannan, S. & Thangaraj, R. (2020). Environment-friendly management of textile mill wastewater sludge using epigeic earthworms: Bioaccumulation of heavy metals and metallothionein production. *J. Environ. Manage.* 254, 109813.

Zhao, F. J., Hamon, R. E. & McLaughlin, M. J. (2001). Root exudates of the hyperaccumulator *Thlaspi caerulescens* do not enhance metal mobilization. *New Phytol.* 151, 613-620.

Zhang, Y., Zhang, X., Zhang, H., He, Q., Zhou, Q., Su, Z. & 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.

Zhen, Y., Qi, J. L., Wang, S. S., Su, J., Xu, G. H., Zhang, M. S., Miao, L. V., Peng, X. X., Tian, D. & Yang, Y. H.(2007). Comparative proteome analysis of differentially expressed proteins induced by Al toxicity in soybean. *Physiol*. *Plant* 131, 542-554.

Zhen, Y., Qi, J.L., Wang, S.S., Su, J., Xu, G.H., Zhang, M.S., Miao, L.V., Peng, X.X., Tian, D. & Yang, Y.H.(2007). Comparative proteome analysis of differentially expressed proteins induced by Al toxicity in soybean. *Physiol*. *Plant* 131, 542-554.

Zhenyan, H. E., Jiangchuan, L. I., Zhang, H. & Ma, M. I. (2005). Different effects of calcium and lanthanum on the expression of phytochelatin synthase gene and cadmium absorption in *Lactuca sativa*. *Plant Sci.* 168,309-318.

Zhou, J. Z., Bruns, M. A. & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62, 316–322.

Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S. G. & Alvarez-Cohen, L. (2015). High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *MBio*.6, e02288–e02214.

Zhu, Y. L., Pilon-Smits, E. A. H, Tarun, A. S., Weber, S. U, Jouanin, L. & Terry, N. (1999). Cadmium tolerance and accumulation in, Indian mustard is enhanced by overexpressing gammaglutamylcysteine synthetase. *Plant*. *Physiol*. 121, 1169-1177.

Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & 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. & 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.

APPENDIX I

COMPOSITION OF MEDIA USED

YEPD Agar (pH 7.1) (1 L)

Peptone	20.0 g
Yeast extract	10.0 g
Dextrose	20.0 g
Agar	20.0 g
Distilled water	1000 ml

Allow medium to cool to55 ºC and then add dextrose (glucose) to 2% (50 ml of a sterile 40% stock solution).

Synthetic Defined media (pH 5.7) (1 L)

Yeast nitrogen base (w/o aa)	6.7 g
Agar	20.0 g
Distilled Water	1000 ml

Add 0.76 g of DO supplement or in case of 10X DO stock, add 990 ml of H_2O and 10 ml of DO supplement.

SOC media (pH 7.2)

Peptone	10.0g
Lactose	5.0g
Dipotassium hydrogen phosphate	2.0g

Luria Broth (pH 7.0) (1 L)

Bacto-tryptone	10.0 g
Bacto-yeast extract	5.0 g
NaCl	5.0 g
D.W.	1000 ml

*To prepare LB+amp plates, add agar (20 g/L) to the media and autoclave it. Cool down the autoclaved media (50 ºC) to add ampicillin (50 µg/ml) and pour the media into plates.

** To prepare LB plates with amp/IPTG /X-Gal, spread 100 µl of IPTG (100mM) and 20 µl of 50 mg/ml X-Gal on LB + amp plates prepared as above. Incubate these plates at 37°C for 30 mins prior to use.

IPTG stock solution (0.1M)

IPTG	1.2 g

Add water to 50 ml final volume. Filter sterilize and store at 4°C.

X-Gal (2ml)

Dissolve 100 mg of 5-bromo-4-chloro-3-indolyl-D-galactoside in 2 ml N,N´-dimethylformamide. Cover with aluminum foil and store at 20°C.

10X Drop out (DO) Solution

L-Adenine hemisulfate salt	200 mg/L
L-Arginine HCl	200 mg/L
L-Histidine HCl	200 mg/L
L-Isoleucine	300 mg/L
L-Leucine	1000 mg/L
L-Lysine HCl	300 mg/L
L-Methionine	200 mg/L
L-Phenylalanine	500 mg/L
L-Threonine	2000 mg/L
L-Tryptophan	200 mg/L
L-Tyrosine	300 mg/L
L-Uracil	200 mg/L
L-Valine	1500 mg/L

To make one liter of 10X –Ura Drop-out Solution, combine all amino acid except uracil.

APPENDIX II

1. COMPOSITION OF REAGENTS USED

Denaturing solution (pH 8)

Guanidine isothiocyanate	4 M
Tris-HCl	10 mM
Na_2EDTA	1 mM

Lytic solution (pH 9)

Tris-HCl	100 mM
Na_2EDTA	20 mM
NaCl	100 mM
SDS	2 **%**

PEG/LiAc solution (polyethylene glycol/lithium acetate)

PEG 4000	40%
TE buffer	1X
LiAc	1X

Use freshly prepared.

TBE Buffer (10x)

Tris-HCl (pH 8)	0.09 M
Boric acid	0.9 M
EDTA (pH 8)	0.02 M

Agarose gel loading dye (6X)

Bromophenol blue	0.25%
Xylene cyanol FF	0.25%
Glycerol in water	30.0%

MOPS (5X)

3-morpholinopropane sulphonic acid	20mM
Sodium acetate	5 mM
Na_2EDTA (pH 7.0)	5 mM

TE buffer 10X

Tris-HCl (pH 8)	0.1 M
Na2EDTA (pH 8)	10 M

Potassium Dichromate Normal Solution (1 N)

Potassium dichromate	49.035 g

Dissolved in distilled water and make the final volume upto one liter.

Ferrous Sulphate solution (0.5 N)

$FeSO_4$	140 g

Dissolved in 0.5 N H_2SO_4 to make 1 liter of solution.

H_2SO_4 solution (0.5 N)

For 0.5 N, 14 ml of concentrated H_2SO_4 was added to water to make 1 L of solution.

Indicator solution for total carbon measurement

Sodium diphenylamine sulphonate	0.25 g
Distilled water	100 ml

Sodium hydroxide solution

NaOH	450g
Distilled water	1000 ml

Alkaline Sodium Sulphide Solution

Sodium sulphide	20 g
H_2O	50 ml
NaOH solution	600ml.

0.5N HCl

For 0.5N HCl, added 41 ml of concentrated HCl in one litre H_2O.

5N H_2SO_4

For 5N H_2SO_4 added 139 ml of concentrated H_2SO_4 in one litre H_2O.

Solution A

Added 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 1000 ml of 5N H_2SO_4. The volume was made up to 2 litres with distilled water and stored in a dark pyrex bottle.

Solution B

Added 1.056 g of ascorbic acid was dissolved in 200 ml of solution A and mixed thoroughly.

Used freshly prepared.

Ammonium Fluoride

NH_4F	37 g
H_2O	1000 ml

Stored in polyethylene bottle.

Extracting Solution

NH_4F (1N)	15 ml
0.5 N HCl	25 ml
H_2O	450 ml

Dickman and Bray's Reagent

Dissolved 15 g of ammonium molybdate ($(NH_4)_6Mo_7O_{24}.4H_2O$) in 300 ml distilled water. It was warmed to 60°C filtered and cooled to room temperature. To that added 34.3 ml of concentrated HCl and final volume made up to 1 litre.

Stannous Chloride Solution

Dissolved 2 g of $SnCl_2$. 2 H_2O crystals was in 8.3 ml Conc. HCl and diluted to 100 ml.

Preparation of Standard Phosphorous Solution

Potassium Dihydrogen Phosphate	0.439 g
Conc. H_2SO_4	25 ml

To make a 2ppm stock, 2 ml from the mixture was added to a beaker and volume was made up to 10 ml

APPENDIX III

Uncultured eukaryote clone PLBe1 ubiquitin fusion protein mRNA, partial cds

```
LOCUS       MH244348         693 bp    mRNA    linear   ENV 22-MAY-2018
DEFINITION  Uncultured eukaryote clone PLBe1 ubiquitin fusion protein
            mRNA, partial cds.
ACCESSION   MH244348
VERSION     MH244348.1
KEYWORDS    ENV.
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Thakur,B., Yadav,R., Fraissinet-Tachet,L., Marmeisee,R. and
            Reddy,S.M.
  TITLE     Isolation and characterization of multi-metal tolerant
            ubiquitin fusion protein from metal contaminated soil by
            metatranscriptomic
            approach
  JOURNAL   Unpublished
REFERENCE   2(bases 1 to 693)
  AUTHORS   Thakur,B., Yadav,R., Fraissinet-Tachet,L., Marmeisee,R. and
            Reddy,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2018) Department of Biotechnology, Thapar
            University, Bhadson Road, Patiala, Punjab 147004, India
COMMENT     ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..693
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="soil sample"
                     /db_xref="taxon:100272"
                     /clone="PLBe1"
                     /environmental_sample
     CDS             <1..590
                     /note="ubiquitin moiety fused with AN-1-like zinc
finger
                     protein"
                     /codon_start=3
                     /product="ubiquitin fusion protein"
                     /protein_id="AWK68379.1"

/translation="REDKEGIPPDQQRLIFAGKQLEDGRTLADYNIQKESTLHLVLRL

RGGMQIFVKTLTGKTITLEVESSDTIENVKQKIQDKEGIPPYQQRLIFAGKQFEDGRT

LADYNIQKESTLHLVLRLRGGGKKTKCCFGECSRRVVLITGDCRFCQSQFCGSHRLPE
```

```
DHACQSMDSCRQAAFDRNKNKLMNEKCVSQKVAST"

ORIGIN
1   agcgggagga caaggagggc atccccccgg accagcagcg tctcatcttc gccggcaagc
61  agctcgagga cggccgtacc ctcgcggact acaacatcca gaaggagtcg actctccacc
121 tcgtcctccg tctgcgtggc ggtatgcaga tcttcgtcaa gaccctgacc ggcaagacca
181 ttaccctcga ggtcgagtct agcgatacca tcgagaacgt gaagcagaag atccaggaca
241 aggagggcat ccccccgtat cagcagcgtc tcatcttcgc tggtaagcag ttcgaggacg
301 gtcgcacgct cgccgactac aacatccaga aggagtcgac cctccacctt gtcctccgcc
361 ttcgtggagg cggcaagaag accaagtgct gcttcggcga gtgctccagg cgcgtggtcc
421 tcatcactgg cgattgccgc ttctgccagt ctcagttctg cggctcgcac cgcctccccg
481 aggatcacgc ctgccagagc atggactcat gccgacaggc ggccttcgat cggaacaaga
541 acaagctcat gaatgagaag tgcgtcagcc agaaggtcgc ctctacctaa agctgcgaga
601 tcctcctcct ccctgtacaa atgactccat ttgtaaacac cccttcgtaa atacagtaca
661 tatccaccaa aaaaaaaaaa aaaaaaaaaa aaa
//
```

Uncultured eukaryote clone PLCe10 von Willebrand factor type D domain protein (VWD) mRNA, complete cds

```
LOCUS       MF770726     1234 bp    mRNA    linear   ENV 14-APR-2018
DEFINITION  Uncultured eukaryote clone PLCe10 von Willebrand factor
            type D domain protein (VWD) mRNA, complete cds.
ACCESSION   MF770726
VERSION     MF770726.1
KEYWORDS    ENV.
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1  (bases 1 to 1234)
  AUTHORS   Thakur,B., Yadav,R., Vallon,L., Frassinet-Tachet,L.,
            Marmeisse,R. and Reddy,S.M.
  TITLE     Characterization and multi-metal resistance of Von
            Willebrand factor type D domain from eukaryotic
            transcript of metal-contaminated site
  JOURNAL   Unpublished
REFERENCE   2(bases 1 to 1234)
  AUTHORS   Thakur,B., Yadav,R., Vallon,L., Frassinet-Tachet,L.,
            Marmeisse,R. and Reddy,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2017) Department of Biotechnology,
            Thapar University, Bhadson Road, Patiala, Punjab 147004,
            India
FEATURES             Location/Qualifiers
     source          1..1234
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="soil"
                     /db_xref="taxon:100272"
                     /clone="PLCe10"
                     /environmental_sample
                     /country="France: northwest of Paris,
Pierrelaye,
                     agro-forestry site"
     gene            1..1234
                     /gene="VWD"
     CDS             12..1028
                     /gene="VWD"
                     /note="involved in metal tolerance"
                     /codon_start=1
                     /product="von Willebrand factor type D domain
protein"
                     /protein_id="AVW85699.1"

/translation="MWEQLREYVQDDEDQAQCSITSKSRRQGRSQVETFDGTKFSAPF

TNCWVVLAKDCGSQQPKFVVMARKSERGQDLKEVKIVTKKHRIQLTPDSAEYNSVKVQ

VNGQQYDPESDQQITENGQVVQIEKDQSTVQVQLPNTGLEVEFDGYAINIQLSQAYRG
```

```
QQCGLCGHFDLESTDEFRNPDFTDEQDIRQFYMNYLIKDGQCKAPQQLTEVCESEECD

KADRSSSSSSSSSSSGSQESQSNEESEIPEKKTKVIEIDDQLCFSTVPIPQCDEEDSY

PMGEKEKRKVAYVCIDQDSQEAEDIERQARSGRRQIPALNNRTPSFTRTEKIPEKCKK
YSRN"
ORIGIN
1   gcgggcagcc catgtgggaa caattgcgcg aatacgtgca agacgacgaa gatcaggccc
61  agtgctccat cactagcaag tcaaggcgac aaggtcgctc tcaagttgag accttcgacg
121 gtaccaaatt cagcgctcca ttcaccaact gctgggtcgt tttggccaaa gattgtggct
181 cgcagcaacc gaaattcgtc gtcatggctc gcaaatcgga acgcggccag gatttgaagg
241 aagtgaaaat cgttaccaag aagcatcgca tccaattgac acccgacagc gccgagtaca
301 attcggtcaa ggtccaagtc aacggccaac agtacgatcc agaaagcgac caacaaatca
361 ccgaaaatgg acaagtcgtt caaatcgaaa aggaccaatc gaccgtccaa gttcaattgc
421 cgaatactgg cctcgaagtc gaattcgacg gctacgctat caacattcaa ttatcgcaag
481 cctaccgtgg tcagcaatgc ggtctttgcg gtcacttcga cttggaatcg actgacgaat
541 tccgcaaccc cgacttcacc gacgaacagg acattcgtca gttctacatg aactacctga
601 tcaaggacgg ccaatgcaag gctccgcaac aattgaccga agtctgcgaa agcgaagagt
661 gcgacaaggc tgaccgtagc agcagcagca gcagcagcag cagcagcagc ggaagccaag
721 agagtcaaag caatgaggaa tcggaaattc ctgaaaaaaa gaccaaggtt atcgaaattg
781 acgatcaact ctgcttctcc accgttccca taccccagtg cgacgaggaa gattcctacc
841 cgatgggaga gaaagagaaa aggaaggtgg cctacgtctg catcgatcaa gacagccaag
901 aagccgagga tatcgaacgc caggctcgtt ccggccgccg ccagatcccg gccttgaata
961 atcgtactcc gtcgttcact cgcacggaga agataccaga gaaatgcaag aagtacagca
1021 ggaactaaag tgttcctcaa ttcgactcac tcccgaagtg agagatgtga tttttgtatc
1081 gtagcgtctc cctttgcgca ctcgatacca tatgatatta tccatattca ttattatgat
1141 tcattttata cgcagtgagt gccagttttt atcagtgtca ttgggatgat gaaaataaat
1201 caactgcatg caaaaaaaaa aaaaaaaaac aaag
//
```

Uncultured eukaryote clone PLCc43 DnaJ-like protein (dnaJ) mRNA, complete cds

```
LOCUS       MK079355    1344 bp    mRNA    linear   ENV 05-FEB-2019
DEFINITION  Uncultured eukaryote clone PLCc43 DnaJ-like protein
            (dnaJ) mRNA,complete cds.
ACCESSION   MK079355
VERSION     MK079355.1
KEYWORDS    ENV.
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1(bases 1 to 1344)
  AUTHORS   Thakur,B., Yadav,R., Vallon,L., Marmeisse,R.,
            Fraissinet-Tachet,L. and Reddy,S.M.
  TITLE     Molecular and functional characterization of eukaryotic
            metatranscript as DnaJ like protein screened from metal
            contaminated site
  JOURNAL   Unpublished
REFERENCE   2(bases 1 to 1344)
  AUTHORS   Thakur,B., Yadav,R., Vallon,L., Marmeisse,R.,Fraissinet-
           Tachet,L. and Reddy,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-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..1344
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="soil"
                     /db_xref="taxon:100272"
                     /clone="PLCc43"
                     /environmental_sample
                     /country="France: North west of Paris,
Pierrelaye"
     gene            1..1344
                     /gene="dnaJ"
     CDS             17..871
                     /gene="dnaJ"
                     /note="Involved in metal tolerance"
                     /codon_start=1
                     /product="DnaJ-like protein"
                     /protein_id="QAX33439.1"

/translation="MAKEFHPDRNPAAGDKFKEISFAYEVLSDPEKREVYDRYGEQRL

KEGPSAGGGGGFDSAEDLFASLFGGGFFGGMSGGGGGGRGRQRKGEDLVHSLEVTMED
```

```
LYNGKVSKLALRKQVICGVCQGKGGRNPDSVKPCDTCRGTGFKVTLRQLGPGMVQQLQ

TVCPKCRGQGEVIREKDRCKRCRGSKVVQEKKVLDIYIDKGMRHKQKIVFTGEADQEP

GVIPGDVIIILKELSHPRFKRNDANLLIEQEITLVEALCGFSFALLHLDSRTLLIXSX
PRRSDSAR"
ORIGIN
1   gcgggggtac tacaaaatgg ccaaagaatt ccatcccgac cgcaaccccg ctgctggcga
61  caagttcaag gagatcagct tcgcctacga ggtgctgtcg gatccagaga agcgggaggt
121 gtacgaccgc tatggtgagc agcgtctcaa ggaaggcccc tcagctggag gaggaggagg
181 cttcgattcg gcggaggacc tcttcgcctc ccttttcggg ggaggcttct ttggcggaat
241 gagcggcggt ggaggcggag gacgcggaag gcaacgcaag ggcgaggacc tcgtccattc
301 cctcgaggtg acgatggagg acctctacaa cggcaaggtg tcaaagcttg ccctccgcaa
361 acaggtcatt tgcggtgttt gccaagggaa aggcgggagg aaccccgact cggtaaagcc
421 ttgcgacacg tgcagaggca caggattcaa ggttactctt cgccagctgg gtcccggaat
481 ggtacagcag ttgcagaccg tatgccccaa atgtcgtgga caaggcgagg tcatacgtga
541 gaaggaccgc tgcaagcgat gtagaggaag taaagttgtg caagaaaaga aagttctgga
601 catatacatc gacaaggga tgcgacacaa acaaaaaatc gtgttcaccg gagaagcaga
661 ccaggagcct ggtgtaattc ccggcgacgt gatcatcatt ctcaaagagc tgtcacatcc
721 acggttcaag cgcaacgacg caaatctgct catcgaacag gagataacgc tggttgaagc
781 gctatgcggc ttttcgtttg ctctcctgca cctcgacagc cgcacgctct tgatcanatc
841 anatccccgg agaagtgatt cggccaggtg acatcaagga agtgcccggc gaagggacgc
901 ccatccacaa gcggccattc gacaaggcct cttggttatc aaattcaaca tcaaattccc
961 cgactccatc aaccaagaac aggctactag gcttcttgcg acgttacccg gtcccgagcc
1021 actgcgactg tcaatggagg gagtggagga agtgttcttg cgcgattacg gcgaggaaga
1081 gacctcaacg caccagcacc agcgccaggc ctacgaagag gacgagcgca ctggatcgaa
1141 actcgattgc tcccagcaat gatgcaaata tagactttt tattttcatt ttttttcgg
1201 catcacgtcg ttatcgtaca agaccactta cactagaatt acatgaaaaa aaaaaaaana
1261 aaaaacaaag ggccnccccg gcnagcggcc gcaatnattt tttttctttt ctctttccca
1321 cccttaacaa aagggccatc acag
//
```

Uncultured eukaryote clone PLCd43 gamma glutamylcysteinyl transferase-like protein mRNA, complete cds

```
LOCUS       MK248844     1393 bp    mRNA    linear   ENV 24-JUN-2019
DEFINITION  Uncultured eukaryote clone PLCd43 gamma
           glutamylcysteinyl transferase-like protein mRNA, complete
           cds.
ACCESSION   MK248844
VERSION     MK248844.1
KEYWORDS    ENV.
SOURCE      uncultured eukaryote
  ORGANISM  uncultured eukaryote
            Eukaryota; environmental samples.
REFERENCE   1  (bases 1 to 1393)
  AUTHORS   Thakur,B., Yadav,R., Vallon,L., Marmeisse,R.,
            Fraissinet-Tachet,L. and Reddy,S.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-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..1393
                     /organism="uncultured eukaryote"
                     /mol_type="mRNA"
                     /isolation_source="Agro-Forestry Soil"
                     /db_xref="taxon:100272"
                     /clone="PLCd43"
                     /environmental_sample
                     /country="France: Pierrelaye, Northwest of
Paris"
     CDS             12..1367
                     /codon_start=1
                     /product="gamma glutamylcysteinyl transferase-
                     like protein"
                     /protein_id="QDD67520.1"

/translation="MKRVSFGAGPMTPRPNYRSFSSCLMSRERHFVVASTPGVCLERI

RWSPFSNWRRDYATQVEKEIVPTEESFHGRPLPPHLSSFSSPEGRMIFKEMVAGKYME

SYWTLAEQFQTQSEPAYCGLAALSMVLNALKIDPKRLWKAPWRWFSEDLLDCCAPLEI

VKKKGITFGEFACNSVGIQLIIEGLAKCNGANIESYPVSESSEDQFREHIKKASSTEG

VHLVVSFSRQMLGQTGDGHFSPIAGYHAEXDLVLVLDTAKFKYPAFWCPVHMLFEAMK

PADQVTSRPRGYFLLTKNFHCNTAICCKVNRDESSLAITFHTLSQQLPPLLKQSSTTT

VSDIVKIVFGNFTSQVNSLLELDGEKGSRPNVEKTELFGIVEETLRGDKNGSSLDAEK
```

```
ATFFLLASPSQLYKELSTELKAEFFQLRNTEAMPKELRNEVTQMRYDMLNLTQCETVY
                       K"
ORIGIN
1   gcgggggaga aatgaagagg gtttcctttg gtgcaggtcc catgacccca cggccaaact
61  atcgttcctt ctcttcgtgt ttgatgagca gagagcgaca ttttgtggtg gcctctactc
121 ctggagtttg tcttgagagg attcggtgga gtcctttttc caactggagg agggattatg
181 caacccaagt ggagaaggaa atagtcccca cagaagagag cttccatgga cgtccattgc
241 ctccccatct cagctccttc tcatctccag aaggacgaat gatattcaaa gagatggtcg
301 cagggaaata catggaatca tattggaccc ttgcagaaca gttccagaca cagagtgaac
361 cagcatattg tggactggca gcgctctcca tggttctaaa tgcgctcaag atagatccaa
421 agaggttgtg gaaggctcca tggaggtggt tcagtgaaga tctcttggac tgctgtgccc
481 ctttggagat cgtgaagaag aaggggatca cttttggtga atttgcttgt aactctgttg
541 gaatccaact catcattgaa ggtctggcaa aatgcaatgg cgcaaacata gagtcttatc
601 ctgtgtctga atcttcagag gatcagtttc gtgagcacat taagaaggca tcttccactg
661 aaggggtcca tcttgtggtc agtttttcta ggcagatgct tggccagacg ggggatggac
721 acttctctcc aattgctgga tatcacgcag aanaggacct cgtgcttgtc ttggatacag
781 ccaagtttaa gtacccagcc ttctggtgtc ccgttcacat gctgttcgag gccatgaaac
841 cagctgatca ggtgacaagc agacctcgtg ggtacttctt gctgacaaag aacttccact
901 gcaataccgc catctgctgc aaggtcaaca gagacgaatc ctcgctggcg attactttcc
961 acacactctc acagcagctc ccacctctat tgaagcagtc aagcaccacc accgtctcag
1021 acatcgtgaa gattgtgttt ggaaacttca cttcacaagt gaattccctt ttggaattgg
1081 atggagagaa aggatcgcgt ccaaatgtgg aaaagacaga actctttggg atcgtggagg
1141 aaactttgcg aggggacaag aatggaagct ctttggatgc agagaaggcc accttcttct
1201 tgcttgcatc tccctctcag ctctacaagg aactgtcaac ggaattgaag gcggaattct
1261 ttcagctgcg gaacacagag gccatgccaa aggagctgag gaatgaggtg acacagatga
1321 gatacgacat gctgaacctc acacagtgtg agacagtgta caaatgaaga aaacacaaaa
1381 aaaaaaaaaa aaa
//
```

PUBLICATION

Journal of Microbiological Methods 152 (2018) 119–125

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Isolation of multi-metal tolerant ubiquitin fusion protein from metal polluted soil by metatranscriptomic approach

Bharti Thakur[a], Rajiv Yadav[b], Laurence Fraissinet-Tachet[b], Roland Marmeisse[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:
Functional Metatranscriptomics
Heavy metals
Ubiquitin fusion protein
Functional complementation
AN1 type zinc finger protein

ABSTRACT

Release of heavy metals into the soil pose a significant threat to the environment and public health because of their toxicity accumulation in the food chain and persistence in nature. The potential of soil microbial diversity of cadmium (Cd) contaminated site was exploited through functional metatranscriptomics by construction of cDNA libraries A (0.1–0.5 kb), B (0.5–1.0 kb), and C (1-4 kb) of variable size, from the eukaryotic mRNA. The cDNA library B was further screened for cadmium tolerant transcripts through yeast complementation system. We are reporting one of the transformants $ycf1^{\Delta}$PLBe1 capable of tolerating high concentrations of Cd (40 μM - 80 μM). Sequence analysis revealed that PLBe1 cDNA showed homology with ubiquitin domain containing protein fused with AN1 type zinc finger protein of *Acanthameoba castellani*. Further, this cDNA was tested for its tolerance towards other heavy metals such as copper (Cu), zinc (Zn) and cobalt (Co). Functional complementation assay of cDNA PLBe1 showed a range of tolerance towards copper (150 μM - 300 μM), zinc (10 mM - 12 mM) and cobalt (2 mM - 4 mM). This study promulgates PLBe1 as credible member of multi-metal tolerant gene in the eukaryotic soil microbial community and can be used as potential member to revitalise the heavy metal contaminated sites or can be used as a biomarker to detect heavy metal contamination in the soil environment.

1. Introduction

Heavy metal pollution in soil environment is an outrageous hurdle, owing to expeditious social and economic expansions over the past years in several parts of the world (Facchinelli et al., 2001; Solgi et al., 2012). In spite of the fact that heavy metals exists in the soil naturally, there are additional and huge contributions from various anthropogenic activities like agriculture, urbanisation, industrialisation and mining (Wei and Yang, 2010). Heavy metals from soil enter into the food chain through contaminated water and dietary intake of plant-derived food and hence result into potentially health-threatening human exposure (Clemens and Ma, 2016). Soil is a rich reservoir of numerous microorganisms which are the primary agents of organic mineralization in soil, and may also convert heavy metal contaminants to stable and/or less toxic forms (Bell et al., 2014). Both uni- and multicellular members of soil ecosystem have been explored to discover a range of tolerance mechanisms at molecular and cellular level, predominantly involving some kind of detoxification processes like chelation, cellular compartmentalization or the excretion of the metals (Clemens, 2006; Blindauer and Leszczyszyn, 2010; Gadd, 2010). It has been reported that the degradation of aberrant proteins or the proteolytic inactivation of regulatory proteins are also major cellular response to heavy metals (Jungmann et al., 1993; Hwang et al., 2007; Medicherla and Goldberg, 2008). Therefore microbial communities of derelict sites such as heavy metal mines were explored to revitalise these sites (Thavamani et al., 2017; Teng et al., 2017).

The major impediment in the study of eukaryotic microbial communities is that majority of the microbes are uncultivable hence functional and ecological roles of individual taxa is unexplored (Prosser, 2007; Fierer et al., 2007). To recognize the eukaryotic functional diversity and *in situ* activities they possess in response to different environment constraints, mRNA pool of soil is explored through metatranscriptomic approach (Bailly et al., 2007; Aguiar-Pulido et al., 2016; Marmeisse et al., 2017). Functional metatranscriptomics is appreciated as commanding tool in identifying various genes playing crucial role in numerous mechanisms and phenomenon to adapt the diverse environmental conditions (Damon et al., 2011; Kellner et al., 2010; de Menezes et al., 2012; Zielinski et al., 2016). Function - driven approaches of

---

* Corresponding author.
E-mail address: msreddy@thapar.edu (M. Sudhakara Reddy).

https://doi.org/10.1016/j.mimet.2018.08.001
Received 30 April 2018; Received in revised form 31 July 2018; Accepted 1 August 2018

metatranscriptomics unveils innumerous genes responsible for various ecologically significant mechanisms or processes such as oligopeptide transporters and metal tolerant genes and cysteine rich proteins (Damon et al., 2011; Lehembre et al., 2013; Ziller et al., 2017). This approach also provides the snapshot of gene expression profiling and diversity of various microbial communities like forest soils (Damon et al., 2011), activated sludge community from municipality wastewater (Yu and Zhang, 2012), phenanthrene degrading community in soil (de Menezes et al., 2012), muskoxen rumen (Qi et al., 2011), acid mine drainage community (Chen et al., 2015). Identification of full length environmental transcripts involves construction of cDNA libraries from polyadenylated RNA isolated from an environmental sample followed by screening of the library and sequencing of full length transcripts (Bailly et al., 2007; Damon et al., 2012). Screening of synthesized cDNA libraries, which is a consortium of protein coding genes, involves their heterologous expression through a suitable eukaryotic host like *Saccharomyces cerevisiae* (Kellner et al., 2010). Hence metatranscriptomic approach has led to discovery of novel genes and mechanisms with a potential to bring revolution in biotechnology and bioremediation.

In the present study, eukaryotic cDNA libraries constructed from the soil samples contaminated with Cd were screened by yeast complementation assay to identify the Cd tolerant genes. Screening of one of the libraries (library B) led to the identification of many genes which are involved in Cd tolerance including ubiquitin fusion protein (UFP). Ubiquitin fusion protein is an eukaryotic protein playing an important role in selective degradation of aberrant proteins translated due to heavy metal contamination (Pena et al., 2007; Forzani et al., 2002) or due to other stress conditions like osmostress, chemicals and heat shock (Leach et al., 2011; Peng et al., 2017; Adler et al., 2017). Present study aims to characterize the UFP and its multi-metal tolerance by using metal sensitive yeast mutants.

2. Materials and methods

2.1. Soil sample collection and analysis

Soil samples were collected from the Agro-forestry site in Pierrelaye (PL) (49°1′45”N, 2°10′32″E) which is located northwest of Paris, France. Earlier, the sampling site was used for maize cultivation and later turned into Agro-forestry land by embracing long trees of Poplar. Prolonged use of untreated wastewater to irrigate this site led to heavy metal contamination. Twenty soil cores were collected from each replicate and equal volumes of each core were mixed together to form composite samples after sieving them through 2.0 mm mesh. Soil sample was analyzed for physicochemical properties *i.e.* pH (1:2 soil: water ratio), organic carbon (Walkley, 1947); total P (Kitson and Mellon, 1944); available P (Olsen et al., 1954) and total nitrogen (Piper, 1966). The metal contents of the soil sample were determined by inductively coupled plasma atomic emission spectroscopy (ICP-AES) (Moor et al., 2001).

2.2. Biological materials

Cadmium -sensitive $ycf1^{\Delta}$, cobalt-sensitive $cot1^{\Delta}$, Zinc-sensitive $zrc1^{\Delta}$ and copper-sensitive $cup1^{\Delta}$ (DTY4) mutant strains of *Saccharomyces cerevisiae* were used in this study along with their respective wild type strains. All these yeast mutant strains and wild strain were procured from Euroscarf (http://www.euroscarf.de). $ycf1^{\Delta}$, $cot1^{\Delta}$ and $zrc1^{\Delta}$ strains were derived from the wild-type strain BY4741 (*MATα*, $his3^{\Delta}1$, *leu2Δ0*, $met15^{\Delta}0$, *ura3Δ0*) and the isogenic wild type strain DTY3 ($cup1^{s}$) (*MATα, leu2–3,* $112his3^{\Delta}1$, *trp1–1, ura3–50, gal1,* $CUP1^{S}$) were used as positive controls. Thus Cd sensitive $ycf1^{\Delta}$ lacks an ABC transporter gene YCF1 (Li et al., 1997) while deletion of the ZRC1, a gene encoding for transporter protein in $zrc1^{\Delta}$ (Miyabe et al., 2001) and deletion of COT1 gene in $cot1^{\Delta}$ (Conklin et al., 1992) has led to the hypersensitivity of these strains towards their respective metals. The copper-sensitive $cup1^{\Delta}$ (DTY4) is lacking the copperthionein gene (Longo et al., 1996). 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. Screening of cDNA library B for cadmium tolerance genes

Total RNA was isolated from soil sample by using the RNA PowerSoil® Total RNA Isolation Kit (Mo Bio laboratories, Carlsbad, CA). The obtained soil RNA sample was treated with RNase-free DNase I. After a final precipitation step, the RNA was dissolved in nuclease-free water. RNA integrity was checked by Bioanalyzer 2100 (Agilent Technologies, USA) electrophoresis and RNA quantity and purity were determined by spectrophotometry (SAFAS UVmc2, SAFAS Monaco). cDNA synthesis and size fractionation of cDNA libraries were performed as described in our earlier work (Yadav et al., 2014). cDNAs were ligated downstream of the *Saccharomyces cerevisiae* PGK1 promoter in a modified pFL61 yeast expression vector containing SfiIA and SfiIB sites (Minet et al., 1992). The cDNA library B (0.5 kb- 1.0 kb) was screened for cadmium tolerant genes by transforming into mutant strain $ycf1^{\Delta}$. The transformation was done by lithium acetate method (Gietz and Schiestl, 2007) and cadmium tolerant transformants were selected by spreading on SD-Ura mediim supplemented with 40 μM $CdSO_4$ and incubated at 30 °C for 48–72 h. Functional complementation experiments were executed by drop out assay and growth assay. Drop out assay involved growth of Cd tolerant transformants in SD-Ura medium at 30 °C and 220 r.p.m. until their optical density (OD_{600}) reached up to 1.0. Dilutions of pre-grown cultures with OD_{600} of 1.0 were made up to 0.0001 and 5.0 μl of each dilution was dropped on SD-Ura control plates and SD-Ura supplemented with 40 μM $CdSO_4$. Simultaneously, growth assay was performed to know the level of expression of Cd tolerant gene by inoculating the $ycf1^{\Delta}$ transformant in 20 ml media supplemented with 40 μM, 60 μM and 80 μM $CdSO_4$ and the growth was measured (OD_{600}) after 48 h. One of the transformants, PLBe1 was selected for further studies based on its tolerance to high concentrations of cadmium.

2.4. Sequence analysis

Plasmid DNA was extracted from the $ycf1^{\Delta}$ transformant carrying PLBe1 by using Zymoprep™ Yeast Plasmid Miniprep II kit (Zymo Research, USA) as per the manufacturer instructions. The plasmid was then electro-transformed into One Shot® TOP10 Electrocomp™ *E. coli* cells (Thermofisher, USA). Plasmids were extracted from electro-transformed TOP10 cells and the insert was sequenced by using vector specific primers NF (5′-CAGATCATCAAGGAAGTAATTATCTAC-3′) and NR (5′-CAGAAAAGCAGGCTGGG AAGC-3′). The sequence was compared for homologous sequences by using BLASTX analysis of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Homologous sequences along with PLBe1 sequence were aligned by multiple sequence alignment using Clustal Omega (http://www.ebi.ac.uk/ Tools/msa/clustalo/). The potential O– and N- linked glycosylation sites were predicted using NetOglyc 3.1 (http://www.cbs.dtu.dk/ services/NetOGlyc/) and NetNGlyc 1.0 server (http://www.cbs.dtu.dk/ services/NetNGlyc/). Potential phosphorylation sites present in the polypeptide were determined by NetPhos 2.0 server (http://www.cbs.dtu.dk/services/ NetPhos/). The cDNA (PLBe1) sequence obtained in this study was deposited at NCBI under the accession number MH244348.

2.5. Screening of PLBe1 gene for multi-metal tolerance

The Cd tolerant PLBe1 cDNA was tested for its ability to tolerate other heavy metals such as Cu, Zn and Co. The recombinant plasmid was transformed to Cu sensitive (*cup1*$^{\Delta}$), Zn sensitive (*zrc1*$^{\Delta}$) and Co sensitive (*cot1*$^{\Delta}$) yeast mutants and were regenerated on metal amended media with respective metal concentrations (Cu 150 µM, Zn 10 mM and Co 2 mM). For positive and negative controls, wild strains and mutant strains were transformed with empty pFL61vector. Furthermore, drop out tests were performed as described above with respective metals. Subsequently, growth assays for respective metal specific transformants was carried out by inoculating log phase cultures of transformed *cup1*$^{\Delta}$, *cot1*$^{\Delta}$ and *zrc1*$^{\Delta}$ cells into flasks containing 20 ml of SD-Ura medium supplemented in a range of metal concentrations Cu (150 µM - 500 µM), Zn (10–12 mM) and Co (2 mM - 4 mM). The growth was measured (OD$_{600}$) after 48 h.

3. Results and discussion

3.1. Soil analysis

The collected soil sample is sandy luvisol and the physicochemical properties of the soil are 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 of the soil samples are as follows: Cd 2.5, Cu 64 and Zn 385 mg/kg of soil.

3.2. Screening for cadmium tolerance

Cadmium hypersensitive mutant *ycf1*$^{\Delta}$ was used to screen the environmental cDNA library B synthesized from total RNA of soil. Restoration of Cd tolerant phenotype of the mutant strain due to transformed environmental cDNA facilitated the selection of transformant *ycf1*$^{\Delta}$PLBe1 as a subject of further experiments (Fig. 1a). The screened transformant *ycf1*$^{\Delta}$PLBe1 was able to tolerate high concentrations of Cd. The maximum growth of this transformant was observed at 40 µM Cd. With increase in concentration of Cd, reduction in the growth was observed. However the growth was significantly higher compared to the wild type strain. No growth was observed when *ycf1*$^{\Delta}$ transformed with empty pFL61 in presence of Cd (Fig. 1b). These results signify the role of PLBe1 cDNA in providing Cd tolerance to the hypersensitive yeast mutant *ycf1*$^{\Delta}$.

Present study demonstrated the significant role of metatranscriptomic approach in revealing metabolic potentials and ecological roles of microorganisms present in various adverse environments like heavy metal contaminated site. The complementation assay using metal sensitive yeast *S. cerevisae* mutants has been successfully used in characterizing a number of metal tolerant genes (Bellion et al., 2007; Ramesh et al., 2009; Reddy et al., 2014, 2016; Ziller et al., 2017). In the yeast complementation assay, metal tolerant genes were under the control of phosphoglycerokinase (PGK) promoter. Metatranscriptomics applied to environmental transcripts led to the identification of fungal genes encoding for acid phosphatase and imidazoleglycerol-phosphate dehydratase from the organic layer of a sugar maple forest soil (Kellner et al., 2010). In a similar study, diverse gene sequences were recovered from soil metatranscriptomic cDNA library, with a potential to restore tolerance towards heavy metals (Lehembre et al., 2013). Various contaminated sites such as deepwater horizon oil spill, acid mine drainage, were explored through metatranscriptomics with a purpose to reveal numerous microbes with metabolic and ecological significance in order to remediate such adverse environmental conditions (Mason et al., 2014; Hua et al., 2015). Ziller et al. (2017) characterised five metatranscriptomic sequences of unknown taxonomic origin as cysteine rich proteins on the basis of their metal binding potential from contaminated soil.

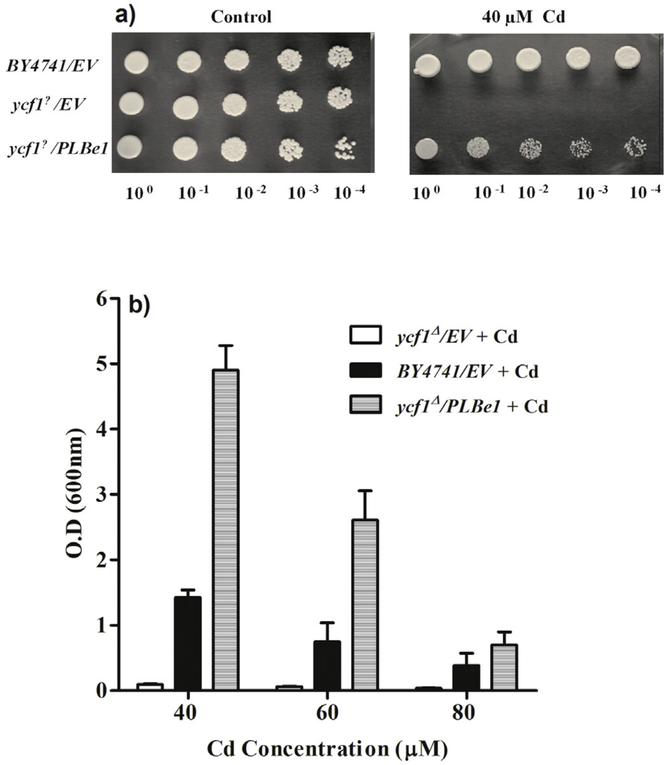
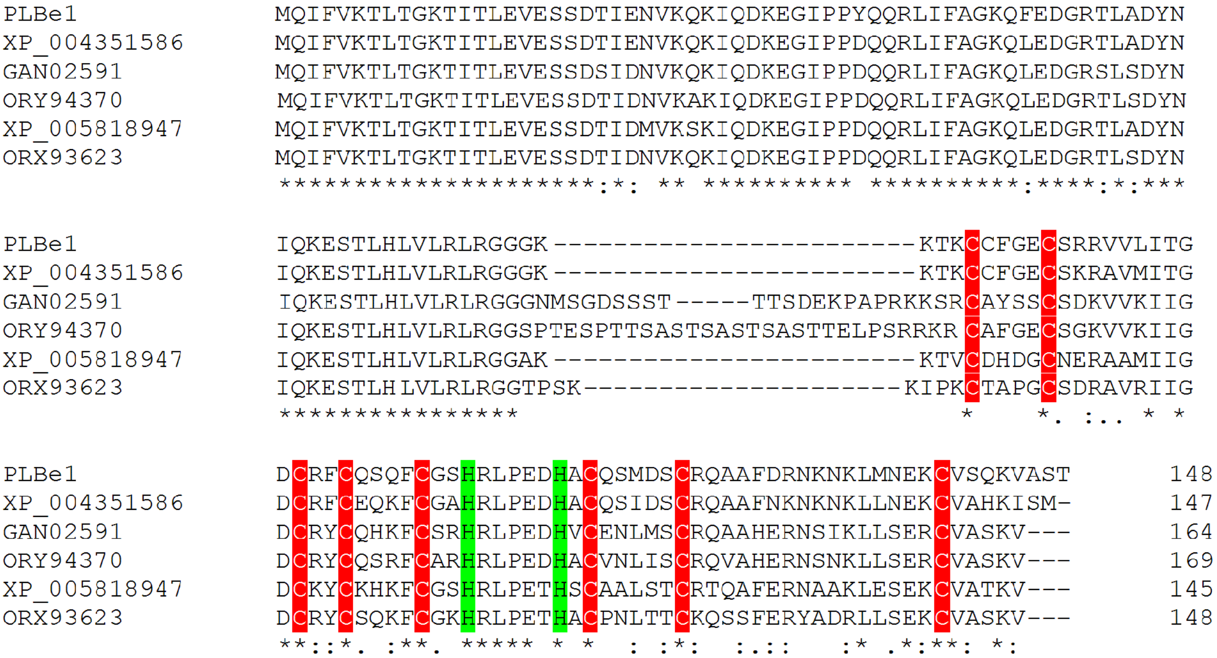


Fig. 1. a) Functional complementation of *Saccharomyces cerevisae* mutant *ycf1*$^{\Delta}$ on selective media. Mutant strains were transformed with empty vector (EV) pFL61, or vector containing PLBe1. Diluted transformant cultures were spotted on SD-Ura medium with or without Cd supplement as indicated. Wild-type strains transformed with EV were used as positive controls. b) Growth assay of mutant strain *ycf1*$^{\Delta}$ expressing PLBe1 in SD-Ura broth with Cd metal supplementation 40 µM - 80 µM for 48 h.

3.3. Sequence analysis

The BLASTX analysis with the deduced nucleotide sequences of PLBe1 revealed 694 bp cDNA showing 97% identity (98% positives) with ubiquitin domain containing protein (UFP) fused with AN1 type Zn finger of *Acanthameoba castellani* (XP_004333233) The encoding polypeptide showed two prime characteristic features: presence of N-terminal 76 aa ubiquitin domain followed by an extension of AN1 type Zn finger (Zf_AN1) located at C-terminal. Amino acid sequence alignment of the deduced polypeptide with various other members of ubiquitin domain containing protein or zinc finger fusion protein showed the presence of conserved domains in the sequence which advocates this polypeptide as a member of UFP (Fig. 2). As shown in Fig. 3, the presence of diglycine residues at the N- terminal of ubiquitin domain in fusion protein UFP signify the proteolytic processing of ubiquitin moiety (Walters et al., 2002) while lysine residue at 48 and 63 positions (K48 and K63) advocates the formation of polyubiquitin chains through chain elongation *via* K48 and K63 after posttranslational modification (Pickart and Fushman, 2004). K48 residue site corresponds as attachment site for other ubiquitin molecules which showed that the ubiquitin moiety of UFP after posttranslational modifications helps in binding of other ubiquitin molecules to form polyubiquitin chains for protein degradation while K63 corresponds to non-degradative signalling pathways for homeostasis. In the present study, UFP has ubiquitin domain along with 66 aa long extension protein with characteristic zinc finger features (Fig. 3). The characteristic pattern of Zn finger (C-X2-C-X (9–12)-C-X(1–2)-C-X4-C-X2-H-X5-H-X-C) at the C terminal of this fusion protein corresponds to Zf_AN1 and structure of Zn ring finger showed two interleaved Zn binding sites (Freemont, 2000). Characterisation of 147 aa long polypeptide also depicts various metabolically significant sites such as 14 potential sites for phosphorylation and three O-linked glycosylation sites. Klug and Rhodes (1987) reported that UFP comprised of two domains where carboxyl group of ubiquitin



```
PLBe1          MQIFVKTLTGKTITLEVESSDTIENVKQKIQDKEGIPPYQQRLIFAGKQFEDGRTLADYN
XP_004351586   MQIFVKTLTGKTITLEVESSDTIENVKQKIQDKEGIPPDQQRLIFAGKQLEDGRTLADYN
GAN02591       MQIFVKTLTGKTITLEVESSDSIDNVKQKIQDKEGIPPDQQRLIFAGKQLEDGRSLSDYN
ORY94370       MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYN
XP_005818947   MQIFVKTLTGKTITLEVESSDTIDMVKSKIQDKEGIPPDQQRLIFAGKQLEDGRTLADYN
ORX93623       MQIFVKTLTGKTITLEVESSDTIDNVKQKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYN
               *********************:*: ** ********** **********:****:*:***

PLBe1          IQKESTLHLVLRLRGGGK------------------------KTKCCFGECSRRVVLITG
XP_004351586   IQKESTLHLVLRLRGGGK------------------------KTKCCFGECSKRAVMITG
GAN02591       IQKESTLHLVLRLRGGGNMSGDSSST-----TTSDEKPAPRKKSRCAYSSCSDKVVKIIG
ORY94370       IQKESTLHLVLRLRGGSPTESPTTSASTSASTSASTTELPSRRKRCAFGECSGKVVKIIG
XP_005818947   IQKESTLHLVLRLRGGAK------------------------KTVCDHDGCNERAAMIIG
ORX93623       IQKESTLHLVLRLRGGTPSK---------------------KIPKCTAPGCSDRAVRIIG
               ***************                              *    *. :.. * *

PLBe1          DCRFCQSQFCGSHRLPEDHACQSMDSCRQAAFDRNKNKLMNEKCVSQKVAST   148
XP_004351586   DCRFCEQKFCGAHRLPEDHACQSIDSCRQAAFNKNKNKLLNEKCVAHKISM-   147
GAN02591       DCRYCQHKFCSRHRLPEDHVCENLMSCRQAAHERNSIKLLSERCVASKV---   164
ORY94370       DCRYCQSRFCARHRLPEDHACVNLISCRQVAHERNSNKLLSERCVASKV---   169
XP_005818947   DCKYCKHKFCGSHRLPETHSCAALSTCRTQAFERNAAKLESEKCVATKV---   145
ORX93623       DCRYCSQKFCGKHRLPETHACPNLTTCKQSSFERYADRLLSEKCVASKV---   148
               **::*.  :**. ***** * *  : :*:  :.::   :* .*:**: *:
```

Fig. 2. Multiple sequence alignment of *PLBe1* gene with other homologous sequences retrieved by BLASTp analysis. Alignment was performed using the Clustal omega program and manually adjusted. Fully conserved cysteine residues (highlighted in red) and histidine residues (highlighted in green) of C– terminal zinc finger domain is shown. Accession numbers are *Mucor ambiguus* (GAN02591) *Syncephalastrum racemosum* (ORY94370), *Acanthameoba castellini* (XP_004351586), *Guillardia theta* CCMP2712 (XP_005818947), *Basidiobolus meristosporus* CBS 931.73 (ORX93623). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

domain can be linked to amino group of one or two small non related proteins varying in length from 52 aa to 80 aa. Generally these C - terminal extension proteins are highly basic; contain a 4 cysteine residue sequence in consensus with the Zn^{2+} finger domain present in nucleic acid binding proteins (Kalderon et al., 1984). Similar fusion proteins encoded by two isoforms of An1 mRNA in *Xenopus laevis* have been characterised by Linnen et al. (1993).

It had been reported that ubiquitin is available in the living beings either as polyubiquitin form or as a fusion protein at C-terminal and their post-translational modifications led to formation of ubiquitin monomers (Finley et al., 1987; Redman and Rechsteiner, 1989). Various studies reported the immediate cleavage of these fusion proteins encoded from genetically engineered ubiquitin fusion genes hence depicted the simultaneous *in vivo* processing of amino peptide linked ubiquitin (Bachmair et al., 1986; Bachmair and Varshavsky, 1989; Butt et al., 1988). The presence of diglycine residue at N- terminal and lysine residue at 48 and 63 positions support UFP encoded by cDNA PLBE1 as potential cofactors for ubiquitin proteosome system (UPS) as in case of yeast Cuz1/YNL155w proteins that have significant role in ubiquitin mediated proteasomal degradation of proteins (Sá-Moura et al., 2013). Ubiquitin- proteasome system (UPS) is a network that helps in degradation of aberrant proteins in eukaryotes (Ravid and Hochstrasser, 2008). It has been reported that ubiquitin target the degradation of abnormal proteins following environmental stress by UPS (Finley et al., 1987; Amerik et al., 2005).

3.4. Multi-metal tolerance of PLBe1

To validate the role of PLBe1 towards the tolerance of other metals such as Cu, Zn and Co, the recombinant plasmid carrying PLBe1 cDNA was transformed into *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ mutant strains of *S. cerevisiae*. It was observed that the Cd tolerant cDNA PLBe1 also conferred tolerance to the hypersensitive yeast mutant *cup1*$^{\Delta}$ (DTY4) at 150 μM Cu while no growth was observed for yeast mutant *cup1*$^{\Delta}$ transformed with empty vector pFL61 (Fig. 4a). In Zn amended (10 mM) media, prolific growth was observed for transformant *zrc1*$^{\Delta}$PLBe1 in comparison to the mutant strain *zrc1*$^{\Delta}$pFL61 (Fig. 4b). Similar results were obtained for cobalt metal (2 mM) while studying the restoring efficiency of cDNA PLBe1 towards cobalt toxicity. The transformant *cot1*$^{\Delta}$PLBe1 grows at a similar rate to that of the wild strain BY4741 in comparison to *cot1*$^{\Delta}$pFL61 (Fig. 4c).

To study the metal tolerant profiling of cDNA PLBe1 for various metals, the transformants were grown in liquid media supplemented with different concentrations of the metals. The transformant *cup1*$^{\Delta}$PLBe1 was able to grow up to 300 μM Cu observing the maximum growth at 150 μM. It was observed that *cup1*$^{\Delta}$PLBe1 conferred four times more tolerance compared to the *cup1*^{s} (DTY3) which possess both copies of *CUP1* gene at 150 μM Cu concentration. The growth of *cup1*$^{\Delta}$PLBe1 was significantly higher than its wild type and *cup1*$^{\Delta}$ transformed with empty vector (Fig. 5a). The growth of *zrc1*$^{\Delta}$PLBe1 also increased its tolerance to high concentrations of Zn (12 mM) showing the maximum growth at 10 mM of Zn. It was observed that the growth of transformant *zrc1*$^{\Delta}$PLBe1 at 10 mM concentration of Zn was 72%

```
MQIFVKTLTGKTITLEVESSDTIENVKQKIQDKEGIPPYQQRLIFAGKQF    50
EDGRTLADYNIQKESTLHLVLRLRGGGKKTKCCFGECSRRVVLITGDCRF   100
CQSQFCGSHRLPEDHACQSMDSCRQAAFDRNKNKLMNEKCVSQKVAST     147
```

Fig. 3. Deduced amino acid sequence of UFP encoded by cDNA PLBe1: Highly conserved lysine residue at 48 position (green highlighted) and diglycine residue in the C-terminal end of ubiquitin moiety at 75 and 76 aa (highlighted in blue). Cysteine and histidine residues of C-terminal zinc finger moiety of the fusion protein are highlighted in red. Phosphorylation sites are highlighted in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

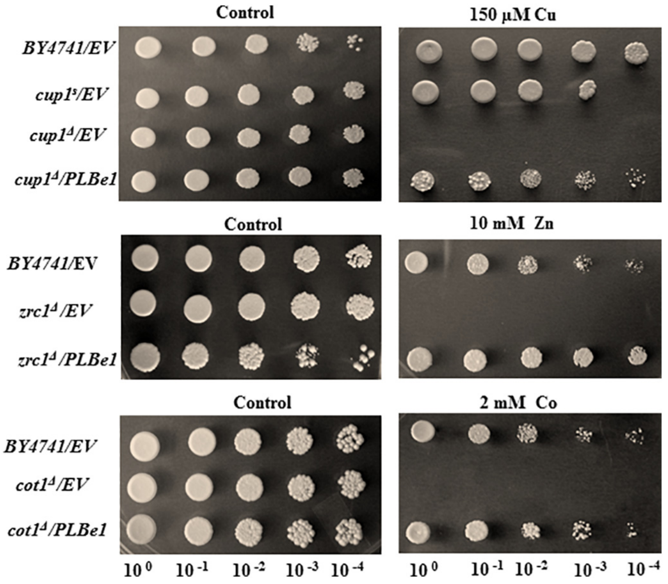


Fig. 4. Functional complementation of *Saccharomyces cerevisae* mutants, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ on selective media. Mutant strains were transformed with empty vector (EV) pFL61, or vector containing *PLBe1*. Diluted transformant cultures were spotted on SD-Ura medium with or without metal supplement as indicated. Wild-type strains *cup1*^{S} and BY4741 transformed with EV were used as positive controls.

more when compared to the wild strain BY4741. The growth of *zrc1*$^{\Delta}$PLBe1 was significantly higher than its wild type and *zrc1*$^{\Delta}$ transformed with empty vector (Fig. 5b). The cobalt sensitive yeast transformant carrying cDNA PLBe1 (*cot1*$^{\Delta}$pFL61) had shown high tolerance to Co with maximum growth at 2 mM concentration. The growth assay had shown that *cot1*$^{\Delta}$pFL61 was 74% more capable of tolerating 2 mM of Co concentration in comparison to yeast mutant. The growth of *cot1*$^{\Delta}$pFL61 was much higher than its wild type and mutant strain transformed with empty vector in all the concentrations of Co tested (Fig. 5c).

Ubiquitin fusion protein encoded by cDNA PLBe1 showed immense tolerance towards all the four metals in the study in an order: Zn > Co > Cu > Cd. Woo et al. (2009) observed that ubiquitin encoding cDNA *OjaUB* showed elevated expression level in order to remove numerous aberrant proteins synthesized after exposure to different concentrations of Cd, Cu, Zn, Ni, Cr and Ag. Various studies have reported the significant role of ubiquitin and Zn finger protein as stress associated proteins. Leach et al. (2011) reported that ubiquitination plays an important role in growth, stress responses and metabolic adaptation of *Candida albicans*. Proteins with Zn - finger are also among the potential members of stress associated proteins both in plants (Mukhopadhyay et al., 2004) and animals (Linnen et al., 1993; Evans et al., 2004; Huang et al., 2004; Heyninck and Beyaert, 2005; Hishiya et al., 2006). Tolerance of UFP towards Zn may also be due to the presence of two interleaved Zn^{2+} binding sites in N - terminal Zn finger extension of cDNA PLBe1, similar to the Zn finger pattern observed by Freemont (2000). This interprets that various metals used in present study induced the expression of UFP encoded by cDNA PLBe1 in response to degradation of numerous proteins damaged due to metal toxicity by ubiquitination process of proteasome - dependent pathway as observed in a study by Woo et al. (2009). Henceforth, this study contributes UFP as a prospective member of metal tolerant proteins in response to metal contamination in soil environment.

4. Conclusion

In conclusion, the cDNA PLBe1 have the potential to tolerate heavy metal contamination in the environment. Presence of ubiquitn domain and zinc finger domain in translated protein UFP signified the role of both domains in abiotic stress conditions like heavy metal contamination, hence can be further explored as biomarker to detect heavy metal contamination. These results altogether categorised functional metatranscriptomics as a modern ‘omic’ approach that helps in exploring the environment for numerous unknown mechanisms and novel elements like UFP, responsible for adapting the unfavourable conditions or abiotic stress. Metatranscriptomic approach revealed high degree of diversity while exploring the nature for countless known and unknown

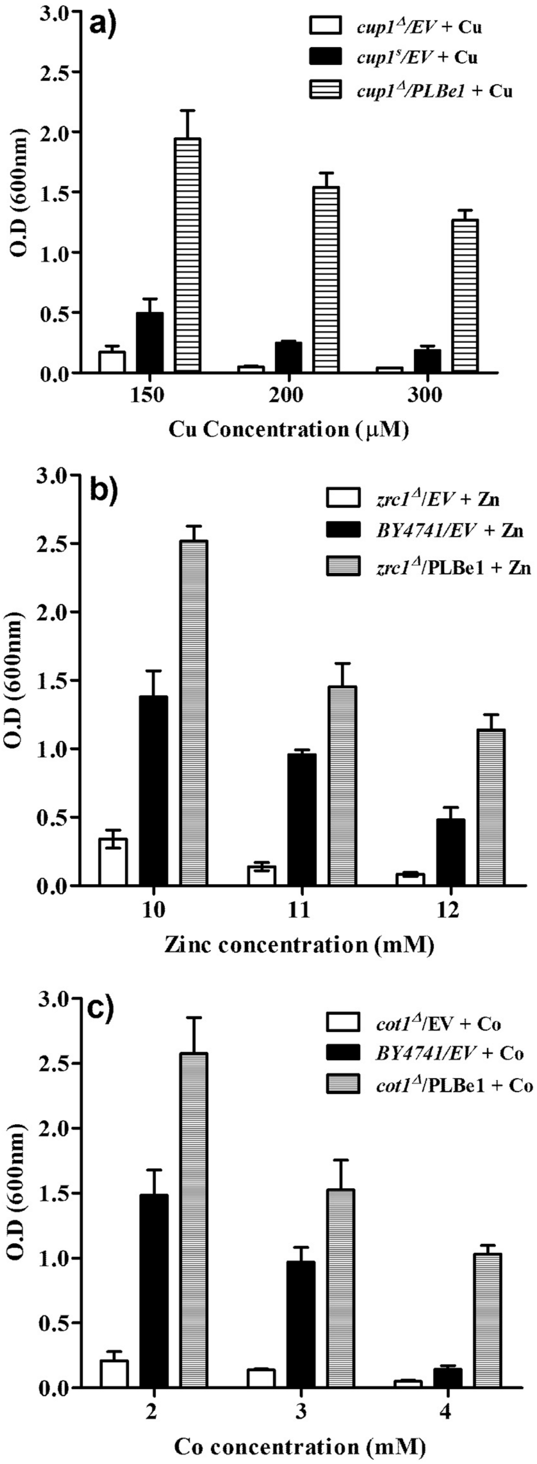


Fig. 5. Growth assay of metal sensitive mutant strains *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLBe1 in SD-Ura broth with metal supplementation for 48 h (a) Copper 150–300 µM, (b) Zinc 10 mM - 12 mM, (c) Cobalt 2 mM - 4 mM.

phenomenonon and its importance in abiotic stress response. Most of the work on ubiquitin or UFP or zinc finger was host specific while in this study; functional metatranscriptomics approach was explored to screen the cDNA PLBe1 encoding UFP and studied the heterologous expression for heavy metal tolerance hence is novel findings of this work. UFP can be used as potential member to revitalise the heavy metal contaminated sites or can be used as a biomarker to detect heavy metal contamination in the soil environment.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This study is funded by Indo-French Centre for the Promotion of Advanced Research (IFCPAR), New Delhi, India under the grant No. 4709-1.

References

Adler, G., Konrad, Z., Zamir, L., Mishra, A.K., Raveh, D., Bar-Zvi, D., 2017. The Arabidopsis paralogs, PUB46 and PUB48, encoding U-box E3 ubiquitin ligases, are essential for plant response to drought stress. BMC Plant. Biol. 17 (1), 8.

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., Narasimhan, G., 2016. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. Evol. Bioinform. Online 12 (Suppl. 1), 5.

Amerik, A.Y., Li, S.-J., Hochstrasser, M., 2005. Analysis of the deubiquitinating enzymes of the yeast *Saccharomyces cerevisiae*. Biol. Chem. 381, 981–992.

Bachmair, A., Varshavsky, A., 1989. The degradation signal in a short-lived protein. Cell 56 (6), 1019–1032.

Bachmair, A., Finley, D., Varshavsky, A., 1986. In vivo half-life of a protein is a function of its amino-terminal residue. Science 234 (4773), 179–186.

Bailly, J., Fraissinet-Tachet, L., Verner, M.C., Debaud, J.C., Lemaire, M., Wésolowski-Louvel, M., Marmeisse, R., 2007. Soil eukaryotic functional diversity, a metatranscriptomicapproach. ISME J. 1, 632–642.

Bell, T.H., Joly, S., Pitre, F.E., Yergeau, E., 2014. Increasing phytoremediation efficiency and reliability using novel omics approaches. Trends Biotechnol. 32 (5), 271–280.

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

Blindauer, C.A., Leszczyszyn, O.I., 2010. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. Nat. Pro. Rep. 27 (5), 720–741.

Butt, T.R., Khan, M.I., Marsh, J., Ecker, D.J., Crooke, S.T., 1988. Ubiquitin -metallothionein fusion protein expression in yeast. A genetic approach for analysis of ubiquitin functions. J. Biol. Chem. 263 (31), 16364–16371.

Chen, L.X., Hu, M., Huang, L.N., Hua, Z.S., Kuang, J.L., Li, S.J., Shu, W.S., 2015. Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. ISME J. 9 (7), 1579.

Clemens, S., 2006. Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. Biochimie 88 (11), 1707–1719.

Clemens, S., Ma, J.F., 2016. Toxic heavy metal and metalloid accumulation in crop plants and foods. Annu. Rev. Plant Biol. 67, 489–512.

Conklin, D.S., McMaster, J.A., Culbertson, M.R., Kung, C., 1992. COT1, a gene involved in cobalt accumulation in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12, 3678–3688.

Damon, C., Vallon, L., Zimmermann, S., Haider, M.Z., Galeote, V., Dequin, S., Luis, P., Fraissinet-Tachet, F., Marmeisse, R., 2011. A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. ISME J. 5, 1871–1880.

Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L., Marmeisse, R., 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. PLoS One 7, e28967.

Evans, P.C., Ovaa, H., Hamon, M., Kilshaw, P.E., Hamm, S., Bauer, S., Ploegh, H.L., Smith, T.S., 2004. Zinc-finger protein A20, a regulator of inflammation and cell survival, has de-ubiquitinating activity. Biochem. J. 378, 727–734.

Facchinelli, A., Sacchi, E., Mallen, L., 2001. Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. Environ. Pollut. 114, 313–324.

Fierer, N., Bradford, M.A., Jackson, R.B., 2007. Toward an ecological classification of soil bacteria. Ecology 88, 1354–1364.

Finley, D., Özkaynak, E., Varshavsky, A., 1987. The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. Cell 48, 1035–1046.

Forzani, C., Lobréaux, S., Mari, S., Briat, J.F., Lebrun, M., 2002. Metal resistance in yeast mediated by the expression of a maize 20S proteasome α subunit. Gene 293 (1), 199–204.

Freemont, P.S., 2000. RING for destruction? Curr. Biol. 10 (2), R84–R87.

Gadd, G.M., 2010. Metals, minerals and microbes: geomicrobiology and bioremediation. Microbiol. 156, 609–643.

Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31–34.

Heyninck, K., Beyaert, R., 2005. A20 inhibits NF-_B activation by dual ubiquitin-editing functions. Trends Biochem. Sci. 30, 1–4.

Hishiya, A., Iemura, S., Natsume, T., Takayama, S., Ikeda, K., Watanabe, K., 2006. A novel ubiquitin-binding protein ZNF216 functioning in muscle atrophy. EMBO J. 25, 554–564.

Hua, Z.S., Han, Y.J., Chen, L.X., Liu, J., Hu, M., Li, S.J., Kuang, J.L., Chain, S.G.P., Huang, L.N., Shu, W.S., 2015. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. ISME J. 9 (6), 1280.

Huang, J., Teng, L., Li, L., Liu, T., Li, L., Chen, D., Xu, L.G., Zhai, Z., Shu, H.B., 2004. ZNF216 is an A20-like and IkappaB kinase gamma- interacting inhibitor of NFkappaB activation. J. Biol. Chem. 279, 16847–16853.

Hwang, G.W., Furuchi, T., Naganuma, A., 2007. Ubiquitin-conjugating enzyme Cdc34 mediates cadmium resistance in budding yeast through ubiquitination of the transcription factor Met4. Biochem. Biophys. Res. Commun. 363, 873–878.

Jungmann, J., Reins, H.A., Schobert, C., Jentsch, S., 1993. Resistance to cadmium mediated by ubiquitindependent proteolysis. Nature 361, 369–371.

Kalderon, D., Roberts, B.L., Richardson, W.D., Smith, A.E., 1984. A short amino acid sequence able to specify nuclear location. Cell 39 (3), 499–509.

Kellner, H., Luis, P., Portetelle, D., Vandenbol, M., 2010. Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. Microbiol. Res. 166, 360–368.

Kitson, R.E., Mellon, M.G., 1944. Colorimetric determination of phosphorus as molybdovanadophosphoric acid. Ind. Eng. Chem. 16, 379–383.

Klug, A., Rhodes, D., 1987. 'Zinc fingers': a novel protein motif for nucleic acid recognition. Trends Biochem. Sci. 12, 464–469.

Leach, M.D., Stead, D.A., Argo, E., MacCallum, D.M., Brown, A.J., 2011. Molecular and proteomic analyses highlight the importance of ubiquitination for the stress resistance, metabolic adaptation, morphogenetic regulation and virulence of *Candida albicans*. Mol. Microbiol. 79 (6), 1574–1593.

Lehembre, F., Doillon, D., David, E., Perrotto, S., Baude, J., Foulon, J., Harfouche, L., Vallon, L., Poulain, J., Da Silva, C., Wincker, P., 2013. Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. Environ. Microbiol. 15, 2829–2840.

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,* YCF1Δ-catalyzed transport of bis(glutathionato) cadmium. Proc. Natl. Acad. Sci. U. S. A. 94, 42–47.

Linnen, J., Bailey, M., Week, C.P., 1993. Two related localized mRNAs from Xenopus Zaevis encode ubiquitin-like fusion proteins. Gene 128, 181–188.

Longo, V.D., Gralla, E.B., Valentine, J.S., 1996. Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae*. J. Biol. Chem. 271, 12275–12280.

Marmeisse, R., Kellner, H., Fraissinet-Tachet, L., Luis, P., 2017. Discovering protein-coding genes from the environment: time for the eukaryotes? Trends Biotechnol. 35 (9), 824–835.

Mason, O.U., Scott, N.M., Gonzalez, A., Robbins-Pianka, A., Bælum, J., Kimbrel, J. Bouskill, Prestat, E., Borglin, S., Joyner, D.C., Fortney, J.L., 2014. Metagenomics reveals sediment microbial community response to Deepwater horizon oil spill. ISME J. 8 (7), 1464.

Medicherla, B., Goldberg, A.L., 2008. Heat shock and oxygen radicals stimulate ubiquitin-dependent degradation mainly of newly synthesized proteins. J. Cell Biol. 182, 663–673.

de Menezes, A., Clipson, N., Doyle, E., 2012. Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. Environ. Microbiol. 14 (9), 2577–2588.

Minet, M., Dufour, M.E., Lacroute, F., 1992. Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. Curr. Genet. 2, 417–422.

Miyabe, S., Izawa, S., Inoue, Y., 2001. The Zrc1 is involved in zinc transport system between vacuole and cytosol in *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun. 282, 79–83.

Moor, C., Lymberopoulou, T., Dietrich, V.J., 2001. Determination of heavy metals in soils, sediments and geological materials by ICP-AES and ICP-MS. Microchim. Acta 136 (3–4), 123–128.

Mukhopadhyay, A., Vij, S., 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. U. S. A. 101, 6309–6314.

Olsen, S.R., Cole, C.V., Watanabe, F.S., Dean, L.A., 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. In: U.S. Government printing office. US Department of Agriculture. Circular No. 939, Washington (DC).

Pena, L.B., Pasquini, L.A., Tomaro, M.L., Gallego, S.M., 2007. 20S proteasome and accumulation of oxidized and ubiquitinated proteins in maize leaves subjected to cadmium stress. Phytochemistry 68 (8), 1139–1146.

Peng, H., Yang, J., Li, G., You, Q., Han, W., Li, T., Gao, D., Xie, X., Lee, B.H., Du, J., Hou, J., 2017. Ubiquitylation of p62/sequestosome1 activates its autophagy receptor function and controls selective autophagy upon ubiquitin stress. Cell Res. 27, 657–674.

Pickart, C.M., Fushman, D., 2004. Polyubiquitin chains: polymeric protein signals. Curr. Opin. Chem. Biol. 8, 610–616.

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.

Prosser, J.I., 2007. Microorganisms Cycling Soil Nutrients and Their Diversity. In: Modern Soil Microbiology, Second Edition. CRC Press, Boca Raton, pp. 237–262.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., Forster, R.J., 2011. Snapshot of the eukaryotic gene expression in muskoxen rumen - a metatranscriptomic approach. PLoS One 6, e20521.

Ramesh, G., Podila, G.K., Gay, G., Marmeisse, R., Reddy, M.S., 2009. Copper and cadmium metallothioneins of the ectomycorrhizal fungus *Hebeloma cylindrosporum* have different patterns of regulation. Appl. Environ. Microbiol. 75, 2266–2274.

Ravid, T., Hochstrasser, M., 2008. Diversity of degradation signals in the ubiquitin proteasome system. Nat. Rev. Mol. Cell Biol. 9, 679–690.

Reddy, M.S., Prasanna, L., Marmeisse, R., Fraissinet-Tachet, L., 2014. Differential expression of metallothioneins in response to heavy metals and their involvement in metal tolerance in the symbiotic basidiomycete *Laccaria bicolor*. Microbiol. 160, 2235–2242.

Reddy, M.S., Kour, M., Aggarwal, S., Ahuja, S., Marmeisse, R., Fraissinet-Tachet, L., 2016. Metal induction of a *Pisolithus albus* metallothionein and its potential involvement in heavy metal tolerance during mycorrhizal symbiosis. Environ. Microbiol. 18, 2446–2454.

Redman, K.L., Rechsteiner, M., 1989. Identification of the long ubiquitin extension as ribosomal protein S27a. Nature 338 (6214), 438.

Sá-Moura, B., Funakoshi, M., Tomko, R.J., Dohmen, R.J., Wu, Z., Peng, J., Hochstrasser, M., 2013. A conserved protein with AN1 zinc finger and ubiquitin-like domains modulates Cdc48 (p97) function in the ubiquitin-proteasome pathway. J. Biol. Chem. 288 (47), 33682–33696.

Solgi, E., Esmaili-Sari, A., Riyahi-Bakhtiari, A., Hadipour, M., 2012. Soil contamination ofmetals in the three industrial estates, arak. Iran. Bull. Environ. Contam. Toxicol. 88, 634–638.

Teng, W., Kuang, J., Luo, Z., Shu, W., 2017. Microbial diversity and community assembly across environmental gradients in acid mine drainage. Fortschr. Mineral. 7 (6), 106.

Thavamani, P., Samkumar, R.A., Satheesh, V., Subashchandrabose, S.R., Kavitha, R., Naidu, R., Venkateswarlu, K., Megharaj, M., 2017. Microbes from mined sites: harnessing their potential for reclamation of derelict mine sites. Environ. Pollut. 230, 495–505.

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.

Walters, K.J., Kleijnen, M.F., Goh, A.M., Wagner, G., Howley, P.M., 2002. Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. Biochemist 41, 1767–1777.

Wei, B.G., Yang, L.S., 2010. A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. Microchem. J. 94, 99–107.

Woo, S., Yum, S., Park, H.S., Lee, T.K., Ryu, J.C., 2009. Effects of heavy metals on antioxidants and stress-responsive gene expression in Javanese medaka *(Oryzias javanicus)*. Comp. Biochem. Physiol. Part C: Toxicol. Pharmacol. 149 (3), 289–299.

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.

Yu, K., Zhang, T., 2012. Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS One 7 (5), e38183.

Zielinski, B.L., Allen, A.E., Carpenter, E.J., Coles, V.J., Crump, B.C., Doherty, M., Foster, R.A., Goes, J.I., Gomes, H.R., Hood, R.R., McCrow, J.P., 2016. Patterns of transcript abundance of eukaryotic biogeochemically-relevant genes in the Amazon River plume. PLoS One 11 (9), e0160929.

Ziller, A., Yadav, R.K., Capdevila, M., Reddy, M.S., Vallon, L., Marmeisse, R., Atrian, S., Palacios, O., 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.

Science of the Total Environment 661 (2019) 432–440

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Multi-metal tolerance of von Willebrand factor type D domain isolated from metal contaminated site by metatranscriptomics approach

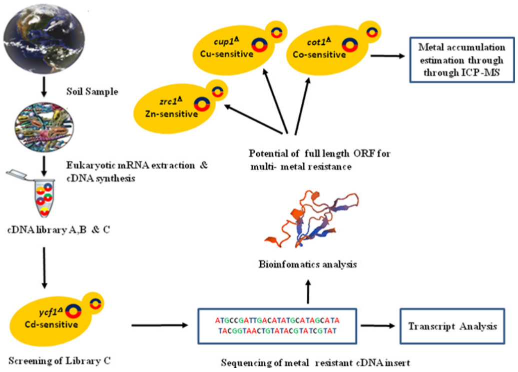
Bharti Thakur [a], Rajiv Yadav [b], Laurent Vallon [b], Roland Marmeisse [b], Laurence Fraissinet-Tachet [b], M. Sudhakara Reddy [a,*]

[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

HIGHLIGHTS

- Sized eukaryotic cDNA libraries were constructed from metal polluted soil.
- Cd tolerant PLCe10 clone was isolated from library C by heterologous expression.
- PLCe10 transcript showed homology to VWD domain of vitellogenin.
- PLCe10 also provided tolerance to Cu, Zn and Co with metal sensitive yeast mutants.
- VWD serves as bioremediating agent and also biomarker in metal contaminated sites.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 16 November 2018
Received in revised form 15 January 2019
Accepted 16 January 2019
Available online 17 January 2019

Editor: Shuzhen Zhang

Keywords:
Metatranscriptomics
Heavy metals
VWD
Vitellogenin
Heavy metal tolerance

ABSTRACT

Environmental pollution through heavy metals is an upcoming universal problem that relentlessly endangers human health, biodiversity and ecosystems. Hence remediating these heavy metal pollutants from the environment by engineering soil microbiome through metatranscriptomics is befitting reply. In the present investigation, we have constructed size fractionated cDNA libraries from eukaryotic mRNA of cadmium (Cd) contaminated soil and screened for Cd tolerant genes by yeast complementation system by using Cd sensitive *ycf1*$^{\Delta}$ mutant. We are reporting one of the transformants PLCe10 (from library C, 1–4 kb) with potential tolerance towards Cd toxicity (40 μM–80 μM). Sequence analysis of PLCe10 transcript showed homology to von Willebrand factor type D domain (VWD) of vitellogenin-6 of *Ascaris suum* encoding 338 amino acids peptide. qPCR analysis revealed that PLCe10 induced in presence of Cd (32 fold) and also accumulated maximum amount of Cd at 60 μM Cd. This cDNA was further tested for its tolerance against other heavy metals like copper (Cu), zinc (Zn) and cobalt (Co). Heterologous complementation assays of cDNA PLCe10 showed a range of tolerance to Cu (150 μM–500 μM), Zn (10 mM–12 mM) and Co (2–4 mM). Results of the present study suggest that cDNA PLCe10 is one of the functional eukaryotic heavy metal tolerant genes present among the soil microbial community and could be exploited to rehabilitate metal contaminated sites.

1. Introduction

Soils are a large dump for various hazardous elements such as heavy metals which are released into the soil through numerous anthropogenic

---

* Corresponding author.
E-mail address: msreddy@thapar.edu (M. Sudhakara Reddy).

https://doi.org/10.1016/j.scitotenv.2019.01.201
0048-9697/

activities such as industrial and agricultural activities (Fernández et al., 2018). Heavy metals not only cause pollution by changing the physical and biochemical properties of soil but also represent severe environmental effects and are considered as health hazards (Mitchell et al., 2011; Tchounwou et al., 2012; Esteban-Vasallo et al., 2012). Conditions which affect the soil properties also affect the diversity and functioning of microorganisms living in that environment. Majority of these organisms possess a rich reservoir of genes which are involved in detoxification mechanisms such as chelation, excretion of metals or cellular compartmentalization (Cobbett and Goldsbrough, 2002; Clemens, 2006; Blindauer and Leszczyszyn, 2010). Discovering these genes and the mechanisms in which they are involved will not only elaborate our understanding towards the interaction between microorganisms and their environment but also be implicated in bio-industry or for bioremediation.

Eukaryotic microorganisms present in the soil carry out numerous biological processes with ecologically significance and many of them were unknown to the science as they withstand the laboratory cultivation (Marmeisse et al., 2017). To explore true functional diversity and activities expressed by eukaryotic strata of soil in situ in response to various adverse factors, a metatranscriptomics approach has been practiced (Carvalhais et al., 2012; Marmeisse et al., 2017). Metatranscriptomics involves the construction of eukaryotic-specific environmental cDNA libraries through polyadenylated mRNA and screening them by heterologous complementation in an appropriate microbial host such as *Saccharomyces cerevisiae* (Grant et al., 2006; Bailly et al., 2007; Thakur et al., 2018). The systematic sequencing of these transcriptomes has led to the recognition of full-length genes involved in different biological processes from soils (Damon et al., 2012; Hesse et al., 2015; Žifčáková et al., 2016; Mauchline et al., 2018), animal digestive tracts (Tartar et al., 2009; Qi et al., 2011; Dai et al., 2015), marine waters (Pearson et al., 2015; Zielinski et al., 2016) and extreme aquatic environments (Edgcomb et al., 2016; Puente-Sánchez et al., 2016). Metatranscriptomics approach has also been established to discover various ecological significant genes from different environments involved in numerous phenomena such as metal tolerance (Lehembre et al., 2013), organic matter degradation (Damon et al., 2011), cyanobacterial blooming in marine bodies (Berg et al., 2018), plant-microbe interaction (Crump et al., 2018), root microbiome (Gonzalez et al., 2018), saline tolerance in deserts microbiome (Ren et al., 2018). Studies have also been reported where metatranscriptomics is explored to remediate the contaminated sites such as degradation of polycyclic aromatic compounds like phenanthrene (de Menezes et al., 2012; Suenaga et al., 2007) and in neutralizing the toxic heavy metals present in soil environment (Ziller et al., 2016).

In this study, eukaryotic cDNA libraries constructed from metal contaminated soil was screened to identify the genes involved in metal tolerance by yeast functional complementation assay. Screening of one of the libraries (library C) yielded several genes involved in metal tolerance including von Willebrand factor type D domain (VWD) of vitellogenin (Vg) protein (PLCe10). Vitellogenin is a multimeric egg yolk protein encoded by multiple Vg genes in a number of species that include insects, fish and frog (Tufail and Takeda, 2008) and provide a potential indicator for monitoring and assessing the toxic effect of heavy metals like Cd, Cu, and As (Hwang et al., 2009). Present study aims to characterize the cDNA PLCe10 and test its multi-metal tolerance characteristics by heterologous complementation.

2. Materials and methods

2.1. Collection of soil samples and analysis

Soil samples were collected from Agro-forestry land in Pierrelaye (49°1′45″N, 2°10′32″E) in northwest of Paris, France. This site was previously maize growing crop field which now has been converted into an Agro-forestry site where Poplars are being cultivated. These sites became polluted with heavy metal due to irrigation with untreated wastewater for many years. Twenty soil cores were collected and equal volumes of each core was mixed together to form composite sample after sieving them through 2.0 mm mesh. The soil sample was analyzed for its pH (1:2 soil: water ratio, v/v), organic carbon (Walkley, 1947); total Phosphorus (Kitson and Mellon, 1944); available Phosphorous (Olsen et al., 1954) and total nitrogen (Piper, 1966). Soil samples were digested with HNO_3 and $HClO_4$ (3:1 v/v) and the metal content was determined by using inductively coupled plasma atomic emission spectroscopy (ICP-AES) (ARCOS, Simultaneous ICP Spectrometer, SPECTRO Analytical Instruments GmbH, Germany) (Moor et al., 2001).

2.2. Biological materials

In the present study, Cd-sensitive *ycf1*$^{\Delta}$, copper-sensitive *cup1*$^{\Delta}$ (DTY4), Zn-sensitive *zrc1*$^{\Delta}$ and cobalt-sensitive *cot1*$^{\Delta}$ mutant strains of *S. cerevisiae* were used. The yeast mutant *ycf1*$^{\Delta}$ lacks *YCF1* gene which is an ABC transporter encoding gene responsible for conferring Cd tolerance through the vacuolar transport of Cd conjugates (Li et al., 1997). Similarly, copper sensitive *cup1*$^{\Delta}$ lacks both the copies of copperthionein gene while its isogenic wild type strain DTY3 (*cup1*^{s}) (*MATα*, *leu2-3*, *112his3*$^{\Delta}$*1*, *trp1-1*, *ura3-50*, *gal1*, *CUP1*^{S}) carries single copy of copperthionein gene *cup1* (Longo et al., 1996). In case of *zrc1*$^{\Delta}$, *ZRC1* gene encoding a transporter protein that helps in sequestering of Zn into vacuoles is knocked out (Li and Kaplan, 1998). The hypersensitivity to Co is due to deletion of *COT1* gene that mediates the efflux of cobalt ions into the vacuole (Conklin et al., 1992). *ycf1*$^{\Delta}$, *cot1*$^{\Delta}$ and *zrc1*$^{\Delta}$ strains derived from the wild-type strain BY4741 (*MATα*, *his3*$^{\Delta}$*1*, *leu2Δ0*, *met15*$^{\Delta}$*0*, *ura3Δ0*) were used as a negative control for metal tolerance. All yeast cultures were maintained on Yeast Peptone Dextrose media (YPD) (Hi-media laboratories, India). For all bacterial transformations, One Shot® TOP10 Electrocomp™ *E. coli* (Thermofisher, USA) was used.

2.3. Screening of cDNA library C for cadmium tolerance genes

Total RNA was extracted from soil sample using the RNA PowerSoil® Total RNA Isolation Kit (Mo Bio laboratories, Carlsbad, CA) according to the manufacturer's instructions. cDNA synthesis and size fractionation of cDNA libraries were performed as described earlier (Yadav et al., 2014). Briefly, cDNA libraries were prepared by ligation of cDNAs between SfiIA and SfiIB sites downstream of the PGK1 promoter of *S. cerevisiae* present in yeast expression vector modified pFL61 (Minet et al., 1992). The largest fraction of cDNA library (Lib C) was screened for cadmium tolerant genes by functional complementation assay using *ycf1*$^{\Delta}$ according to the method described in Thakur et al. (2018). For positive transformants, heterologous complementation assay was performed by drop out assay and growth assay. To perform drop out assay, yeast cells were grown in SD-ura broth until $OD_{600} = 1.0$ and serial dilutions were made (0.1–0.0001). Five μl of each dilution was subsequently used to perform drop out assay on SD-ura with and without Cd (40 μM $CdSO_4$) plates. The growth assay was performed to know the range of tolerance of tolerant cDNA by inoculating the respective yeast transformant in 20 ml medium with log phase culture in range of 40–80 μM $CdSO_4$ and growth was recorded after 48 h. One of the transformants, PLCe10 was chosen for further studies, based on its tolerance towards higher Cd concentrations.

2.4. Sequence analysis

Plasmid DNA was extracted from *ycf1*$^{\Delta}$/PLCe10 cells by using Zymoprep™ Yeast Plasmid Miniprep II (Zymo Research, USA) and electro-transformed (3–4 μl) into One Shot® TOP10 Electrocomp™ *E. coli* cells (Thermofisher, USA). The insert was sequenced with vector specific primers NF and NR (Table 1). BLASTX analysis was performed to find homologous sequences for PLCe10 via NCBI web server (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The resulting polypeptide sequences

Table 1
Primers used to amplify different genes in this study.

pFL61 vector specific primers	
pFL61 NF	5′-CAGATCATCAAGGAAGTAATTATCTAC-3′
pFL61 NR	5′-CAGAAAAGCAGGCTGGG AAGC-3′
qPCR primers for yeast (*S. cerevisae*)	
Actin	qACTF: 5′-CGAATTGAGAGTTGCCCCAG-3′ qACTR: 5′-CAAGGACAAAACGGCTTGGA-3′
TAF10	qTAF10F: 5′-GCTAACAACAGTCAGGCGAG-3′ qTAF10R: 5′-GAGCCCGTATTCAGCAACAG-3′
TFC1	qTFC1F: 5′-ACGACGCTGCTTTGGAAAAT-3′ qTFC1R: 5′-GCTTTTCATTGTTTCCGCGG-3′
PLCe10	qvitF: 5′-ACCGTCCAAGTTCAATTGCC-3′ qvitR: 5′-GAATGTCCTGTTCGTCGGTG-3′

were aligned by multiple sequence alignment using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Potential O- and N-linked glycosylation sites were determined using NetOglyc 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc/) and NetNGlyc 1.0 server (http://www.cbs.dtu.dkservices/NetNGlyc/) and phosphorylation sites by NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/). The sequence of PLCe10 cDNA was deposited at NCBI database under the accession number MF770726.

2.5. qPCR analysis and metal uptake

The induction level of PLCe10 in response to Cd stress was quantified by qPCR analysis. The *ycf1*$^{\Delta}$ cells carrying PLCe10 were grown in SD-ura broth supplemented with different concentrations of Cd (0, 40, 60 and 80 μM) for 48 h at 30 °C. Cells were centrifuged at 3000 rpm for 5 min and the pellet was ground with liquid nitrogen. Total RNA was extracted by using Trizol reagent followed by use of RNase-free DNase I (Thermo Scientific, USA) to remove genomic DNA contamination. The purity and integrity of RNA was determined by NanoDrop™ 1000 (Thermo Scientific, USA) and formaldehyde agarose gel electrophoresis. First strand cDNA was synthesized from total RNA by using PrimeScript™ first strand cDNA synthesis kit (Takara, Japan) as per the instruction manual. Gene expression analysis of Cd tolerant cDNA PLCe10 was performed by using SYBR® Premix Ex Taq™ (Tli RNase H Plus) (Takara, Japan) as directed in user's manual in a Mastercycler® ep realplex system (Eppendorf AG, Hamburg, Germany) with a final volume of 20 μl (100 times diluted RT cDNA 6.8 μl, 0.2 μM forward primer, 0.2 μM reverse primer (Table 1), 2× SYBR Green mix 10 μl and H_2O for final volume make up). Amplification programme used was: 95 °C for 2 min (1 cycle), 95 °C for 15 s, 53 °C for 15 s and 68 °C for 20 s (40 cycles). The amplification efficiency (E) was calculated in accordance to the equation $E = [10^{(-1/\mathrm{slope})}] - 1$ and efficiency values lies between 0.94 and 1.27. These efficiency values were further used to compute Ct_1 values by using equation $Ct_1 = C_{te}x\ [\log(1 + E) / \log 2]$. The primers (Table 1) were designed for reference genes from *ACT* (Actin; SGD: S000001855), *TAF10* (TATA binding protein-associated factor; SGD: S000002574) and *TFC1* (Transcription factor class C; SGD: S000000327) of *S. cerevisiae* (Teste et al., 2009). Analysis via Normfinder suggested that most stable expression and minimum stability value under Cd stress recorded was for *TFC1*. Hence, *TFC1* was used as reference genes to calculate comparative expression level of genes by using the formula $2^{-\Delta\Delta C_T}$ (Livak and Schmittgen, 2001). The $-\Delta\Delta C_T$ is calculated by subtracting the normalised Ct value for test data from normalised Ct value for corresponding control. For every measurement, RNA was extracted from three independent biological samples (experimental replicates) and for each experimental replicate, three technical replicates were also performed.

The Cd accumulation by *ycf1*$^{\Delta}$/PLCe10 cells were determined by ICP-MS (Element XR, Thermo Fisher Scientific, Germany) analysis by growing these cells at different concentrations of Cd (40 μM to 80 μM $CdSO_4$). Cells were harvested and acid digested 1 g of cell pellet with HNO_3 and $HClO_4$ (3:1 v/v). Acid digested samples (final volume of 50 ml with H_2O) were analyzed by ICP-MS. The wild type strain BY4741 transformed with empty vector was used as positive control.

2.6. Tolerance of PLCe10 for different metals

The plasmid pFL61 containing cDNA PLCe10 was transformed into yeast mutants hypersensitive to specific heavy metal like Cu (*cup1*$^{\Delta}$), Zn (*zrc1*$^{\Delta}$) and Co (*cot1*$^{\Delta}$) and their respective wild type strains with empty vector. The transformants were regenerated on SD-ura media containing specific metal i.e. Cu 150 μM, Zn 10 mM and Co 2 mM. In parallel experiments for studying growth under the influence of metal, each flasks containing 20 ml of SD-ura medium with different range of metal concentrations of Cu (150 μM–500 μM), Zn (10 mM–12 mM) and Co (2 mM–4 mM) were inoculated with transformed *cup1*$^{\Delta}$, *cot1*$^{\Delta}$, *zrc1*$^{\Delta}$ cells and incubated at 30 °C at 220 rpm for 48 h. In a similar way, yeast transformant cultures (*cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$) were grown in 50 ml of SD-ura supplemented with range of different metals i.e. Cu (150–500 μM), Zn (10–12 mM) and Co (2–4 mM) and the uptake of metals were determined by ICP-MS as described above.

3. Results and discussion

3.1. Soil analysis

The soil collected from Pierrelaye region is sandy luvisol with a pH of 7.1. The other chemical properties of the soil were: organic carbon 1.6%; total P 291 mg/kg; available P 14.2 mg/kg and total nitrogen 0.12%. The Cd content was 2.5 mg/kg, Cu 64 mg/kg and Zn 385 mg/kg of the soil.

3.2. Cadmium tolerance genes

One of the main challenges in cDNA library construction is to obtain a high number of clones having inserts more than 1.0 kb size (Yadav et al., 2014). In the present study, the cDNA quantities were slightly increased by amplifying through three PCR cycles without affecting significantly their average length. Then the cDNAs were separated by bidimensional agarose gel electrophoresis to separate them and minimize cross-contamination of size fractions. Three different size fractions such as fraction A (0.1–0.5 kb), fraction B (0.5–1 kb) and fraction C (1–4 kb) were recovered and the respective cDNA libraries were constructed as described in Yadav et al. (2014). Prior to library construction, the fractions were tested for the absence of cross contaminations by performing PCR amplification of four eukaryotic genes of different length classes (Yadav et al., 2014). In the present study, library C (1–4 kb) was screened for presence of metal tolerant gene present in soil environment. The transformants from cDNA library C were selected based on their ability to rescue the Cd-hypersensitive mutant strain *ycf1*$^{\Delta}$. One of the transformants, *ycf1*$^{\Delta}$PLCe10, which showed high tolerance to Cd while performing drop test, was selected for further studies (Fig. 1a). The *ycf1*$^{\Delta}$PLCe10 showed significantly higher growth in all the concentrations compared to the wild type strain transformed with empty vector. With increase in Cd concentration, reduction in the growth was observed both in *ycf1*$^{\Delta}$PLCe10 and wild type strain (Fig. 1b). The metabolic and ecological roles of microorganisms present in various adverse environments including metal contaminated site could be studied by metatranscriptomics approach. Lehembre et al. (2013) reported different categories of genes such as metallothioneins (MTs), cysteine rich protein (CRPs) responsible for Cd tolerance and *BOLA* genes responsible for Zn, Mn and Co tolerance by using soil metatranscriptomics. While exploring metatranscriptomics for novel genes, Ziller et al. (2016) reported metal binding properties of five new CRPs (environmental MTs) conferring tolerance to heavy metals like Zn and Cd. Similarly, Thakur et al. (2018) isolated environmental transcript PLBe1 encoding an ubiquitin fusion protein responsible for Cd tolerance.

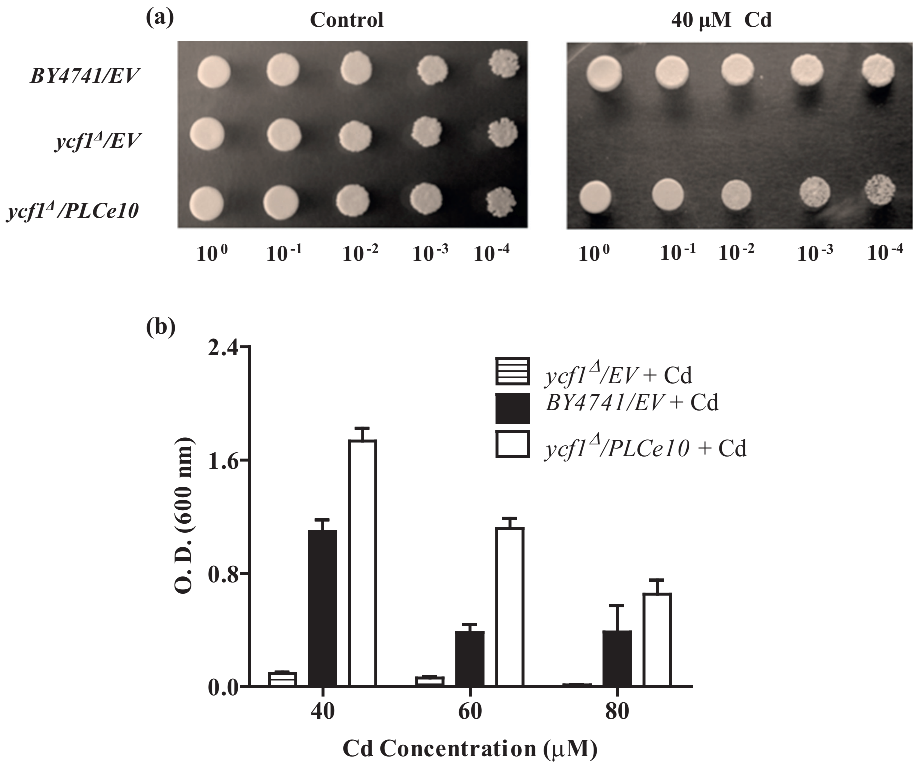


Fig. 1. Functional complementation of PLCe10 (a) Drop assay of mutant strain *ycf1*$^{\Delta}$ expressing PLCe10 on Cd 40 μM. (b) Growth assay of mutant strain *ycf1*$^{\Delta}$ expressing PLCe10 in SD-ura broth with different concentrations of Cd. Wild type BY4741 and *ycf1*$^{\Delta}$ transformed with empty vector pFL61 (EV) were used as controls. Errors bars are ±SD.

3.3. Sequence analysis

The BLASTX analysis with the deduced nucleotide sequences of PLCe10 revealed 30% identity (50% positives) to von Willebrand factor type D (VWD) of vitellogenin-6 of nematode *Ascaris suum* with the open reading frame of 1017 bp, encoding 338 amino acids. The predicted molecular mass of peptide was 38.96 kDa (average mass) with pI of 4.77. The deduced peptide showed homology with amino acid sequence of *A. suum* from 1428 to 1719 amino acids denoted C-terminal of the Vg-6 which is an egg yolk protein. It had been reported that in *Paracylopina nana*, C-terminal region of the Vg1 from amino acid sequence 1571–1739 and the Vg2 i.e. from 1481 to 1658 amino acids codes for von Willebrand factor type D domain (VWD). Hwang et al. (2010) reported that the BLAST analysis with the inferred amino acid sequences of *Paracyclopina nana* Vg1 and Vg2 had shown that in amino acid residues strongly matched with the intertidal copepod Vg1 (35% identities/56% positives) and *Lepeophtheirus salmonis* Vg2 (31% identities/52% positives), respectively. Multiple sequence alignment of homologous sequences showed that the 1017 bp gene belongs to protein family smart00216 which is also named as VWD. The deduced polypeptide identified as VWD like protein, contained characteristic phosphorylation sites at 32 amino acid sequences and 14 potential motifs for O-linked glycosylation sites were also found while no N-linked glycosylation sites (Fig. 2). Finn (2007) observed that before exporting Vg from hepatocytes, phosphorylation of precursor protein takes place in the Golgi bodies and the presence of phosphorylated serine residues in the Vg polypeptide are assumed to avert premature degradation and assist the solubility of Vg molecules in the blood. Various phosphorylation sites within the predicted VWD like polypeptide sequence indicate that the wide majority of serine residues present in the polypeptide of VWD like domain may be phosphorylated along with threonine and tyrosine. In addition to this, comparison of VWD like polypeptide to other vitellogenin showed a high degree of conservation; contain two RXXR consensus cleavage sites each at C-terminal and N-terminal of the polypeptide, a well characterised the GL/ICG motif and conserved polycysteine residues at C-terminal (Fig. 3).

Tufail et al. (2010) characterised Vg from insect *Nilaparvata lugens* and found high degree of conservation for the motif GL/ICG. In a study, Shu et al. (2009) aligned Vg of *Spodoptera litura* with Vg of other species and characterised it by presence of highly conserved cleavage signal RXXR at N terminal, C-terminal the GL/ICG motif followed by cysteine residues. Similarly in a previous research, it was concluded that in C-terminal of Vg, VWD contain a signature amino acid sequence GL/ICG followed by nine cysteine residues (Tufail and Takeda, 2008). Salmela et al. (2016) characterised three Vg like proteins and found that all of them share structural similarities with Vg but each of them also has unique gene/protein architecture and distinct pattern of positive selection, which are indicative of divergence of functions. Presence of hexameric polyadenylation signal (PAS) AATAAA at 12 bp upstream the poly-A tail in ORF PLCe10, also characterised this cDNA as ORF with a potential for coding a protein as described by Beaudoing et al. (2000). From all the findings of bioinformatic analysis, we assumed that the cDNA sequence PLCe10 conferring tolerance to Cd metal obtained in this study encode for VWD like domain of putative protein Vg. Jubeaux et al. (2012) clearly demonstrated the benefit of moderate phylogenetic conservation of Vg like protein such as Vg proteotypic peptides, to directly exploit them in developing biomarkers to identify more or less close phylogenetically related un-sequenced species.

3.4. qPCR analysis and metal uptake

Induction of cDNA PLCe10 encoding VWD like protein in presence of Cd metal was studied. Induction levels of PLCe10 increased significantly due to Cd stress and maximum mRNA accumulation was observed at 60 μM of $CdSO_4$, which was 32 fold higher than control (Fig. 4a). Similar findings were reported for induction of metallothioneins in fungi on exposure to heavy metals and observed that expression of different

```
PLCe10      CGLCGHFDLESTDEFRNPDFTDEQDIRQFYMNYLIKDGQCKAPQQLTEVCESEECDKADRSSSSSSSS-S----------S
KHN77828    CGICGHYDGEKLDELRAADNGLTDDVEEYSRSYFDKGDECNIEEDIVSEKNSYRFDDDID---------ENFDIDDVEEE-
ERG83573    CGICGHYDGEKQYELRMADNELTDDLEQYSRSYFNNDEECDIEENIVKEKMNYRLEDDSN---------EFFGEDEDEEH-
CCD65563    CGLCGNNDDESTNEFYTSDNTETEDIEEFHRSYLLKNEECEAEEERLSEKKNYRKYERDEEQSDEYSSEETYDYE--QENT
ABV46673    CGLCGHFDEDKDSEFYTPNKEYTSDVEEFHKSYLLT-DKCEVEKEFSFEKKDYAVETNEER--SDDWL-STYDDDNTSND-
CDJ93502    CGLCGHFDEEKDNEFYTPKKEYTDDIMEFHKSYLLN-DECEVEKELPREKKHYRLEKSESSSSEEDWL-DFYENDDKRKEM
KHJ90957    CGLCGHYDEEKDNEFLTPKMERTTDIMEFHKSYLLD-EECEMEKELINEKKHYKLEEREH---DDDFL-DIFFEEEDIDT-

PLCe10      SGSQESQSNEESEIPEKKTKVIEIDDQLCFSTVPIPQCDEEDSYPMGEKEKRKVAYVCIDQDSQEAEDIERQARSGRRQIP
KHN77828    -------EEREIKKPILRTKIIEHNYEVCFSMKAVPHCPHKTYARNEMKTRKKVPFVCLPRTNHEATNFINRA---LHEVL
ERG83573    -------TESNIKKPILKTKIIEYNNEVCFSMKPVKECPQQTYAREDNKTQRKVPFACLPRSTYETTKLLNRA---RFEVL
CCD65563    KKSQKNQRSQKKSDLVEKTQIKEFSHRICFSVEPVAECRRGYE--VEQQQQRKIRFTCLQRHNRDASRLLKES---RQQPL
ABV46673    -------LKDFDEEPLKTTHVMEFPHRVCFSLEPVRKCRKNEK--MDDMVEKKVRFTCLPRSSHETRQLLHKA---RTSVL
CDJ93502    RRTYKYRAESEEEEILEKHHVIEYPHQVCFSLEPVRVCRRNEV--KGDTMDKKVRFTCLPRSSREARELLHKV---RKNVV
KHJ90957    -----KNKREKDDEILKRNHVMEFSERVCFSREPVRVCRKNEE--MDEVVDMKVRFTCLPRTSHEARQLLHKV---RNNVL

PLCe10      ALNNRTPSFTRTEKIPEKCKKYSRN
KHN77828    DVNDYKKSFVEEVTVPDSCRAF---
ERG83573    NMNKYKASFTEEVSIPETCRTF---
CCD65563    QLDDYPVSFVESVKVPTACVAY---
ABV46673    ELNDYPISFVENLRVPLHALSTKLM
CDJ93502    DLSRHPISFVETIEVPRTCTVY---
KHJ90957    NLHDYPVSFVETIQVPRTCVVY---
```

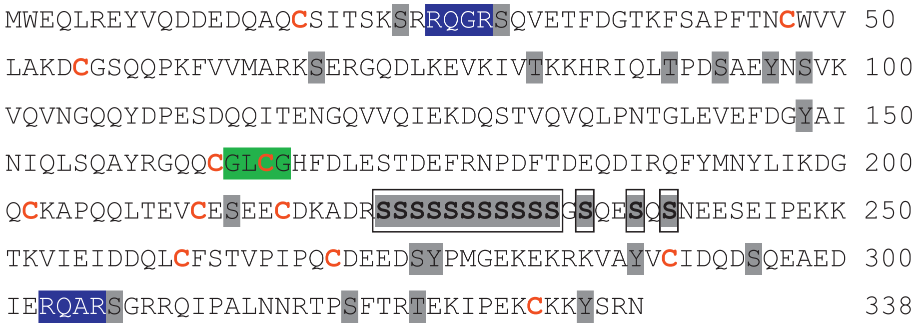
Fig. 2. Alignment of amino acid sequences of different organisms showing highly conserved GL/ICG motif (red highlighted) to the C-terminal ends. Accession numbers are in parenthesis *Toxocara canis* (KHN77828), *Ascaris suum* (ERG83573), *Caenorhabditis elegans* (CCD65563), *Angiostrongylus vasorum* (ABV46673), *Haemonchus contortus* (CDJ93502) and *Oesophagostomum dentatum* (KHJ90957). Cysteine residues are highlighted with green colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

metallothioneins helps in providing tolerance towards different heavy metals (Ramesh et al., 2009; Reddy et al., 2016; Kalsotra et al., 2018). The *ycf1*$^{\Delta}$PLCe10 cells accumulated different levels of Cd when grown in presence of various Cd concentrations. It has been observed that the amount of Cd metal accumulated increased up to 60 μM of Cd concentration and decreased at 80 μM Cd (Fig. 4b).

3.5. Multi-metal tolerance of PLCe10

The cDNA PLCe10 was tested for its tolerance to other heavy metals such as Cu, Zn and Co. As shown in Fig. 5a, Cu sensitive mutant *cup1*$^{\Delta}$ with empty pFL61 was sensitive to 150 μM $CuSO_4$ while *cup1*$^{\Delta}$ cells carrying plasmid PLCe10 (*cup1*$^{\Delta}$PLCe10) were showing tolerance to a recognisable level in comparison to *cup1*^{s}pFL61 which carries single copy of *CUP1* gene only. Likewise, *zrc1*$^{\Delta}$ cells carrying PLCe10 plasmid were tolerant at 10 mM $ZnCl_2$ whereas this concentration was toxic for *zrc1*$^{\Delta}$pFL61 as shown in Fig. 5b. Similarly, 2 mM $CoCl_2$ inhibited the growth of *cot1*$^{\Delta}$pFL61 where as *cot1*$^{\Delta}$ cells carrying plasmid PLCe10 (*cot1*$^{\Delta}$PLCe10) were showing tolerance to Co as in case of wild strain BY4741pFL61 (Fig. 5c).

The growth assays of different transformants were showed positive results on different range of concentration of metals. When screened for Cu, it was observed that *cup1*$^{\Delta}$PLCe10 cells were able to grow up to 500 μM compared to mutant strain *cup1*$^{\Delta}$pFL61 and wild type strain *cup1*-^{s}pFL61 (Fig. 6a). In context to Cu, it was reported that estrogen induced vitellogenin was tolerant to oxidation reactions induced by Cu and chelate low concentrations of Cu ion. Antioxidant property of vitellogenin protected the copper-stimulated oxidation of very low density lipoprotein (VLDL). Vitellogenin believed to be served as transition metals-binding lipoprotein which results into extensive depressing of free-radical reactions in the oocytes (Ando and Yanagida, 1999). While studying tolerance for Zn metal tolerance, it was found that transformant has shown more tolerance towards Zn (10 mM to 12 mM $ZnSO_4$) in comparison to both wild strain BY4741pFL61 and mutant strain *zrc1*$^{\Delta}$pFL61 (Fig. 6b). Vitellogenin mediates the transfer of Zn from the liver to the maturing oocytes in the turkey hen, ultimately resulting in Zn deposition into yolk of the newly formed egg (Richards, 1989). It was also reported that Vg is a macromolecule phospholipid glycoprotein containing Ca and Zn ligands (Wallace, 1985; Montorzi et al., 1994; Denslow et al., 1999). Furthermore, it was reported that among honeybee workers and queens, Vg plays role as a Zn carrier that protects them from oxidative stress (Amdam et al., 2004; Seehuus et al., 2006) and contributes to queen longevity (Corona et al., 2007). Similarly in cobalt, the gene provided tolerance



```
MWEQLREYVQDDEDQAQCSITSKSRRQGRSQVETFDGTKFSAPFTNCWVV  50
LAKDCGSQQPKFVVMARKSERGQDLKEVKIVTKKHRIQLTPDSAEYNSVK 100
VQVNGQQYDPESDQQITENGQVVQIEKDQSTVQVQLPNTGLEVEFDGYAI 150
NIQLSQAYRGQQCGLCGHFDLESTDEFRNPDFTDEQDIRQFYMNYLIKDG 200
QCKAPQQLTEVCESEECDKADRSSSSSSSSSSSGSQESQSNEESEIPEKK 250
TKVIEIDDQLCFSTVPIPQCDEEDSYPMGEKEKRKVAYVCIDQDSQEAED 300
IERQARSGRRQIPALNNRTPSFTRTEKIPEKCKKYSRN             338
```

Fig. 3. Deduced amino acid sequence of VWD protein: Highly conserved GL/ICG motif (highlighted in green) to the C-terminal end. Cysteine residues are given in bold red colour while RXXR cleavage sites at the N-terminal end are highlighted in blue with white font. Bold and boxed amino acids are sites for O-linked glycosylation and grey highlighted are phosphorylation sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

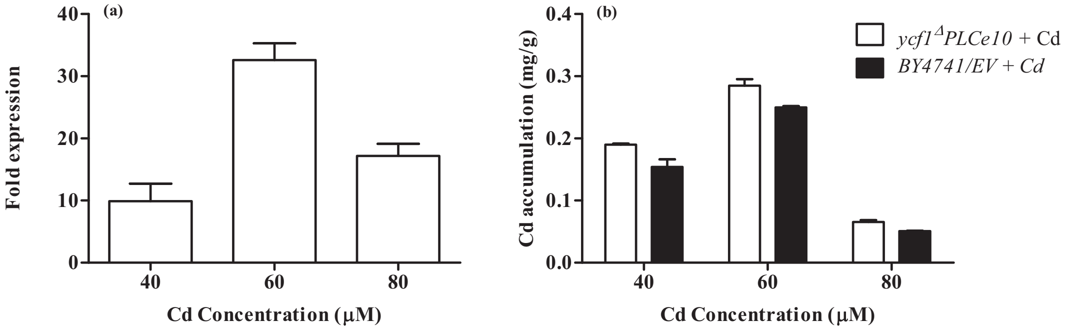


Fig. 4. a) Fold increase in expression levels of cDNA PLCe10 in *ycf1*$^{\Delta}$ after 48 h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$PLCe10 without metal treatment), b) Accumulation of Cd in *ycf1*$^{\Delta}$ carrying cDNA PLCe10 and BY4741/EV. EV stands for empty vector pFL61. Error bars are ±SD.

to mutant yeast from range of 2 mM to 4 mM. Increase in tolerance was observed with the increase in range of Co concentrations (Fig. 6c). Henceforth, this study showed that gene PLCe10 was responsible for providing tolerance to all the four metals. Concurrently, while estimating the accumulated metal by yeast mutants transformed with cDNA PLCe10, it has been observed that cDNA PLCe10 is capable of accumulating a higher range of metals (Cu, Zn and Co) in comparison to wild strain BY4741 transformed with empty vector pFL61. For Cu, the level of metal accumulation increased with increase in concentration as shown in Fig. 6d, while for Zn and Co metal, the level of accumulation increases up to 11 mM of $ZnCl_2$ and 3 mM $CoCl_2$ respectively and decreases with further increase in Zn or Co concentration (Fig. 6e, f). Such variation in order of tolerance and accumulation of heavy metals could potentially be an explanation for distinct metal-binding abilities of the deduced VWD like protein. In general, increased growth and metal accumulation was observed in mutant yeast cells carrying PLCe10 compared to the respective wild type strains. The increase in growth might be due to protection of PLCe10 to the yeast mutants under metal stress and the accumulation of metal in the yeast strains also might to due increased growth.

In heavy metal contaminated aquatic environment, induction of Vg among aquatic animals on exposure to heavy metals and endocrine disruption chemicals has been used as a biomarker (Hwang et al., 2010; Lee-Pow et al., 2016). As on bioinformatics analysis of VWD like protein,

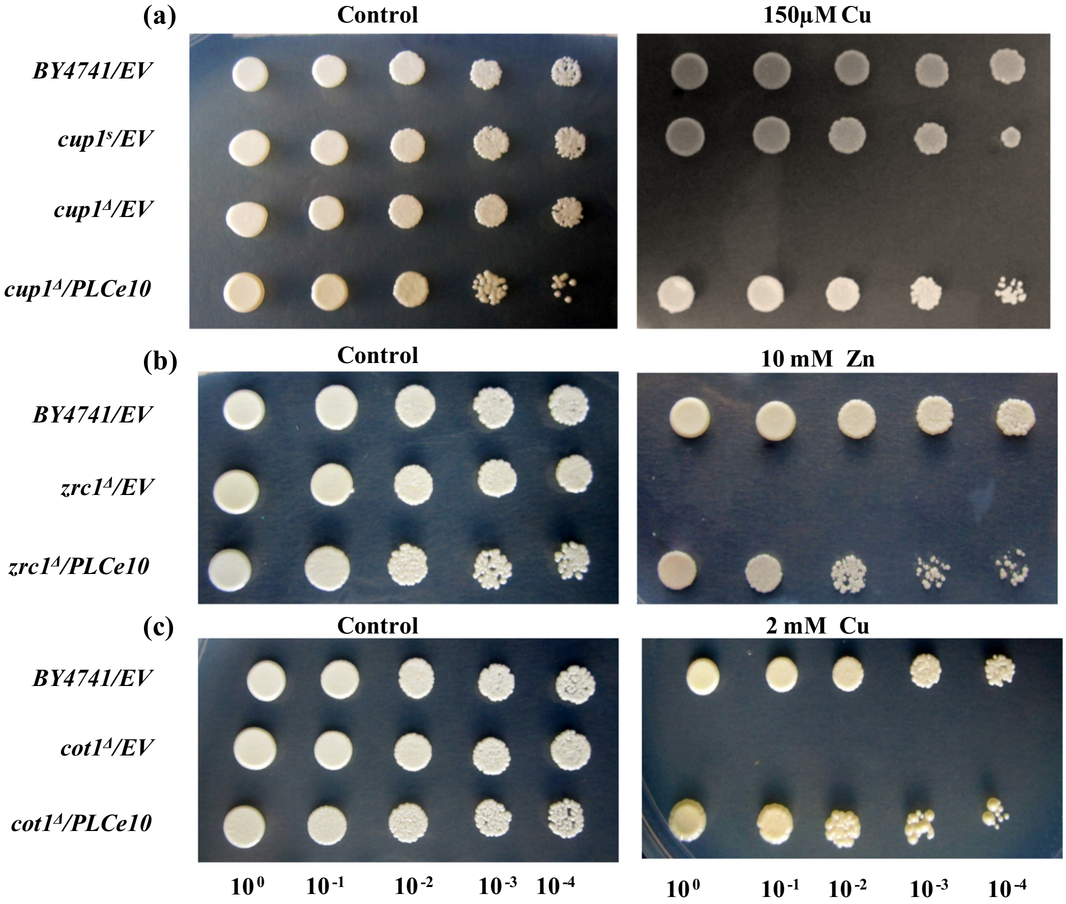


Fig. 5. Drop assay of metal sensitive mutant strains *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLCe10 at various concentrations a) Cu 150 μM, b) Zn 10 mM and c) Co 2 mM. Wild strain BY4741 and *cup1*^{s} are used as control and EV denotes to empty vector pFL61.

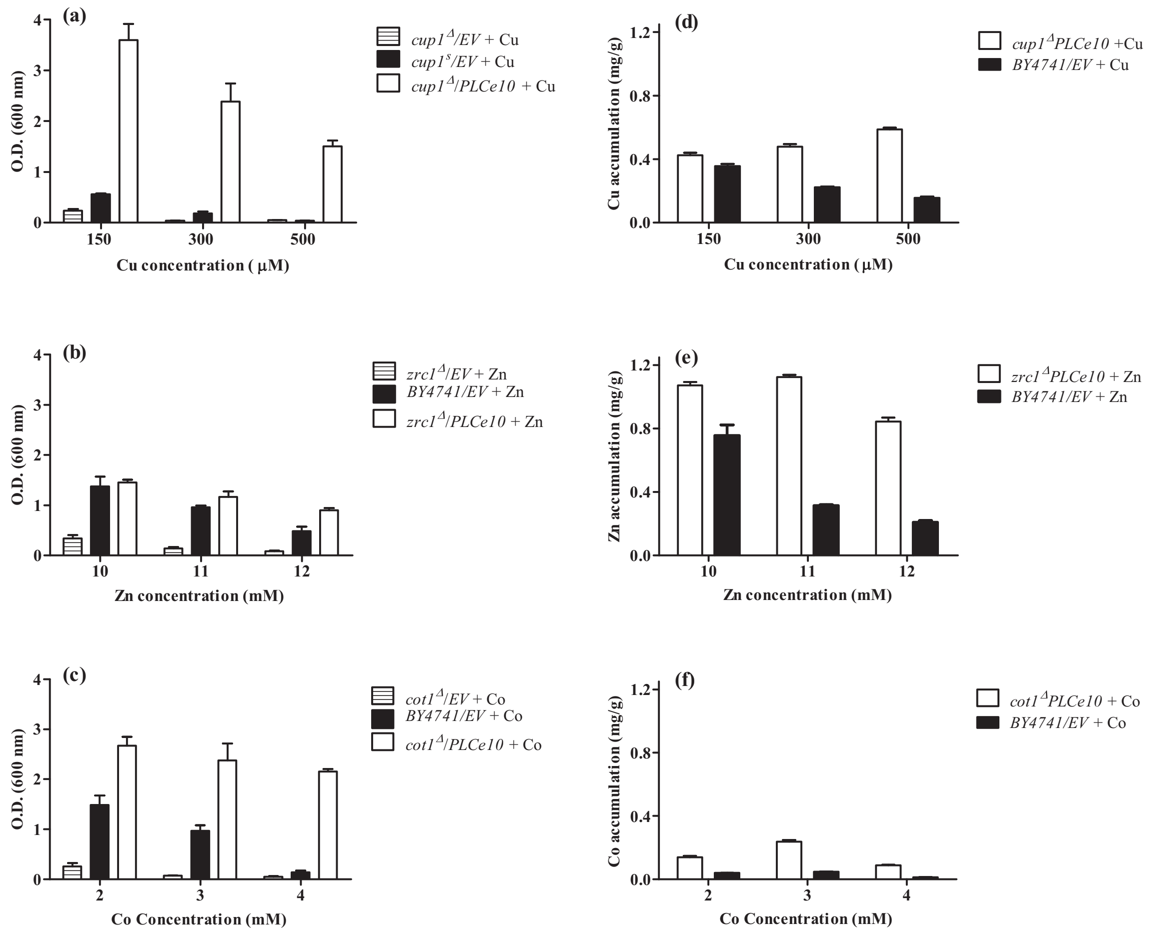


Fig. 6. Growth assay of metal sensitive mutant strains $cup1^{\Delta}$, $zrc1^{\Delta}$ and $cot1^{\Delta}$ expressing PLCe10 in SD-ura broth with metal supplementation for 48 h. a) Cu 150–500 μM, b) Zn 10–12 mM and c) Co 2–4 mM. Accumulation of d) Cu, e) Zn and f) Co in mutant yeast cells $cup1^{\Delta}$, $zrc1^{\Delta}$ and $cot1^{\Delta}$ carrying cDNA PLCe10 and BY4741/EV. Wild type BY4741 and mutant strain transformed with only empty vector pFL61 (EV) were used as controls. Error bars are ±SD.

it was found that the cDNA screened has shown GL/ICG conserved domain (Fig. 2). The presence of GL/ICG motif followed by number of cysteine residues at conserved locations was entitled as a general trait among various invertebrate and vertebrate species (Lee et al., 2000). Consequently, there could be a possibility to gain benefit of this peptide motif conservation to develop biomarkers applicable for groups of phylogenetically related species as these days mass spectrophotometry-based procedures necessitate small peptides i.e. from 5 to 20 amino acids. Hence, these findings reported VWD like protein as a potential member of metal tolerant gene family of eukaryotic community which increases the possibility of using such peptides as biomarker for metal contamination.

4. Conclusions

The present study results conclude that the eukaryotic transcript PLCe10 have potential to tolerate the toxicity of different metals in the environment. Functional metatranscriptomic approach is suitable to characterize novel genes like VWD like protein whose over-expression restored tolerance to all the four heavy metals in hypersensitive yeast mutants. This study strongly supports the existence of some unknown mechanisms and biomolecules that play vital role in adaptation of biodiversity of an ecosystem against metal contaminated environment. Metatranscriptomics approach has major contribution in providing information about various unknown phenomena occurring in soil and revealed high degree of diversity. Most of the work on vitellogenin is contributed to aquatic oviparous vertebrates and invertebrates, while in present study VWD like protein screened through metatranscriptomic cDNA library of metal polluted soil environment is novel finding of this approach. In this respect, VWD like protein could be used to probe gene related processes and to assess their involvement in the overall capacity of organisms to survive in heavy metal enriched environment. This gene could be developed as biomarker to determine heavy metals in the environment principally in soil and to exploit the levels of tolerance or resistance among the species of concern for the bioremediation of polluted lands.

Acknowledgements

Authors are thankful to Indo-French Centre for the Promotion of Advanced Research, New Delhi for granting the research project (4709-1) titled "Gene resources from polluted soils".

References

Amdam, G.V., Norberg, K., Fondrk, M.K., Page, R.E., 2004. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. Proc. Natl. Acad. Sci. U. S. A. 101, 11350–11355.

Ando, S., Yanagida, K., 1999. Susceptibility to oxidation of copper induced plasma lipoproteins from Japanese eel, protective effect of vitellogenin on the oxidation of very low density lipoprotein. Comp. Biochem. Physiol. 123, 1–7.

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

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., Gautheret, D., 2000. Patterns of variant polyadenylation signal usage in human genes. Genome Res. 10, 1001–1010.

Berg, C., Dupont, C.L., Asplund-Samuelsson, J., Celepli, N.A., Eiler, A., Allen, A.E., Ekman, M., Bergman, B., Ininbergs, K., 2018. Dissection of microbial community functions during a cyanobacterial bloom in the Baltic Sea via metatranscriptomics. Front. Mar. Sci. 5, 55.

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.

Carvalhais, L.C., Dennis, P.G., Tyson, G.W., Schenk, P.M., 2012. Application of metatranscriptomics to soil environments. J. Microbiol. Methods 91, 246–251.

Clemens, S., 2006. Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. Biochimie 88, 1707–1719.

Cobbett, C., Goldsbrough, P., 2002. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. Annu. Rev. Plant Biol. 53, 159–182.

Conklin, D.S., McMaster, J.A., Culbertson, M.R., Kung, C., 1992. COT1, a gene involved in cobalt accumulation in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12, 3678–3688.

Corona, M., Velarde, R.A., Remolina, S., Moran-Lauter, A., Wang, Y., Hughes, K.A., Robinson, G.E., 2007. Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. Proc. Natl. Acad. Sci. U. S. A. 104, 7128–7133.

Crump, B.C., Wojahn, J.M., Tomas, F., Mueller, R.S., 2018. Metatranscriptomics and amplicon sequencing reveal mutualisms in seagrass microbiomes. Front. Microbiol. 9, 388.

Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., Wang, J., 2015. Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. Appl. Environ. Microbiol. 81, 1375–1386.

Damon, C., Vallon, L., Zimmermann, S., Haider, M.Z., Galeote, V., Dequin, S., Marmeisse, R., 2011. A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. ISME J. 5, 1871–1880.

Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L., Marmeisse, R., 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. PLoS One 7, e28967.

de Menezes, A., Clipson, N., Doyle, E., 2012. Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. Environ. Microbiol. 14, 2577–2588.

Denslow, N.D., Chow, M.C., Kroll, K.J., Green, L., 1999. Vitellogenin as a biomarker of exposure for estrogen or estrogen mimics. Ecotoxicology 8, 385–398.

Edgcomb, V.P., Pachiadaki, M.G., Mara, P., Kormas, K.A., Leadbetter, E.R., Bernhard, J.M., 2016. Gene expression profiling of microbial activities and interactions in sediments under haloclines of E. Mediterranean deep hypersaline anoxic basins. ISME J. 10, 2643–2657.

Esteban-Vasallo, M.D., Aragonés, N., Pollan, M., López- Abente G, Perez-Gomez B., 2012. Mercury, cadmium and lead levels in human placenta: a systematic review. Environ. Health Perspect. 120, 1369–1377.

Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J., Ordóñez, C., 2018. Geographically Weighted Principal Components Analysis to assess diffuse pollution sources of soil heavy metal: application to rough mountain areas in Northwest Spain. Geoderma 311, 120–129.

Finn, R.N., 2007. Vertebrate yolk complexes and the functional implications of phosvitins and other subdomains in vitellogenins. Biol. Reprod. 76, 926–935.

Gonzalez, E., Pitre, F.E., Pagé, A.P., Marleau, J., Nissim, W.G., St-Arnaud, M., Labrecque, M., Joly, S., Yergeau, E., Brereton, N.J.B., 2018. Trees, fungi and bacteria: tripartite metatranscriptomics of a root microbiome responding to soil contamination. Microbiome 6, 53.

Grant, S., Grant, W.D., Cowan, D.A., Jones, B.E., Ma, Y., Ventosa, A., Heaphy, S., 2006. Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. Appl. Environ. Microbiol. 72, 135–143.

Hesse, C.N., Mueller, R.C., Vuyisich, M., Gallegos-Graves, L.V., Gleasner, C.D., Zak, D.R., Kuske, C.R., 2015. Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. Front. Microbiol. 6, 337–406.

Hwang, D.S., Lee, K.W., Lee, J.S., 2009. Cloning and expression of vitellogenin 2 gene from the intertidal copepod *Tigriopus japonicus*. Ann. N. Y. Acad. Sci. 1163, 417–420.

Hwang, D.S., Lee, K.W., Han, J., Park, H.G., Lee, J., Lee, Y.M., Lee, J.S., 2010. Molecular characterization and expression of vitellogenin (Vg) genes from the cyclopoid copepod, *Paracyclopina nana* exposed to heavy metals. Comp. Biochem. Physiol. 151, 360–368.

Jubeaux, G., Audourd-Combe, F., Simon, R., Tutundijn, R., Salvador, A., Geffard, O., Chaumot, A., 2012. Vitellogenin- like proteins among invertebrate species diversity: potential of proteomic mass spectrometry for biomarker development. Environ. Sci. Technol. 46, 6315–6323.

Kalsotra, T., Khullar, S., Agnihotri, R., 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.

Kitson, R.E., Mellon, M.G., 1944. Colorimetric determination of phosphorus as molybdovanadophosphoric acid. Ind. Eng. Chem. 16, 379–383.

Lee, J.M., Nishimori, Y., Hatakeyama, M., Bae, T.W., Oishi, K., 2000. Vitellogenin of the cicada *Graptopsaltria nigrofuscata* (Homoptera), analysis of its primary structure. Insect Biochem. Mol. Biol. 30, 1–7.

Lee-Pow, C.S., Yost, E.E., Aday, D.D., Kullman, S.W., 2016. Sharing the roles: an assessment of Japanese medaka estrogen receptors in vitellogenin induction. Environ. Sci. Technol. 50, 8886–8895.

Lehembre, F., Doillon, D., David, E., Perrotto, S., Baude, J., Foulon, J., Harfouche, L., Vallon, L., Poulain, J., Da Silva, C., Wincker, P., Oger-Desfeux, C., Richaud, P., Colpaert, V., Chalot, M., Fraissinet-Tachet, L., Blaudez, D., Marmeisse, R., 2013. Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. Environ. Microbiol. 15, 2829–2840.

Li, L., Kaplan, J., 1998. Defects in the yeast high affinity iron transport system result in increased metal sensitivity because of the increased expression of transporters with a broad transition metal specificity. J. Biol. Chem. 273, 22181–22187.

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*, YCF1Δ-catalyzed transport of bis(glutathionato) cadmium. Proc. Natl. Acad. Sci. U. S. A. 94, 42–47.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. Methods 25, 402–408.

Longo, V.D., Gralla, E.B., Valentine, J.S., 1996. Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae*. J. Biol. Chem. 271, 12275–12280.

Marmeisse, R., Kellner, H., Fraissinet-Tachet, L., Luis, P., 2017. Discovering protein-coding genes from the environment: time for the eukaryotes? Trends Biotechnol. 35, 824–835.

Mauchline, T.H., Hayat, R., Clark, I.M., Hirsch, P.R., 2018. Old meets new: most probable number validation of metagenomic and metatranscriptomic datasets in soil. Lett. Appl. Microbiol. 66, 14–18.

Minet, M., Dufour, M.E., Lacroute, F., 1992. Complementation of Saccharomyces cerevisiae auxotrophic mutants by Arabidopsis thaliana cDNAs. Curr. Genet. 2, 417–422.

Mitchell, E., Frisbie, S., Sarkar, B., 2011. Exposure to multiple metals from groundwater-a global crisis: geology, climate. Metallomics 3, 874–908.

Montorzi, M., Falchuk, K.H., Vallee, B.L., 1994. *Xenopus laevis* vitellogenin is a zinc protein. Biochem. Biophys. Res. Commun. 200, 1407–1413.

Moor, C., Lymberopoulou, T., Dietrich, V.J., 2001. Determination of heavy metals in soils, sediments and geological materials by ICP-AES and ICP-MS. Microchim. Acta 136, 123–128.

Olsen, S.R., Cole, C.V., Watanabe, F.S., Dean, L.A., 1954. Estimation of Available Phosphorus in Soils by Extraction With Sodium Bicarbonate. U.S. Government printing office. US Department of Agriculture, Washington (DC) (Circular No. 939).

Pearson, G.A., Lago-Leston, A., Cánovas, F., Cox, C.J., Verret, F., Lasternas, S., Serrão, E.A., 2015. Metatranscriptomes reveal functional variation in diatom communities from the Antarctic Peninsula. ISME J. 9, 2275–2289.

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.

Puente-Sánchez, F., Olsson, S., Gómez-Rodriguez, M., Souza-Egipsy, V., Altamirano-Jeschke, M., Amils, R., Aguilera, A., 2016. Solar radiation stress in natural acidophilic biofilms of Euglena mutabilis revealed by metatranscriptomics and PAM fluorometry. Protist 167, 67–81.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., Forster, R.J., 2011. Snapshot of the eukaryotic gene expression in muskoxen rumen - a metatranscriptomic approach. PLoS One 6, e20521.

Ramesh, G., Podila, G.K., Gay, G., Marmeisse, R., Reddy, M.S., 2009. Different patterns of regulation for the copper and cadmium metallothioneins of the ectomycorrhizal fungus *Hebeloma cylindrosporum*. Appl. Environ. Microbiol. 75, 2266–2274.

Reddy, M.S., Kour, M., Aggarwal, S., Ahuja, S., Marmeisse, R., Fraissinet-Tachet, L., 2016. Metal induction of a *Pisolithus albus* metallothionein and its potential involvement in heavy metal tolerance during mycorrhizal symbiosis. Environ. Microbiol. 18, 2446–2454.

Ren, M., Zhang, Z., Wang, X., Zhou, Z., Chen, D., Zeng, H., Zhao, S., Chen, L., Hu, Y., Zhang, C., Liang, Y., 2018. Diversity and contributions to nitrogen cycling and carbon fixation of soil salinity shaped microbial communities in Tarim Basin. Front. Microbiol. 9, 431.

Richards, M.P., 1989. Influence of egg production on zinc, copper and iron metabolism in the Turkey hen (Meleagris gallopavo). Comp. Biochem. Physiol. 93A, 811–817.

Salmela, H., Stark, T., Stucki, D., Fuchs, S., Freitak, D., Dey, A., Kent, C.F., Zayed, A., Dhayegude, K., Hokkanen, H., Sundstorm, L., 2016. Ancient duplications have led to functional divergence of vitellogenin-like genes potentially involved in inflammation and oxidative stress in honey bees. Genome Biol. Evol. 8, 495–506.

Seehuus, S.C., Norberg, K., Gimsa, U., Krekling, T., Amdam, G.V., 2006. Reproductive protein protects functionally sterile honey bee workers from oxidative stress. Proc. Natl. Acad. Sci. U. S. A. 103, 962–967.

Shu, Y., Zhou, J., Tang, W., Lu, K., Zhou, Q., Zhang, G., 2009. Molecular characterization and expression pattern of Spodoptera litura (Lepidoptera, Noctuidae) vitellogenin, and its response to lead stress. J. Insect Physiol. 55, 608–616.

Suenaga, H., Ohnuki, T., Miyazaki, K., 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environ. Microbiol. 9, 2289–2297.

Tartar, A., Wheeler, M.M., Zhou, X., Coy, M.R., Boucias, D.G., Scharf, M.E., 2009. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. Biotechnol. Biofuel. 2, 1–19.

Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J., 2012. Heavy metal toxicity and the environment. Mol. Clin. Environ. Toxicol. 133–164.

Teste, M.A., Duquenne, M., François, J.M., Parrou, J.L., 2009. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. BMC Mol. Biol. 10, 99.

Thakur, B., Yadav, R., Fraissinet-Tachet, L., Mermeisse, 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.

Tufail, M., Takeda, M., 2008. Molecular characteristics of insect vitellogenins. J. Insect Physiol. 54, 1447–1458.

Tufail, M., Naeemullah, M., Elmogy, M., Sharma, P.N., Takeda, M., Nakamura, C., 2010. Molecular cloning, transcriptional regulation and differential expression profiling of vitellogenin in two wing morphs of the brown planthopper, *Nilaparvata ligens* Stal (Hemiptera, Delphacidae). Insect Mol. Biol. 19, 787–798.

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.

Wallace, R.A., 1985. Vitellogenesis and oocyte growth in non-mammalian vertebrates. Dev. Biol. 1, 127–177.

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.

Zielinski, B.L., Allen, A.E., Carpenter, E.J., Coles, V.J., Crump, B.C., Doherty, M., McCrow, J.P., 2016. Patterns of transcript abundance of eukaryotic biogeochemically-relevant genes in the Amazon River plume. PLoS One 11, e0160929.

Žifčáková, L., Větrovský, T., Howe, A., Baldrian, P., 2016. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. Environ. Microbiol. 18, 288–301.

Ziller, A., Yadav, R.K., Capdevila, M., Reddy, M.S., Vallon, L., Marmeisse, R., Atrian, S., Palacios, O., Frassinet-Tachet, L., 2016. Metagenomics analysis reveals a new metallothionein family, sequence and metal-binding features of new environmental cysteine-rich proteins. J. Inorg. Biochem. 167, 1–11.

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-020-12152-6

RESEARCH ARTICLE

Protection from metal toxicity by Hsp40-like protein isolated from contaminated soil using functional metagenomic approach

Bharti Thakur[1] · Rajiv Yadav[2,3] · Arkadeep Mukherjee[1] · Delphine Melayah[2] · Roland Marmeisse[2] · Laurence Fraissinet-Tachet[2] · Mondem Sudhakara Reddy[1]

Received: 8 August 2020 / Accepted: 16 December 2020

Abstract

Pollution in the environment due to accumulation of potentially toxic metals results in deterioration of soil and water quality, thus impacting health of all living organisms including microbes. In the present investigation, a functional metagenomics approach was adopted to mine functional genes involved in metal tolerance from potentially toxic metal contaminated site. Eukaryotic cDNA library (1.0–4.0 kb) was screened for the genes providing tolerance to cadmium (Cd) toxicity through a functional complementation assay using Cd-sensitive *Saccharomyces cerevisiae* mutant *ycf1*$^{\Delta}$. Out of the 98 clones able to recover growth on Cd-supplemented selective medium, one clone designated as PLCc43 showed more tolerance to Cd along with some other clones. Sequence analysis revealed that cDNA PLCc43 encodes a 284 amino acid protein harbouring four characteristic zinc finger motif repeats (CXXCXGXG) and showing partial homology with heat shock protein (Hsp40) of *Acanthamoeba castellanii*. qPCR analysis revealed the induction of PLCc43 in the presence of Cd, which was further supported by accumulation of Cd in *ycf1*$^{\Delta}$/PLCc43 mutant. Cu-sensitive (*cup1*$^{\Delta}$), Zn-sensitive (*zrc1*$^{\Delta}$) and Co-sensitive (*cot1*$^{\Delta}$) yeast mutant strains were rescued from sensitivity when transformed with cDNA PLCc43 indicating its ability to confer tolerance to various potentially toxic metals. Oxidative stress tolerance potential of PLCc43 was also confirmed in the presence of H_2O_2. Present study results suggest that PLCc43 originating from a functional eukaryotic gene of soil community play an important role in detoxification of potentially toxic metals and may be used as biomarker in various contaminated sites.

Keywords Metatranscriptomics · Functional metagenomics · Metal tolerance · Type I Hsp40 · Metal uptake · Cadmium · Yeast complementation

Introduction

Metal contamination due to increasing anthropogenic activities is a significant problem of the industrial era because of the persistent and irreversible nature of these contaminants (Fernández et al. 2018). Soil contamination by potentially toxic metals not only modify soil physico-chemical properties but also deteriorates its microbial quality, thus leading directly or indirectly to harmful effects on the health of all members of food chain including animals and human beings (Dong et al. 2011; Nabulo et al. 2010). Soil is a rich habitat of highly diverse microorganisms, many of which are still unknown and uncultivable as of yet, but have a high metabolic response potential to survive in polluted environments. Exploring functional microbial communities has come across various cellular mechanisms involved in maintaining homeostasis in potentially toxic metal polluted environment, as chelation, efflux, or cellular compartmentalisation (Clemens 2006; Blindauer and Leszczyszyn 2010; Gadd 2010; Reddy et al. 2014, 2016). To harness the hidden functional diversity and in situ mechanisms for revitalizing the polluted soil by attenuation of the detrimental effects of accumulated toxic metals, metatranscriptomics approach has been introduced recently.

Laurence Fraissinet-Tachet and Mondem Sudhakara Reddy contributed equally to this work.

Responsible editor: Robert Duran

✉ Mondem Sudhakara Reddy
msreddy@thapar.edu

[1] Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India

[2] Univ Lyon, Université Claude Bernard Lyon 1, CNRS, INRAE, VetAgro Sup, UMR Ecologie Microbienne, F-69622 Villeurbanne, France

[3] Department of Botany, University of Allahabad, Prayagraj, Uttar Pradesh, India

Metatranscriptomics is a powerful tool in exploring complex biological system as this approach is designed without any limitation to living organisms and opened to any strata of life (Gonzalez et al. 2018; Thavamani et al. 2017). Function-based approach of metagenomics involves conversion of mRNAs into complementary DNA libraries followed by transcription and expression in suitable hosts (Bailly et al. 2007; Grant et al. 2006). It has been reported that function-based metagenomic approach may explore the functional diversity by directly using mRNAs, thus providing actively expressed gene repertoire in environmental samples (Bailly et al. 2007; Marmeisse et al. 2017; Thakur et al. 2018). This approach is a function-driven approach and recognized as an authoritative tool in unearthing various functional genes playing fundamental role in numerous phenomena including biological response mechanisms to environmental conditions (Damon et al. 2011; de Menezes et al. 2012; Zielinski et al. 2016). Function-based metagenomics has identified numerous ecologically significant genes as those encoding for oligopeptide transporters (Damon et al. 2011), cysteine-rich metallothioneins (Ziller et al. 2017) and candidate proteins involved in metal tolerance (Thakur et al. 2019). This approach provided an unprecedented overview of gene expression profiling in microbial communities, thus providing huge information on functional microbial diversity. The functional microbial diversity of muskoxen rumen (Qi et al. 2011), forest soils (Damon et al. 2011), phenanthrene-contaminated soil (de Menezes et al. 2012), activated sludge from wastewater (Yu and Zhang 2012), acid mine drainage and root microbiome (Chen et al. 2015; Gonzalez et al. 2018; Teng et al. 2017) are some of the examples.

Metal stress usually alters cell homeostasis by inactivating enzymes and by suppressing protein functions (Hossain et al. 2012). Induction of heat shock proteins (HSPs) may represent a critical protective, eco-physiologically adaptive and genetically conserved response in some living organisms that undergo environmental stress (Hasan et al. 2017). HSPs play a key function by restoring protein conformation and consequently cell homeostasis (Rhee et al. 2009). Among the major categories of HSPs, Hsp40 proteins play the role of molecular chaperones that act either in an independent way or as co-chaperones with Hsp70 (Craig et al. 2006). Hsp40, also known as J- proteins, are categorized into three different classes, types I, II, and III, depending upon the presence or absence of one or more of four structurally defined domains. Type I Hsp40 proteins possess all the four domains similar to those found in *E. coli* DNAJ, i.e. an N-terminal J-domain linked by a glycine/phenylalanine (G/F)-rich region to a cysteine-rich (CXXCXGXG) zinc finger region and followed by a C-terminal substrate-binding domain. Type II Hsp40s are similar to the class I proteins, except the absence of zinc finger domain in their structure, whereas type III Hsp40 proteins only retain the J domain (Tamadaddi and Sahi 2016). In addition, Hsp40 proteins provide tolerance to multiple types of stressors and protect living organisms from oxidative stress (Jiang et al. 2009). These proteins accumulate in a wide range of plant and animal species in response to environmental stressors to maintain cellular homeostasis (Gillis et al. 2013; Gupta et al. 2010).

Screening of yeast cDNA library constructed from RNA isolated from a metal-contaminated site allowed us to identify several genes involved in Cd tolerance when expressed in yeast (Thakur et al. 2018, 2019; Mukherjee et al. 2019). In the present study, one of the cDNAs, PLCc43, whose deduced protein sequence showed partial sequence similarity to eukaryotic type I Hsp40 proteins, was characterized and its role in providing tolerance to potentially toxic metals in metal-sensitive yeast mutants was assessed.

Materials and methods

Soil sampling and analysis

Soil samples were collected from the metal contaminated site of Pierrelaye (49° 1′ 45″ N; 2° 10′ 32″ E), northwest of Paris, France. Poplar plants are cultivated in this site, which was earlier an agricultural land and became contaminated with potentially toxic metals due to extended use of untreated wastewater for irrigation. Twenty soil cores were collected, mixed in equal volumes to make a composite sample, sieved through 2.0-mm mesh, and preserved in dry ice and then stored at − 70 °C until further experiments. The physicochemical properties of the soil were determined as described in Thakur et al. (2019) and were as follows: pH 7.15; total phosphorus 291 mg/kg; available phosphorus 14.2 mg/kg; organic carbon 1.6%; and total nitrogen 0.12%. The metal contents (mg/kg) of the soil samples were as follows: Cd 2.5, Zn 385 and Cu 64.

Strains, culture conditions and transformation

The mutant strains of *Saccharomyces cerevisiae* sensitive to cadmium (*ycf1*$^{\Delta}$), copper (*cup1*$^{\Delta}$), zinc (*zrc1*$^{\Delta}$) and cobalt (*cot1*$^{\Delta}$) were used for transformation (Gietz and Schiestl 2007). The *ycf1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ strains are derived from the wild-type strain BY4741 (*MAT*α, *his3*$^{\Delta}$*1, met15*$^{\Delta}$*0, leu2*Δ*0, ura3*Δ*0*). The *ycf1*$^{\Delta}$ yeast strain lacks an ABC transporter encoded by gene *YCF1* that mediates the transport of Cd conjugates into vacuoles (Li et al. 1997), while *zrc1*$^{\Delta}$ lacks the transporter protein *ZRC1p* responsible for Zn sequestration into vacuoles (Li and Kaplan 1998). The hypersensitivity of *cot1*$^{\Delta}$ towards Co metal is due to deletion of *COT1* gene that helps in efflux of cobalt ions into the vacuole (Conklin et al. 1992). Similarly, *cup1*$^{\Delta}$ mutant strain is derived from knocking out the copperthionein gene (Longo et al. 1996) from its isogenic wild-type strain *cup1^s* (*MATα, leu2-3, 112his3^△1, trp1-1, ura3-50, gal1, CUP1^S*). Yeast strains BY4742 and its derivative superoxide dismutase mutant *sod2^△* (BY4742 *sod2*::*KanMX*) were used for oxidative stress experiments. For yeast complementation assays, minimal medium supplemented with 2% glucose (Hi-Media Laboratories, India) containing dropout supplement without uracil (SD-Ura) (Takara-Clontech, Japan) was used. Yeast strains were maintained on yeast extract peptone dextrose (YEPD) (Hi-Media Laboratories, India). Bacterial transformations were carried out by using One Shot® TOP10 Electrocomp™ *E. coli* cells (Thermofisher, USA). The yeast strains were grown at 30 °C while bacteria at 37 °C.

Construction of sized cDNA library and screening for Cd tolerance

Total RNA was isolated from the soil composite samples using the RNA PowerSoil® Total RNA Isolation Kit (MoBio Laboratories, Carlsbad, CA) according to the manufacturer's instructions. Extracted total RNA was checked for quality by agarose gel electrophoresis and Bioanalyzer (Agilent Technologies, USA). The extraction yields ranged from 330 to 980 ng/g of soil and at least 3 μg of total soil RNA were obtained from each sample. cDNAs were synthesized from total soil RNA with Mint-2 cDNA synthesis kit (Evrogen, Moscow, Russia). Size fractionation of cDNAs was performed as described in Yadav et al. (2014). Size fractions of cDNAs, A (< 0.5 kb), B (0.5–1 kb) and C (> 1 kb) were ligated downstream of PGK1 promoter of *S. cerevisiae* shuttle plasmid vector pFL61 modified with SfiA and SfiB restriction sites (Minet et al. 1992). 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 medium) were pooled to represent each of the cDNA size libraries. Plasmid DNA from the pooled cells representing the cDNA library C (> 1.0 kb) was extracted with QIAprep Spin Miniprep Kit (Qiagen, USA) and transformed into Cd-sensitive *ycf1^△* yeast mutant by lithium acetate method (Gietz and Schiestl 2007) and were regenerated on SD-Ura medium amended with 40 μM Cd. The selected clones were treated with 5-fluoroorotic acid (5-FOA) to eliminate false-positive colonies. One of the positive clones, PLCc43, was selected for further characterization based on its ability to tolerate high concentrations of Cd.

Sequence analysis

Plasmid DNA from yeast *ycf1^△*/PLCc43 was isolated with Zymoprep yeast plasmid Miniprep II kit (Zymo Research) and sequenced the insert PLCc43 with vector-specific primers NF and NR (Table 1) by using Sanger sequencing technology. Sequence homology search was performed by using BLASTX (https://blast.ncbi.nlm.nih.gov/Blast.cgi/). Open reading frame of the sequence was determined by using the webserver https://www.ncbi.nlm.nih.gov/orffinder/ and molecular weight and theoretical isoelectric point (pI) for the putative encoded protein by http://web.expasy.org/ compute_pi/. Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) was used to align the homologous sequences along with PLCc43. Phosphorylation sites within the deduced polypeptide PLCc43 was predicted with NetPhos 2.0 server (http://www.cbs.dtu.dk/services/ NetPhos/).

Relative quantification of PLCc43

Cadmium-sensitive yeast mutant *ycf1^△* carrying PLCc43 was tested for its relative quantification in presence of different concentrations of Cd. To accomplish this, flasks containing 50 ml of SD-Ura broth amended with different concentrations of Cd (0, 40, 60 and 80 μM) was inoculated with *ycf1^△*/PLCc43 cells and incubated for 48 h at 30 °C. The cells were harvested and grounded with liquid nitrogen. Total RNA was isolated by Trizol reagent, and RNase-free DNaseI (Thermo Scientific, USA) was used to remove genomic DNA contamination. cDNA was synthesized from total RNA by using PrimeScript™ first strand cDNA synthesis kit (Takara, Japan) as per manufacturer's instructions. To assess the relative abundance of PLCc43 and their relative basal levels, 25 μl qPCR reaction mixtures were prepared. The reaction contained 12.5 μl SYBR® Premix Ex Taq™ (Tli RNase H Plus) (Takara, Japan) as recommended by the manufacturer, 0.2 μM specific primers (Table 1), 1.0 μL diluted cDNA (20 fold dilution; approximately 50 ng) and H_2O to make the final volume. The qPCR analysis was performed on the Mastercycler® ep realplex system (Eppendorf AG, Hamburg, Germany). The PCR conditions were 95 °C for 2 min (1 cycle), followed by 35 cycles of 95 °C for 15 s, 53 °C for 15 s and 68 °C for 20 s. The E values for PLCc43 were within the range of 0.94 to 1.27. Three reference genes, actin (*ACT*; SGD: S000001855), TATA-binding protein–associated factor (*TAF10*; SGD:S000002574) and transcription factor class C *(TFC1*; SGD: S000000327) were used in this study (Teste et al. 2009). The relative amount of PLCc43 was determined according to $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) by using TFC1 as reference gene because of its minimum stability than other two reference genes. The base level expression of PLCc43 in control cells (without Cd augmentation) was considered as one and calculated the fold increase in Cd augmented cells. To confirm that there were no traces of DNA present in RNA extracts, PCR was performed with nonreverse-transcribed RNA samples as control. The reaction mixture without cDNA served as negative control. All qPCR measurements were performed on RNA extracted from three independent biological samples (experimental

Table 1 Primers used to amplify different genes in this study.

Gene	Name	Sequence
pFL61	pFL61 NF	5'-CAGATCATCAAGGAAGTAATTATCTAC-3'
	pFL61 NR	5'-CAGAAAAGCAGGCTGGG AAGC-3'
Actin	qACTF	5'-CGAATTGAGAGTTGCCCCAG-3'
	qACTR	5'-CAAGGACAAAACGGCTTGGA-3'
TAF10	qTAF10F	5'-GCTAACAACAGTCAGGCGAG-3'
	qTAF10R	5'-GAGCCCGTATTCAGCAACAG-3'
TFC1	qTFC1F	5'-ACGACGCTGCTTTGGAAAAT-3'
	qTFC1R	5'-GCTTTTCATTGTTTCCGCGG-3'
PLCc43	qHspF	5'- GACCGTATGCCCCAAATGTC 3'
	qHspR	5'-GGGAATTACACCAGGCTCCT-3'

replicates), and for each RNA extract, three technical replicates were performed.

Tolerance to metal toxicity and metal content

Cd-sensitive (*ycf1*$^{\Delta}$), Cu-sensitive (*cup1*$^{\Delta}$), Zn-sensitive (*zrc1*$^{\Delta}$) and Co-sensitive (*cot1*$^{\Delta}$) yeast mutants expressing PLCc43 were tested for tolerance towards Cd, Cu, Zn and Co, respectively as follows: serial dilutions (0.1–0.0001) were prepared from calibrated overnight grown transformant yeast cells and empty vector pFL61 carrying control cultures (wild-type BY4741 and mutant strains). Five microliters of each dilution was dropped on SD-Ura plates supplemented with respective metal concentrations, i.e. $CdSO_4$ 40 μM, $CuSO_4$ 150 μM, $ZnSO_4$ 10 mM and $CoCl_2$ 2 mM. Liquid broth assays were performed in parallel to validate the broad spectrum of metal tolerance. To do this, 20 ml of SD-Ura were inoculated with exponentially cultured cells to achieve an initial OD_{600} = 0.02 and were allowed to grow at 30 °C for 3 h. Then, varying concentrations of $CdSO_4$ (40 μM–80 μM), $CuSO_4$ (150 μM–500 μM), $ZnSO_4$ (10 mM–12 mM) and $CoCl_2$ (2 mM–4 mM) were added separately to the respective yeast cell cultures. Growth was monitored via OD measurements at 600 nm at intervals of 3 h for 48 h. Empty vector pFL61 transformed yeast mutants and wild-type BY4741 served as controls.

For metal accumulation studies, all yeast cells were allowed to grow overnight in SD-Ura medium till the OD_{600} = 1.0 was attained. Cells were harvested by centrifugation and re-suspended in sterile water. One percent inoculum was used to inoculate 50 ml of SD-Ura broth in 250 ml flasks and was incubated at 30 °C at 200 rpm. After 5 h, various concentrations of $CdSO_4$ (40 μM–80 μM), $CuSO_4$ (150 μM–500 μM), $ZnSO_4$ (10 mM–12 mM) and $CoCl_2$ (2 mM–4 mM) were amended in the media similar to growth kinetics study. After 40 h of growth, cells were harvested and washed 3 times with 10 mM EDTA solution. Final washing was done with sterile deionized water to eliminate residual EDTA and dried for 24 h at 80 °C. A combination of HNO_3 and $HClO_4$ (ratio 2:1 v/v) was used to digest dry yeast biomass (Hseu (2004) and the metal content was determined through ICP-MS (Element XR, Thermo Fisher Scientific, Germany). Experiments were conducted in a 3 × 3 factorial design with yeast strains and concentrations of the metals arranged in a completely randomized block design with three replicates. The experiment consisted of three yeast inoculation treatments (*ycf1*$^{\Delta}$/pFL61, *ycf1*$^{\Delta}$/PLCg39 and BY4741) and three concentrations of $CdSO_4$ (40, 60 and 80 μM). Similar design of experiments were conducted for Cu, Zn and Co using the respective yeast mutants.

Furthermore, the involvement of PLCc43 in providing the tolerance to metal toxicity was tested by transforming PLCc43 into the wild-type strain BY4741, which already consists of gene for yeast Hsp40 protein. The transformed cells were grown in presence of 60 μM of $CdSO_4$. The wild-type strain BY4741 transformed with empty vector pFL61 was used as positive control. Calibrated inoculums consisting of 2×10^6 cells from actively growing cultures were used as starting materials and the growth of both the strains was recorded for 36 h with an interval of 3 h.

Oxidative stress

The superoxide dismutase yeast mutant *sod2*$^{\Delta}$ was used to examine the role of PLCc43 in oxidative stress. To do this, functional complementation of *sod2*$^{\Delta}$ using PLCc43 was evaluated in SD-Ura medium supplemented with H_2O_2 (0–0.5 mM) following drop assays and growth kinetics study. For controls, the parent-type BY4742 and *sod2*$^{\Delta}$ both transformed with empty vector pFL61 were used. Growth response was determined by recording the OD_{600} until 30 h at an interval of 5 h.

Statistical analysis

The statistical tests were carried out by using GraphPad Prism (version 5.04) (GraphPad Prism, San Diego California, USA) software. Two-way analysis of variance was performed to examine the effect of various concentrations of metals on the cell growth and metal uptake of different yeast strains and their interactions. The significant differences among the means were compared with Tukey's test at $P < 0.05$ when one way-ANOVA was used

Results and discussion

Screening of cadmium tolerant cDNAs

Yeast mutant sensitive to Cd ($\mathit{ycf1}^{\Delta}$) was used for screening of the environmental cDNAs (library C, > 1.0 kb insert size) synthesized from total soil RNA and the quality of total RNA was determined by agarose gel. Among various clones screened through functional complementation by drop assay, 98 clones were retrieved, which are capable of growing in Cd-amended medium (Supplementary figure S1). These results indicated that some of the transformed clones were able to rescue $\mathit{ycf1}^{\Delta}$ cells from Cd sensitivity. The clone $\mathit{ycf1}^{\Delta}$/PLCc43 was selected for further studies based on its tolerance to high Cd concentrations. Thakur et al. (2018) reported a 695-bp-long eukaryotic transcript PLBe1 providing tolerance towards potentially toxic metals such as Co, Cu, Zn and Cd. Numerous ecologically important genes playing a significant role in many biological processes have been identified using function-based metagenomics approach (Damon et al. 2012; Lehembre et al. 2013; Ren et al. 2018; Ziller et al. 2017).

Sequence analysis of PLCc43

Sequence analysis of PLCc43 revealed that it consisted of 1269 bp with 855 bp open reading frame encoding a putative protein of 284 amino acids with theoretical molecular mass of 31.2 kDa and calculated isoelectric point of 8.99. The sequence showed 51% sequence similarity (65% positives) with Hsp40 protein of *Acanthamoeba castellanii*. Generally, Hsp40 proteins are categorised based on presence of different combinations of three domains, namely, J domain that helps in stimulating the ATPase activity of Hsp70, long stretch of G/F amino acid–rich region and cysteine-rich region that contains zinc-binding CXXCXGXG motif repeats and a carboxy-terminal substrate-binding domain (Caplan and Douglas 1991; Qiu et al. 2006). Putative PLCc43p may display the fundamental structural similarities with type I Hsp40 proteins, i.e. a partial J domain with the conserved tri-peptide HPD (histidine, proline and aspartic acid) motif, a stretch of glycine (G)/phenylalanine (F)-rich region and four zinc finger motif repeats although the final G amino acid in the fourth stretch CXXCXGXG was missing (Fig. 1a). Sequence alignment with other homologous type I Hsp40 proteins showed highly conserved pattern of PLCc43 (Fig. 1b).

Hsp40 proteins are highly conserved proteins that play an important role in protein degradation, translocation, folding and unfolding of protein by stimulating the ATPase activity of another 70-kDa chaperone, Hsp70, which is induced by stress (Caplan and Douglas 1991; Cyr et al. 1994; Qiu et al. 2006). Biochemical and genetic studies have reported the functional redundancy between *E. coli* DnaJ proteins and eukaryotic Hsp40 proteins; hence, the concept of regulation of Hsp70 by Hsp40 via stimulating ATPase activity with the help of J domain as reported in *E. coli* has also been conserved in higher organisms as well (Cyr et al. 1994). Similarly, the tri-peptide motif, HPD, has also been evidenced to play crucial role in functional regulation of Hsp70 as mutations in this conserved motif block the activity of the Hsp70 ATPase (Tsai and Douglas 1996; Wall et al. 1994). Although J domain is considered as an essential feature for functioning of Hsp40 proteins, studies have identified Hsp40 proteins that do not need J domain to perform all their functions (Hageman et al. 2010; Sahi et al. 2010). This further suggests that Hsp40 proteins can perform J domain–independent functions and not all the functions of Hsp40 proteins are Hsp70 dependent. For instance, DnaJA1, a class I Hsp40 protein in mammals, which acts as host factor for influenza A virus replication, could associate itself with the PA and PB2 subunits of RNA polymerase by C-terminal substrate–binding domain, which lacks J domain (Cao et al. 2014). Further, cysteine-rich zinc finger domains of polypeptide encoded by cDNA PLCc43 possess same sequence pattern as in type I Hsp40 protein of *S. cerevisiae* Ydj1 (Caplan and Douglas 1991). When analysing the structure of Hsp40 proteins, Banecki et al. (1996) reported that the four cysteine-rich zinc finger domains combine by pairs, each of the pairs binding to one unit of Zn^{2+} ion, hence allowing the Hsp40 proteins to complex two Zn^{2+} ions. Zinc finger domains known to prevent the aggregation of denatured proteins, is also a part of peptide-binding domain, which is responsible for chaperone function of Hsp40 proteins (Banecki et al. 1996; Szabo et al. 1996). The fourth domain found in the carboxyl-terminal part of the protein was less conserved and plays a significant role in substrate binding (Banecki et al. 1996). Lu and Cyr (1998) reported that the mutant $\text{Ydj1}_{179\text{-}384}$ protein with completely lacking J domain, G/F region and a large part of cysteine rich zinc finger region, was able to suppress the aggregation of rhodanese in vitro thus arguing that some functions rely on the C-terminal substrate-binding fragment only. Similarly, Sahi et al.

Fig. 1 **a** Deduced amino acid sequence of type I Hsp40-like protein encoded by cDNA PLCc43: highly conserved tri-peptide HPD (green highlighted) and G/F-rich region (boxes in bold) are shown. Four cysteine-rich domains CXXCXGXG (yellow-highlighted bold letters) are underlined in the polypeptide sequence. Putative phosphorylation sites are highlighted in red. **b** Alignment of sequences of type I Hsp40 proteins. The fully conserved tripeptide HPD is highlighted in yellow; the cysteine residues from cysteine-rich domains CXXCXGXG are coloured in red, respectively. Hsp40 from *Acanthamoeba castellini* (XP_004349556) and DnaJ homolog members from *Danio rerio* (NP_998658) and *Xenopus tropicalis* (NP_001004807) are shown

(2013) advocates that C-terminal region of Hsp40 Class I proteins is the most evolved region of Hsp40 proteins, which results into a more diverse and specific Hsp40 class I protein family with binding capability towards different polypeptides.

Polypeptide analysis of PLCc43p coding for Hsp40-like protein also marked the presence of 22 metabolically important phosphorylation sites (Fig. 1). The sequence characterization of cDNA PLCc43 thus evidenced features supporting that the encoded polypeptide may belong to the class I of Hsp40 proteins. The sequence of PLCc43 obtained in this study was deposited at NCBI database under the accession number MK079355.

Relative quantification of PLCc43

Expression levels of PLCc43 increased with increase in Cd concentration and the maximum expression of 68-fold increase was observed at 80 μM of $CdSO_4$. Eighteen-fold increase at 40 μM and 55-fold increases at 60 μM $CdSO_4$ was recorded, respectively, in *ycf1*$^{\Delta}$/PLCc43 cells compared to control cells (Fig. 2). Several studies have reported the overexpression of Hsp40 proteins along with other heat shock proteins and their role in degradation and refolding of proteins in various organisms in response to various accumulated metals (Barque et al. 1996; Cai and Zhang 2000; Fusco et al. 2005).

Tolerance to metal toxicity and metal content

The role of PLCc43 in conferring metal tolerance when overexpressed within different hypersensitive mutants of *S. cerevisiae* was studied. The heterologous expression of PLCc43 allowed all hypersensitive yeast mutant strains used (*ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$and *cot1*$^{\Delta}$) to tolerate towards Cd, Cu, Zn and Co toxicity compared to yeast strains carrying empty vector (Fig. 3A, B, C & D). The tolerance potential of PLCc43 towards the four metals used in this study was further authenticated by growing the transformed yeast mutants into metal-amended liquid media (Fig.4). The transformant *ycf1*$^{\Delta}$/

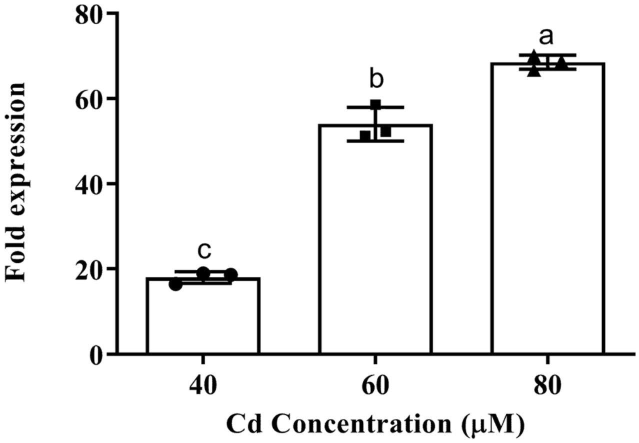


Fig. 2 Fold increase in expression levels of cDNA *ycf1*$^{\Delta}$/PLCc43 after 48-h incubation in medium supplemented with different concentrations of Cd. Values plotted are referred to the control condition (expression level in transformant *ycf1*$^{\Delta}$/PLCc43 without metal treatment). Individual data points are shown in each bar. Error bars are ± SD. Bars sharing a common letter are not significant at $P < 0.05$

PLCc43 showed high tolerance towards Cd in the medium compared to wild-type strain BY4741 and *ycf1*$^{\Delta}$ transformed with empty vector. However, with increase in external metal concentration, the growth was significantly reduced. Maximum growth was observed when *ycf1*$^{\Delta}$/PLCc43 cells were grown in presence of 40 μM of Cd compared to other concentrations (Fig. 4a). The growth of yeast cells decreased with increase in concentrations of external $CuSO_4$ in the medium. The *cup1*$^{\Delta}$/PLCc43 cells showed more growth in presence of $CuSO_4$ compared to wild-type strain BY4741, the maximum growth being at 150 μM of $CuSO_4$ (Fig.4B). The growth of *zrc1*$^{\Delta}$ cells carrying PLCc43 showed more tolerance to Zn compared to wild-type strain as well as *zrc1*$^{\Delta}$ cells carrying empty vector. The growth of yeast cells decreased significantly with increase in concentrations and the maximum growth was observed in *zrc1*$^{\Delta}$/PLCc43 cells at 10 mM $ZnSO_4$ (Fig. 4c). In the case of Co, over-expression of PLCc43 in hypersensitive yeast mutant *cot1*$^{\Delta}$ provided more tolerance against Co toxicity compared to the wild-type strain BY4741 and the *cot1*$^{\Delta}$ cells transformed with empty vector. *cot1*$^{\Delta}$ cells transformed with PLCc43 showed maximum growth at 2 mM of $CoCl_2$ compared to other concentrations (Fig. 4d). Significant variation among the treatments of yeast strains and concentrations of metals along with their interaction effect was observed when analysed by ANOVA (Table 2).

Accumulation of metal by respective yeast mutant strain carrying cDNA PLCc43 was further estimated by ICP-MS. The *ycf1*$^{\Delta}$ cells transformed with PLCc43 showed significantly higher accumulation of Cd at different concentrations compared to wild-type BY4741 as well as *ycf1*$^{\Delta}$ cells transformed with empty vector in all concentrations of Cd. The accumulation of Cd levels increased up to 60 μM of $CdSO_4$ and decreased at 80 μM of $CdSO_4$, which might be due to decrease in the growth of yeast cells (Fig. 5a). Similarly, the level of Cu accumulation increased with concentrations up to 300 μM of $CuSO_4$ after which a slight but not significant decrease was observed at 500 μM of Cu concentration (Fig. 5b). In case of

Table 2 The factorial ANOVA results indicating the effects of different concentrations of metals on cell growth and metal accumulation by different yeast strains. The *P* values < 0.001 are highly significant

Variables		Yeast strains (YS)	Metal concentrations (MC)	YS x MC
			Cell growth (O.D)	
	df	2, 18	2, 18	4, 18
In presence of Cd		< 0.001	< 0.001	< 0.001
In presence of Cu		< 0.001	< 0.001	< 0.001
In presence of Zn		< 0.001	< 0.001	< 0.001
In presence of Co		< 0.001	< 0.001	< 0.001
			Metal accumulation	
Cd accumulation		< 0.001	< 0.001	< 0.001
Cu accumulation		< 0.001	< 0.001	< 0.001
Zn accumulation		< 0.001	< 0.001	< 0.001
Co accumulation		< 0.001	< 0.001	< 0.001
			Oxidative stress	
	df	2, 42	6, 42	12, 42
		Yeast strains (YS)	Days (D)	YS x D
Control		0.141	< 0.001	0.349
0.25 mM H_2O_2		< 0.001	< 0.001	< 0.001
0.5 mM H_2O_2		< 0.001	< 0.001	< 0.001

df degrees of freedom

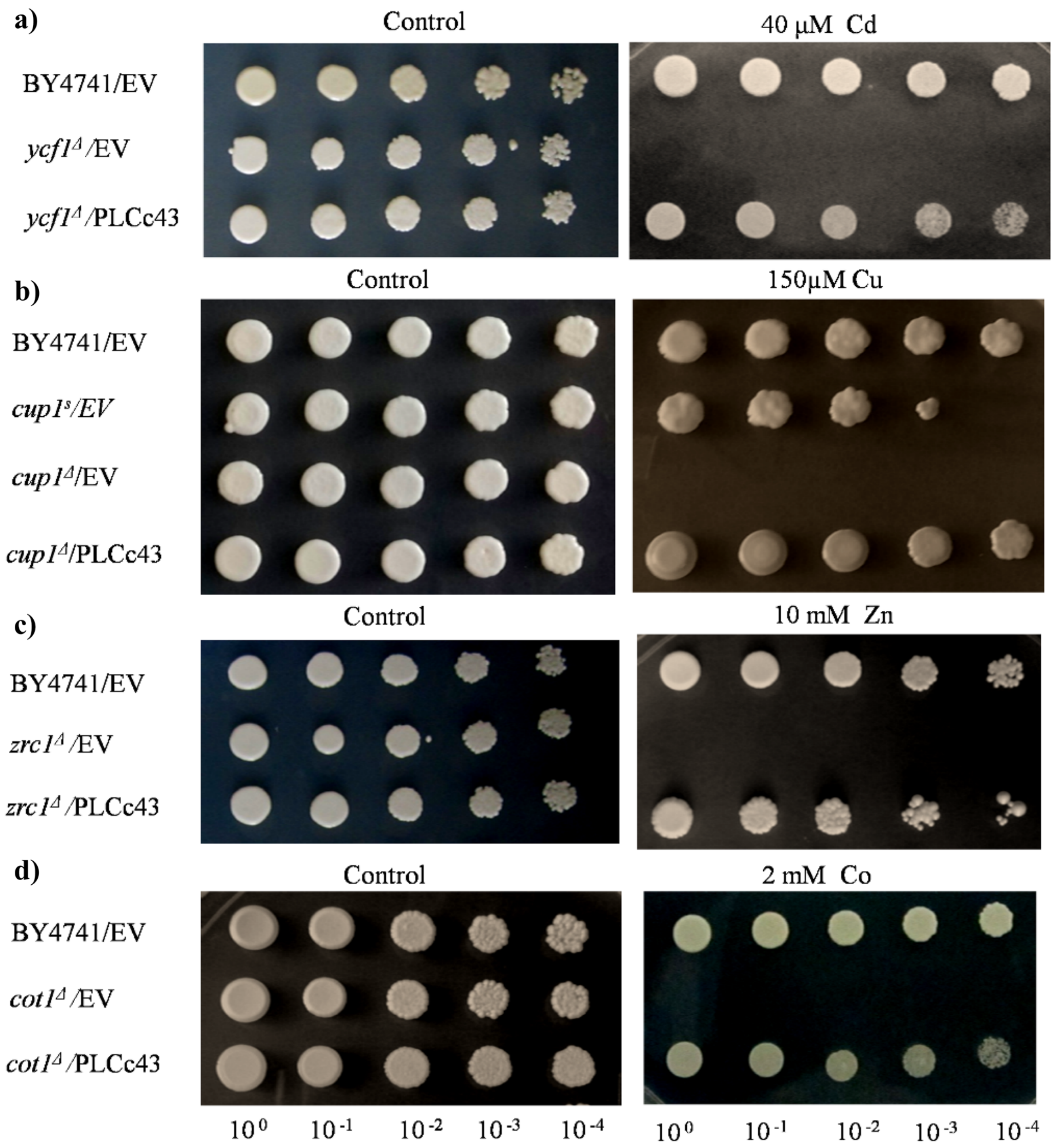


Fig. 3 Drop assay of metal-sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing PLCc43 at various concentrations **a** Cd 40 μM, **b** Cu 150 μM, **c** Zn 10 mM and **d** Co 2 mM. Wild strain BY4741/*EV*, *cup1*^{s}/*EV* serves as positive controls and *ycf1*$^{\Delta}$/EV, *cup1*$^{\Delta}$/*EV*, *zrc1*$^{\Delta}$/*EV* and *cot1*$^{\Delta}$/*EV* as negative controls, respectively. EV stands for empty vector

Zn and Co metals, maximum metal accumulation was observed for 10 mM of $ZnSO_4$ and 2 mM of $CoCl_2$, respectively, after which a sharp decrease was observed for either concentration of metal used (Fig. 5c and d). Overall, metal accumulation levels were significantly higher in yeast mutants expressing PLCc43 than in mutants or wild-type strain transformed with empty vector. A significant variation was observed among the yeast strains, metal concentration, and their interaction in relation to the metal accumulation (Table 2).

To investigate the combined effect of the yeast Hsp40 of BY4741 and PLCc43, the BY4741 transformed cells were grown in SD-Ura amended with 60 μM of $CdSO_4$. When compared to BY4741 carrying empty vector, growth of BY4741 transformed with PLCc43 showed significantly higher growth. This result clearly indicated that PLCc43 is providing additional tolerance to yeast cells compared to control in presence of 60 μM of $CdSO_4$ (Fig. 6).

Nover (1991) reported that the divalent cations of transition metal series such as Cd, Cu, Zn and Hg act as heat stress–like chemical stressors, which induce the production of various stress combating proteins such as Hsp40s. In many plant species, high expression of HSPs especially Hsp40s was

Fig. 4 Growth assay of metal-sensitive mutant strains *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ expressing cDNA PLCc43 in SD-Ura medium supplemented with **a** Cd 40–80 μM, **b** Cu 150–500 μM, **c** Zn 10–12 mM, **d** Co 2–4 mM. Wild-type strain BY4741 and the corresponding mutants transformed with empty vector (EV) are shown as controls. Individual data points are shown in each bar. Error bars are ± SD

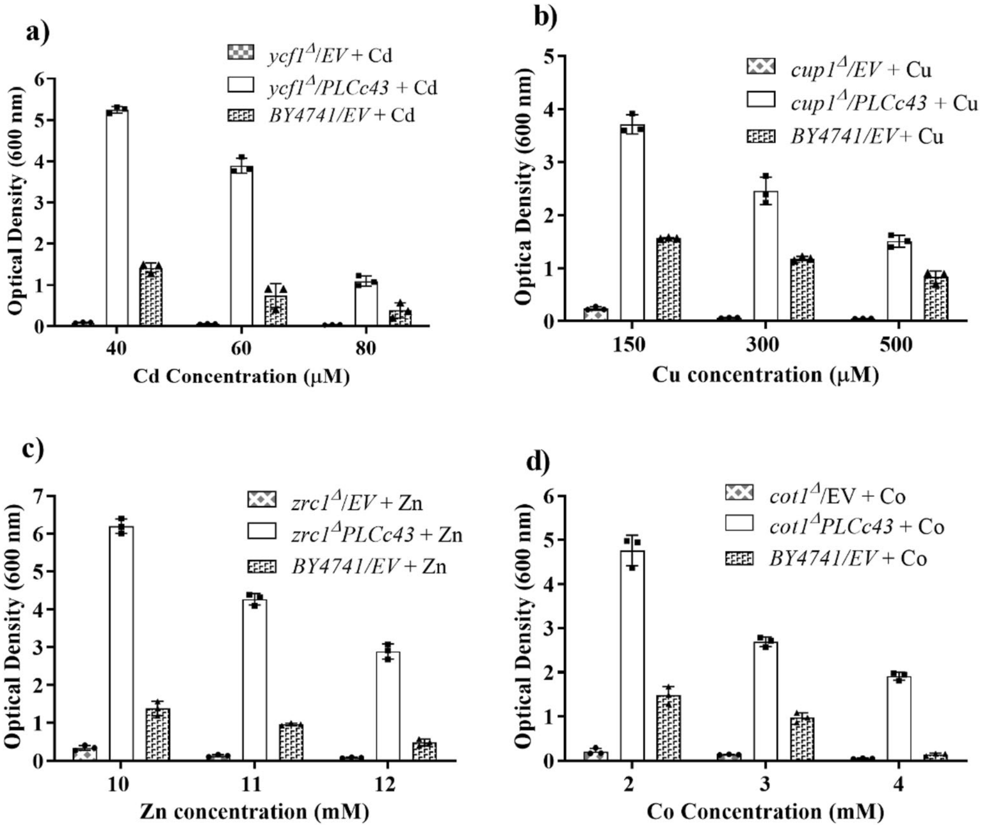
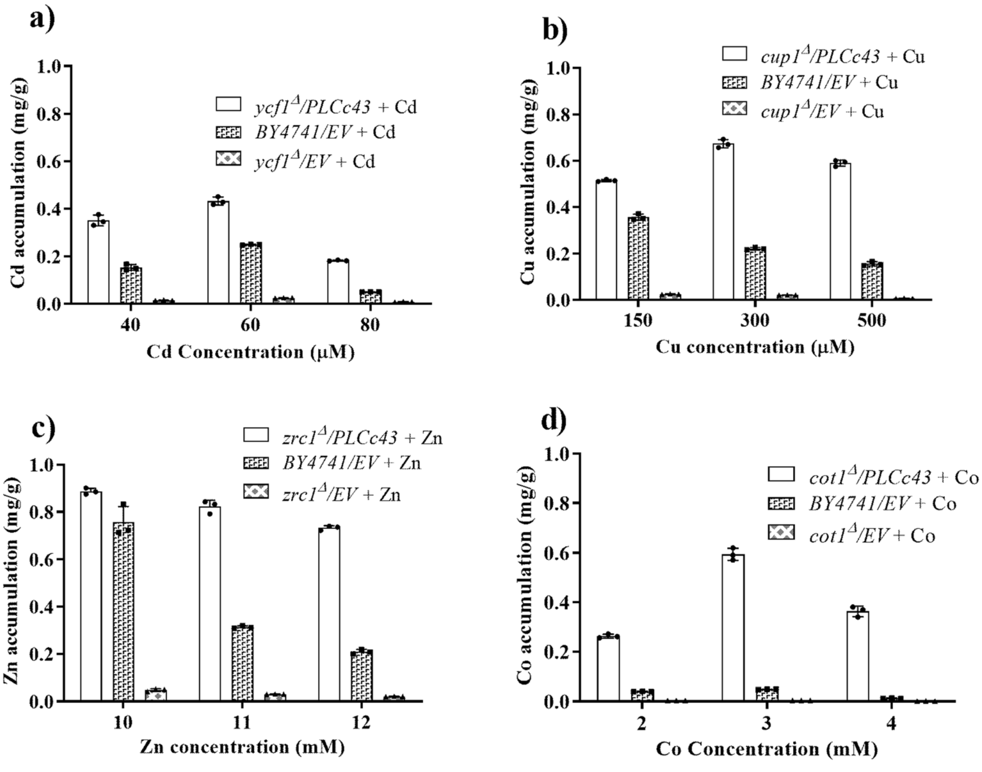


Fig. 5 Accumulation of **a** Cd, **b** Cu, **c** Zn and **d** Co in mutant yeast cells *ycf1*$^{\Delta}$, *cup1*$^{\Delta}$, *zrc1*$^{\Delta}$ and *cot1*$^{\Delta}$ carrying cDNA PLCc43 and BY4741/EV. Wild-type BY4741 transformed with only empty vector pFL61 (EV) and mutant strains with EV were used as positive and negative controls, respectively. Individual data points are shown in each bar. Error bars are ± SD



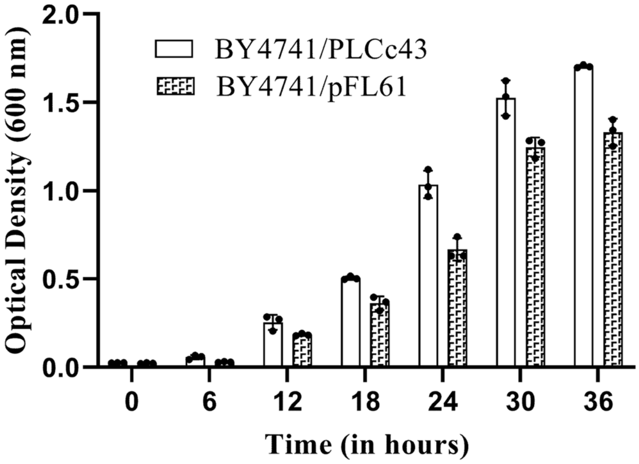


Fig. 6 Growth response of BY4741 transformed with PLCc43 and empty vector pFL61in SD-Ura medium supplemented with 60 µM of Cd. Individual data points are shown in each bar. Error bars are ±SD

observed under a variety of metal stress. Degradation of non-native proteins or proteolytic inactivation of regulatory proteins is a major cellular response of living organisms in metal-contaminated environments. It has been reported that Hsp40 proteins are involved in maintaining the cell homeostasis through translocation, refolding or degradation of aberrant proteins during cellular stress such as metal toxicity (Cyr et al. 1994; Suzuki et al. 2001). It was observed that Hsp40 in association with Hsp70 prevents aggregation of proteins and assist in refolding of non-native proteins under both normal and stressed environments (Frydman 2001; Hartl 1996). In a study, a total of twenty-one cDNAs including Hsp40 were reported in birch *Betula pendula* playing significant role in maintaining the homeostasis and in alleviating the toxic effects of accumulated Cu metal (Keinänen et al. 2007). Similarly, higher expression of Hsp40-like protein encoding gene conferred tolerance to beans when exposed to mercuric chloride (Chai et al. 2000). The mechanism behind this might be direct binding and trafficking, which has also been observed experimentally with copper (Field et al. 2002), or by refolding of denatured proteins due to the stress conditions (Meacham et al. 1999; Shen and Hendershot 2005). The expression of Hsp40 and Hsp70 limits the proteotoxic symptoms of metal ions and helps the detoxification and sequestration of these metal ions by metallothioneins (Haap et al. 2016).

Furthermore, in a study, deletion of Hsp40 protein Mdj1 in the filamentous entomopathogenic fungus *Beauveria bassiana* resulted into sensitivity towards various metals such as Zn, Cu and Fe along with some growth defects such as impaired cell wall integrity, reduced osmo-tolerance and anti-oxidant activities (Wang et al. 2017). Similarly, DnJC3 protein, an Hsp40 protein in honeybee, *Apis cerana cerana* has regulated the ill effects of various stress factors including metal ions (Zhang et al. 2019). Hence, overall, significant tolerance towards potentially toxic metals used in this study, i.e. Cd, Cu, Zn and Co was shown in model eukaryotic living organism that expressed soil originating cDNA, PLCc43.

Oxidative stress

The potential role of PLCc43 in response to oxidative stress was examined using the superoxide dismutase yeast mutant *sod2*$^{\triangle}$ expressing the cDNA PLCc43. The growth of wild-type BY472, *sod2*$^{\triangle}$ transformed with empty vector and *sod2*$^{\triangle}$/PLCc43 showed similar growth pattern when grown in absence of H_2O_2. The *sod2*$^{\triangle}$ cells expressing PLCc43 were able to confer tolerance to 0.25 mM and 0.5 mM of H_2O_2. The growth of BY4742 and *sod2*$^{\triangle}$/PLCc43 were comparable at these concentrations and were significantly higher than the *sod2*$^{\triangle}$ transformed with empty vector (Fig. 7). Significant variation among the yeast strains and the days along with their interaction was observed (Table 2) Oxidative stress is a result of uncontrolled oxidation process, which results in the accumulation of reactive oxygen species (ROS). The generation of ROS because of metal toxicity is very common. Mostly, metals cause inhibition of metabolic processes and are involved in direct electron transfer involving metal cations (Grimsrud et al. 2008). The abiotic stress due to formation of ROS is confronted by various stress-responsive proteins, out of which heat shock proteins play an important role. In a study by Qiu et al. (2006), over-expression of *LeCDJ1* DnaJ gene in tomato not only resulted in improving heat tolerance but also in increasing superoxide dismutase activity, which reduced accumulation of O_2^-. Similar studies on the effects of HSPs on ROS scavenging proteins and related to heat stress have been investigated. In *Arabidopsis*, over-expression of *RcHSP17.8* improved SOD activity (Jiang et al. 2009). Also, in tobacco, when the HSP *ZmHSP16.9* was over-expressed, it confirmed increase in the levels of peroxidase, catalase and superoxide dismutase activity (Sun et al. 2012). A study by Kim et al. (2008) indicated a possible mechanism of H_2O_2 resistance by HSPs. It was concluded that the increase in resistance against oxidative stress conferred by HSPs was due to the prevention of proteasome activity, which ultimately provided resistance against cell toxicity. Furthermore, these findings illustrate the role of proteins like type I Hsp40 in combating the metal toxicity and advocate the significance of functional approach of functional metagenomics in mining, the hidden treasures of soil for remediation of various contaminated sites.

Conclusions

The potential of cDNA PLCc43 to tolerate toxic concentrations of toxic metals was quite evident from the experimental observations. The translated protein had the characteristics of a type I Hsp40-like protein. Functional complementation of

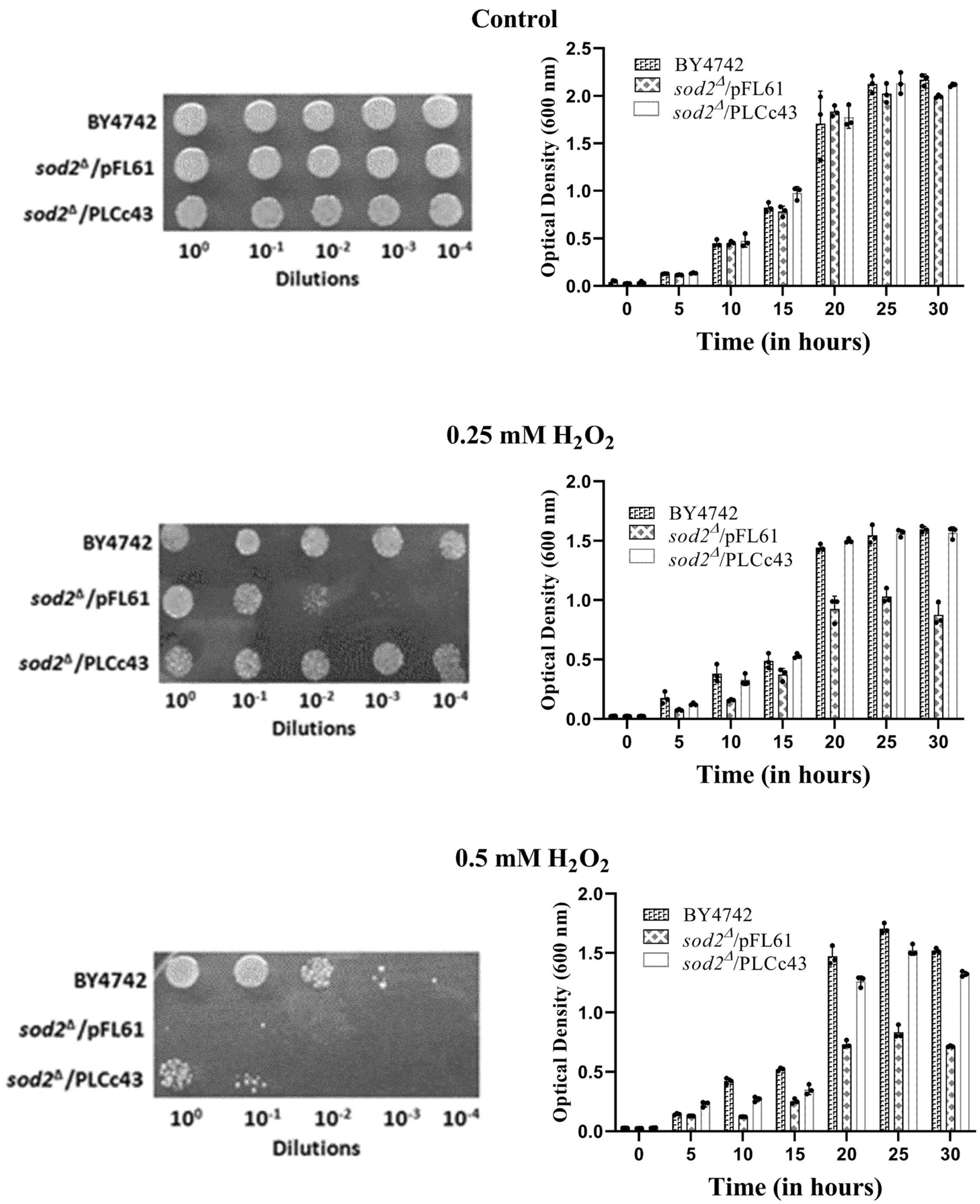


Fig. 7 Drop assay and growth response of yeast mutant *sod2*$^{\Delta}$ strain expressing cDNA PLCc43 in presence of different concentrations of H_2O_2. Wild-type strain BY4742 and *sod2*$^{\Delta}$ transformed with empty were used as controls. Individual data points are shown in each bar. Error bars are ±SD

S. cerevisiae using PLCc43 confers metal tolerance to the mutant strains of yeast, which categorize PLCc43 as potential member of functional repertoire participating in metal tolerance. Hsp40-like proteins might play an important role not only in targeted specific hosts as previously evidenced in numerous studies, but also in still uncharacterized living organisms that prevail in environments such as metal contaminated soils. This study described Hsp40-like protein as a protein originating from metal-polluted soil that may help living organisms to maintain homeostasis between flora and fauna of the soil environment through tolerating high concentrations of metals. Thus, function-based metagenomics approach can reveal diversity of functions related to stress tolerance, when applied to complex environments like polluted soil.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11356-020-12152-6.

Acknowledgements The authors thank the Indo-French Centre for the Promotion of Advanced Research (IFCPAR), New Delhi for financial support.

Authors' contributions RM, LFT, DM and MSR conceived the idea of the experiments and provided technical guidance. RY, BT and AM had a major contribution to overall preparation and carrying out of the research. MSR, DM, BT and AM revised and edited the manuscript.

Funding This work was supported by IFCPAR under the project 4709-1

Data availability The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

Ethics approval and consent to participate All authors declare that manuscripts reporting studies do not involve any human participants, human data, or human tissue. Therefore, it is not applicable.

Consent for publication This manuscript does not contain data from any individual person, so it is "not applicable."

References

Bailly J, Fraissinet-Tachet L, Verner MC, Debaud JC, Lemaire M, Wésolowski-Louvel M, Marmeisse R (2007) Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME J 1:632–642. https://doi.org/10.1038/ismej.2007.68

Banecki B, Liberek K, Wall D, Wawrzynów A, Georgopoulos C, Bertoli E, Tanfani F, Zylicz M (1996) Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone. J Biol Chem 271: 14840–14848. https://doi.org/10.1074/jbc.271.25.14840

Barque JP, Abahamid A, Chacun H, Bonaly J (1996) Different heat-shock proteins are constitutively overexpressed in cadmium and pentachlorophenol adapted *Euglena gracilis* cells. Biochem Biophys Res Commun 223:7–11. https://doi.org/10.1006/bbrc.1996.0837

Blindauer CA, Leszczyszyn OI (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

Cai T, Zhang Y (2000) DNaJ-like protein cDNA cloning and its expression under heavy metal stress. Adv Nature Sci 10:135–140

Cao M, Wei C, Zhao L, Wang J, Jia Q, Wang X, Jin Q, Deng T (2014) DnaJA1/Hsp40 is co-opted by influenza A virus to enhance its viral RNA polymerase activity. J Virol 88:14078–14089. https://doi.org/10.1128/JVI.02475-14

Caplan AJ, Douglas MG (1991) Characterization of YDJ1: a yeast homologue of the bacterial dnaJ protein. J Cell Biol 114:609–621. https://doi.org/10.1083/jcb.114.4.609

Chai TY, Zhang YX, Zhao WM (2000) Cloning of cDNA and expression analysis of a DnaJ-like gene under heavy metal stress in bean. Prog Nat Sci 10:198–203

Chen LX, Hu M, Huang LN, Hua ZS, Kuang JL, Li SJ, Shu WS (2015) Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. ISME J 9:1579. https://doi.org/10.1038/ismej.2014.245

Clemens S (2006) Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. Biochimie 88:1707–1719. https://doi.org/10.1016/j.biochi.2006.07.003

Conklin DS, McMaster JA, Culbertson MR, Kung C (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

Craig EA, Huang P, Aron R, Andrew A (2006) The diverse roles of J proteins, the obligate Hsp70 co-chaperone. Rev Physiol Biochem Pharmacol 156:1–21. https://doi.org/10.1007/s10254-005-0001-8

Cyr DM, Langer T, Douglas MG (1994) DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends Biochem Sci 19:176–181. https://doi.org/10.1016/0968-0004(94)90281-X

Damon C, Vallon L, Zimmermann S, Haider MZ, 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

Damon C, Lehembre F, Oger-Desfeux C, Luis P, Ranger J, Fraissinet-Tachet L, Marmeisse R (2012) Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. PLoSOne 7: e28967. https://doi.org/10.1371/journal.pone.0028967

de Menezes A, Clipson N, Doyle E (2012) Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. Environ Microbiol 14:2577–2588. https://doi.org/10.1111/j.1462-2920.2012.02781.x

Dong J, Yang QW, Sun LN, Zeng Q, Liu SJ, Pan J, Liu XL (2011) Assessing the concentration and potential dietary risk of heavy metals in vegetables at a Pb/Zn mine site. China Environ Earth Sci 64:1317–1321. https://doi.org/10.1007/s12665-011-0992-1

Fernández S, Cotos-Yáñez T, Roca-Pardiñas J, Ordóñez C (2018) Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: application to rough mountain areas in Northwest Spain. Geoderma 311:120–129. https://doi.org/10.1016/j.geoderma.2016.10.012

Field LS, Luk E, Culotta VC (2002) Copper chaperones: personal escorts for metal ions. J Bioenerg Biomembr 34:373–379. https://doi.org/10.1023/A:1021202119942

Frydman J (2001) Folding of newly translated proteins i9n vivo: the role of molecular chaperones. Annu Rev Biochem 70:603–647. https://doi.org/10.1146/annurev.biochem.70.1.603

Fusco N, Micheletto L, Dal Corso G, Borgato L, Furini A (2005) Identification of cadmium-regulated genes by cDNA-AFLP in the heavy metal accumulator *Brassica juncea* L. J Exp Bot 56:3017–3027. https://doi.org/10.1093/jxb/eri299

Gadd GM (2010) Metals, minerals and microbes: geomicrobiology and bioremediation. Microbiol 156:609–643. https://doi.org/10.1099/mic.0.037143-0

Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2:31–34. https://doi.org/10.1038/nprot.2007.13

Gillis J, Schipper-Krom S, Juenemann K, Gruber A, Coolen S, van den Nieuwendijk R, van Veen H, Overkleeft H, Goedhart J, Kampinga HH, Reits EA (2013) The DNAJB6 and DNAJB8 protein chaperones prevent intracellular aggregation of polyglutamine peptides. J Biol Chem 288:17225–17237. https://doi.org/10.1074/jbc.M112.421685

Gonzalez E, Pitre FE, Pagé AP, Marleau J, Nissim WG, St-Arnaud M, Labrecque M, Joly S, Yergeau E, Brereton NJB (2018) Trees, fungi and bacteria: tripartite metatranscriptomics of a root microbiome responding to soil contamination. Microbiome 6:53. https://doi.org/10.1186/s40168-018-0432-5

Grant S, Grant WD, Cowan DA, Jones BE, Ma Y, Ventosa A, Heaphy S (2006) Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. Appl Environ Microbiol 72:135–143. https://doi.org/10.1128/AEM.72.1.135-143.2006

Grimsrud PA, Xie H, Griffin TJ, Bernlohr DA (2008) Oxidative stress and covalent modification of protein with bioactive aldehydes. J Biol Chem 283:21837–21841. https://doi.org/10.1074/jbc.R700019200

Gupta SC, Sharma A, Mishra M, Mishra RK, Chowdhuri DK (2010) Heat shock proteins in toxicology: how close and how far? Life Sci 86:377–384. https://doi.org/10.1016/j.lfs.2009.12.015

Haap T, Schwarz S, Köhler HR (2016) Metallothionein and Hsp70 trade-off against one another in *Daphnia magna* cross-tolerance to cadmium and heat stress. Aquat Toxicol 170:112–119. https://doi.org/10.1016/j.aquatox.2015.11.008

Hageman J, Rujano MA, Van Waarde MA, Kakkar V, Dirks RP, Govorukhina N, Oosterveld-Hut HM, Lubsen NH, Kampinga HH (2010) A DNAJB chaperone subfamily with HDAC-dependent activities suppresses toxic protein aggregation. Mol Cell 37:355–369. https://doi.org/10.1016/j.molcel.2010.01.001

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381:571–580. https://doi.org/10.1038/381571a0

Hasan MK, Cheng Y, Kanwar MK, Chu XY, Ahammed GJ, Qi ZY (2017) Responses of plant proteins to heavy metal stress- a review. Front Plant Sci 8:1492. https://doi.org/10.3389/fpls.2017.01492

Hossain Z, Makino T, Komatsu S (2012) Proteomic study of b-aminobutyric acid-mediated cadmium stress alleviation in soybean. J Proteome 75:4151–4164. https://doi.org/10.1016/j.jprot.2012.05.037

Hseu ZY (2004) Evaluating heavy metal contents in nine composts using four digestion methods. Bioresour Technol 95:53–59. https://doi.org/10.1016/j.biortech.2004.02.008

Jiang C, Xu J, Zhang H, Zhang X, Shi J, Li M, Ming F (2009) A cytosolic class I small heat shock protein, RcHSP17.8, of *Rosa chinensis* confers resistance to a variety of stresses to *Escherichia coli*, yeast and *Arabidopsis thaliana*. Plant Cell Environ 32:1046–1059. https://doi.org/10.1111/j.1365-3040.2009.01987.x

Keinänen SI, Hassinen VH, Kärenlampi SO, Tervahauta AI (2007) Isolation of genes up-regulated by copper in a copper-tolerant birch (*Betula pendula*) clone. Tree Physiol 27:1243–1252. https://doi.org/10.1093/treephys/27.9.1243

Kim SA, Chang S, Yoon JH, Ahn SG (2008) TAT-Hsp40 inhibits oxidative stress-mediated cytotoxicity via the inhibition of Hsp70 ubiquitination. FEBS Lett 582:734–740. https://doi.org/10.1016/j.febslet.2008.01.053

Lehembre F, Doillon D, David E, Perrotto S, Baude J, Foulon J, Harfouche L, Vallon L, Poulain J, Da Silva C, Wincker P (2013) Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. Environ Microbiol 15:2829–2840. https://doi.org/10.1111/1462-2920.12143

Li L, Kaplan J (1998) Defects in the yeast high affinity iron transport system result in increased metal sensitivity because of the increased expression of transporters with a broad transition metal specificity. J Biol Chem 273:22181–22187. https://doi.org/10.1074/jbc.273.35.22181

Li ZS, Lu YP, Zhen RG, Szczypka M, Thiele DJ, Rea PA (1997) A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae,* YCF1Δ-catalyzed transport of bis(glutathionato) cadmium. Proc Natl Acad Sci U S A 94:42–47. https://doi.org/10.1073/pnas.94.1.42

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ method. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

Longo VD, Gralla EB, Valentine JS (1996) Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae*. J Biol Chem 271:12275–12280. https://doi.org/10.1074/jbc.271.21.12275

Lu Z, Cyr DM (1998) The conserved carboxyl terminus and zinc finger like domain of the co-chaperone Ydj1 assist Hsp70 in protein folding. J Biol Chem 273:5970–5978. https://doi.org/10.1074/jbc.273.10.5970

Marmeisse R, Kellner H, Fraissinet-Tachet L, Luis P (2017) Discovering protein-coding genes from the environment: time for the eukaryotes? Trends Biotechnol 35:824–835. https://doi.org/10.1016/j.tibtech.2017.02.003

Meacham GC, Lu Z, King S, Sorscher E, Tousson A, Cyr DM (1999) The Hdj-2/Hsc70 chaperone pair facilitates early steps in CFTR biogenesis. EMBO J 18:1492–1505. https://doi.org/10.1093/emboj/18.6.1492

Minet M, Dufour ME, Lacroute F (1992) Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. Plant J 2:417–422. https://doi.org/10.1046/j.1365-313X.1992.t01-38-00999.x

Mukherjee A, Yadav R, Marmeisse R, Fraissinet-Tachet L, Reddy MS (2019) Heavy metal hypotolerant eukaryotic aldehyde dehydrogenase isolated from metal contaminated soil by metatranscriptomic approach. Biochemie 160:183–192. https://doi.org/10.1016/j.biochi.2019.03.010

Nabulo G, Young SD, Black CR (2010) Assessing risk to human health from tropical leafy vegetables grown on contaminated urban soils. Sci Total Environ 408:5338–5351. https://doi.org/10.1016/j.scitotenv.2010.06.034

Nover L (1991) Heat shock response. CRC Press, Boca Raton

Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB, Forster RJ (2011) Snapshot of the eukaryotic gene expression in muskoxen rumen - a metatranscriptomic approach. PLoS One 6:e20521. https://doi.org/10.1371/journal.pone.0020521

Qiu XB, Shao YM, Miao S, Wang L (2006) The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell Mol Life Sci 63:2560–2570. https://doi.org/10.1007/s00018-006-6192-6\

Reddy MS, Prasanna L, Marmeisse R, Fraissinet-Tachet L (2014) Differential expression of metallothioneins in response to heavy metals and their involvement in metal tolerance in the symbiotic basidiomycete *Laccaria bicolor*. Microbiol 160:2235–2242. https://doi.org/10.1099/mic.0.080218-0

Reddy MS, Kour M, Aggarwal S, Ahuja S, Marmeisse R, Fraissinet-Tachet L (2016) Metal induction of a *Pisolithus albus* metallothionein and its potential involvement in heavy metal tolerance during mycorrhizal symbiosis. Environ Microbiol 18:2446–2454. https://doi.org/10.1111/1462-2920.13149

Ren M, Zhang Z, Wang X, Zhou Z, Chen D, Zeng H, Zhao S, Chen L, Hu Y, Zhang C, Liang Y (2018) Diversity and contributions to nitrogen cycling and carbon fixation of soil salinity shaped microbial communities in Tarim Basin. Front Microbiol 9:431. https://doi.org/10.3389/fmicb.2018.00431

Rhee JS, Raisuddin S, Lee KW, Seo JS, Ki JS, Kim IC, Park HG, Lee JS (2009) Heat shock protein (Hsp) gene responses of the intertidal copepod *Tigriopusjaponicus* to environmental toxicants. Comp Biochem Physiol C Toxicol Pharmacol 149:104–112. https://doi.org/10.1016/j.cbpc.2008.07.009

Sahi C, Lee T, Inada M, Pleiss JA, Craig EA (2010) Cwc23, an essential J protein critical for pre-mRNA splicing with a dispensable J domain. Mol Cell Biol 30:33–42. https://doi.org/10.1128/MCB.00842-09

Sahi C, Kominek J, Ziegelhoffer T, Yu HY, Baranowski M, Marszalek J, Craig EA (2013) Sequential duplications of an ancient member of the DnaJ-family expanded the functional chaperone network in the eukaryotic cytosol. Mol Biol Evol 30:985–998. https://doi.org/10.1093/molbev/mst008

Shen Y, Hendershot LM (2005) ERdj3, a stress-inducible endoplasmic reticulum DnaJ homologue, serves as a cofactor for BiP's interactions with unfolded substrates. Mol Biol Cell 16:40–50. https://doi.org/10.1091/mbc.e04-05-0434

Sun L, Liu Y, Kong X, Zhang D, Pan J, Zhou Y, Wang L, Li D, Yang X (2012) ZmHSP16.9, a cytosolic class I small heat shock protein in maize (*Zea mays*), confers heat tolerance in transgenic tobacco. Plant Cell Rep 31:1473–1484. https://doi.org/10.1007/s00299-012-1262-8

Suzuki N, Koizumi N, Sano H (2001) Screening of cadmium responsive genes in *Arabidopsis thaliana*. Plant Cell Environ 24:1177–1188. https://doi.org/10.1046/j.1365-3040.2001.00773.x

Szabo A, Korszun R, Hartl FU, Flanagan J (1996) A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. The EMBO J 15:408–417. https://doi.org/10.1046/j.1365-3040.2001.00773.x

Tamadaddi CA, Sahi C (2016) J domain independent functions of J proteins. Cell Stress Chaperones 21:563–570. https://doi.org/10.1007/s12192-016-0697-1

Teng W, Kuang J, Luo Z, Shu W (2017) Microbial diversity and community assembly across environmental gradients in acid mine drainage. Minerals 7:106. https://doi.org/10.3390/min7060106

Teste MA, Duquenne M, François JM, Parrou JL (2009) Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. BMC Mol Biol 10:99. https://doi.org/10.1186/1471-2199-10-99

Thakur B, Yadav R, Fraissinet-Tachet L, Marmeisse R, Reddy MS (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

Thakur B, Yadav R, Vallon L, Marmeisse R, Fraissinet-Tachet L, Reddy MS (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. https://doi.org/10.1016/j.scitotenv.2019.01.201

Thavamani P, Samkumar RA, Satheesh V, Subashchandrabose SR, Kavitha R, Naidu R, Venkateswarlu K, Megharaj M (2017) Microbes from mined sites: harnessing their potential for reclamation of derelict mine sites. Environ Pollut 230:495–505. https://doi.org/10.1016/j.envpol.2017.06.056

Tsai J, Douglas MG (1996) A conserved HPD sequence of the J domain is necessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding. J Biol Chem 271:9347–9354. https://doi.org/10.1074/jbc.271.16.9347

Wall D, Zylicz M, Georgopoulos C (1994) The NH2-terminal 108 amino acids of the *Escherichia coli* DnaJ protein stimulate the ATPase activity of DnaK and are sufficient for lambda replication. J Biol Chem 269:5446–5451

Wang J, Ying SH, Hu Y, Feng MG (2017) Vital role for the J-domain protein Mdj1 in asexual development, multiple stress tolerance, and virulence of *Beauveria bassiana*. Appl Microbiol Biotechnol 101:185–195. https://doi.org/10.1007/s00253-020-10659-z

Yadav RK, Barbi F, Ziller A, Luis P, Marmeisse R, Reddy MS, 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

Yu K, Zhang T (2012) Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS One 7:e38183. https://doi.org/10.1371/journal.pone.0038183

Zhang X, Li G, Yang X, Wang L, Wang Y, Guo X, Li H, Xu B (2019) Identification of a DnaJC3 gene in *Apis cerana cerana* and its involvement in various stress responses. Pestic Biochem Physiol 160:171–180. https://doi.org/10.1016/j.pestbp.2019.08.007

Zielinski BL, Allen AE, Carpenter EJ, Coles VJ, Crump BC, Doherty M, Foster RA, Goes JI, Gomes HR, Hood RR, McCrow JP, Montoya JP, Moustafa A, Satinsky BM, Sharma S, Smith CB, Yager PL, Paul JH (2016) Patterns of transcript abundance of eukaryotic biogeochemically-relevant genes in the Amazon River Plume. PLoS One 11:e0160929. https://doi.org/10.1371/journal.pone.0160929

Ziller A, Yadav RK, Capdevila M, Reddy MS, Vallon L, Marmeisse R, Atrian S, Palacios Ò, 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. https://doi.org/10.1016/j.jinorgbio.2016.11.017