
UNDERSTANDING PHYTASE THERMOSTABILITY: BIOPROSPECTING AND MODULATION

A THESIS

SUBMITTED TO

SVKM'S NARSEE MONJEE INSTITUTE OF MANAGEMENT STUDIES

(NMIMS) DEEMED-TO-UNIVERSITY

FOR THE DEGREE OF

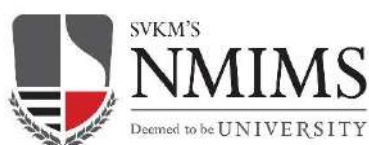
**DOCTOR OF PHILOSOPHY
IN
BIOLOGICAL SCIENCE**

BY

Ms. ASMITA DEEPAK KAMBLE

UNDER THE SUPERVISION OF

Dr. HARINDER SINGH



**SUNANDAN DIVATIA
SCHOOL OF SCIENCE**

SUNANDAN DIVATIA SCHOOL OF SCIENCE

SVKM's Narsee Monjee Institute of Management Studies (NMIMS) Deemed-to-University,

V.L. Mehta Road, Vile Parle (West), Mumbai – 400056, India.

DECEMBER 2023

UNDERSTANDING PHYTASE THERMOSTABILITY: BIOPROSPECTING AND MODULATION

A THESIS
SUBMITTED TO
SVKM'S NARSEE MONJEE INSTITUTE OF MANAGEMENT STUDIES
(NMIMS) DEEMED-TO-UNIVERSITY

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
BIOLOGICAL SCIENCES

BY
Ms. ASMITA DEEPAK KAMBLE



Project Guide
Dr. Harinder Singh



In-charge Dean
Dr. Purvi Bhatt



**SUNANDAN DIVATIA
SCHOOL OF SCIENCE**

SUNANDAN DIVATIA SCHOOL OF SCIENCE

SVKM's Narsee Monjee Institute of Management Studies (NMIMS) Deemed-to-University,
V.L. Mehta Road, Vile Parle (West), Mumbai – 400056, India.

DECEMBER 2023

DECLARATION BY THE STUDENT

This is to certify work embodied in the thesis 'UNDERSTANDING PHYTASE THERMOSTABILITY: BIOPROSPECTING AND MODULATION' for the award of the Degree of Doctor of Philosophy in Biological Sciences is my contribution to the research work carried out under the supervision of 'Dr. Harinder Singh'. The work has not been submitted for the award of any other degree/to any other University. Wherever a reference has been made to earlier reported findings, it has been cited in the thesis. The thesis fulfills the requirements of the ordinance relating to the award of the Ph.D. degree of the University.

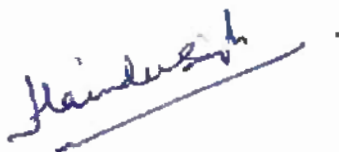


Ms. Asmita Deepak Kamble

Place: Mumbai

Date:

Forwarded by



Dr. Harinder Singh

Assistant Professor.

Department of Biological Sciences.

Sunandan Divatia School of Science.

SVKM's Narsee Monjee Institute of Management Studies (NMIMS) Deemed-to-University,

V.L. Mehta Road,

Vile Parle (West),

Mumbai – 400056, India.

CERTIFICATE

This is to certify that the work described in this thesis entitled "UNDERSTANDING PHYTASE THERMOSTABILITY: BIOPROSPECTING AND MODULATION" has been carried out by Ms. Asmita Deepak Kamble under my supervision. I certify that this is her Bonafide work. The work described is original and has not been submitted for any degree to this or any other University.



Dr. Purvi Bhatt

In-charge Dean

Date:

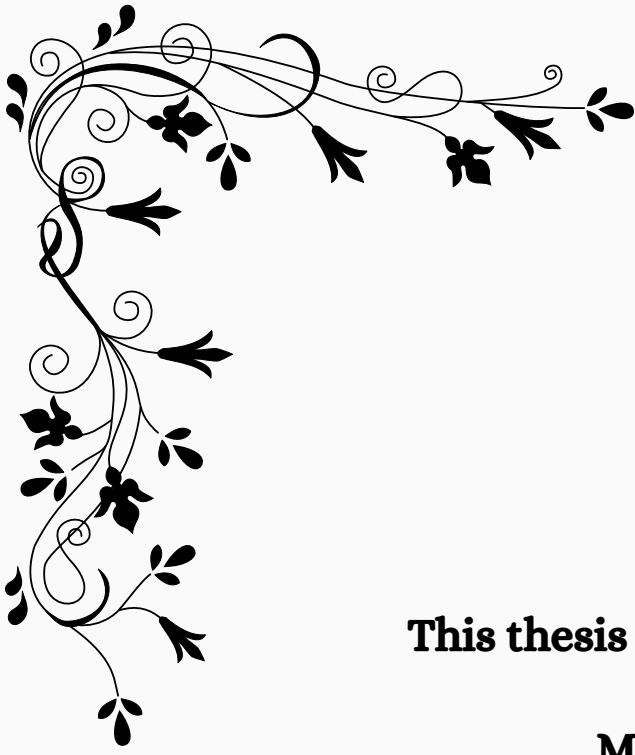


Dr. Harinder Singh

Project Guide

Date:

SUNANDAN DIVATIA SCHOOL OF SCIENCE
SVKM's Narsee Monjee Institute of Management Studies (NMIMS)
Deemed-to-University.
V. L. Mehta Road,
Vile-Parle (West),
Mumbai – 400056, India



This thesis is solely dedicated to

MY PARENTS

DEEPAK KAMBLE AND DEEPIKA KAMBLE

&

MY SISTER ANKITA

MY BROTHER SHUBHAM

MY BROTHER-IN-LAW RANJEET



Acknowledgments

In this pivotal phase and significant chapter of my life, there are individuals without whom this reality would have remained an unattainable dream. Countless people haven't just offered guidance but have actively steered and supported me, guiding, and clearing the path ahead.

I extend my gratitude to **Dr. Harinder Singh**, my mentor, for invaluable guidance and unwavering support during my doctoral studies. I'm especially grateful for his patience, always being there to listen and engage in discussions that have truly shaped the direction of my work. His encouragement during the tough times was like a steady hand helping me navigate the stormy seas. What really stood out was his mentorship style, which taught me to see challenges not as roadblocks but as chances to grow. He didn't just support my ideas; he actively pushed me to explore different angles and possibilities, creating an environment where my thoughts could flourish. I owe a lot to him for recognizing and nurturing my potential, coaxing out the best in me. Dr. Singh's influence has been pivotal, not just in my academic journey but in shaping me as a researcher and thinker.

I want to extend my heartfelt thanks to **Dr. Purvi Bhatt**, In-charge Dean and HOD. Her consistent support has been the backbone of our research work in the lab. I'd also like to express my gratitude to **Dr. Aparna Khanna**, our former Dean, for always being there with support and positive energy during our TAC meetings. A special thank you goes to **Dr. Dhananjaya Saranath**, our esteemed former Emeritus Professor at the Sunandan Divatia School of Science. Her teachings went beyond academics; she taught us how to think critically, fueled our curiosity, and emphasized the importance of discipline in our academic pursuits.

I wish to express my gratitude to my doctoral committee members, namely, **Dr. Kayzad Nilgiriwala**, serving as the Assistant Director - Research at the Foundation of Medical Research, Mumbai, and **Dr. Binoj Kutty**, an Assistant Professor at St. Xavier's College, Mumbai. Their timely assessments and valuable recommendations throughout my Ph.D. tenure significantly contributed to the development of this thesis into its present form.

I extend my sincere appreciation to **Dr. Rajkumar Singh**, a Senior Postdoctoral Researcher at the Karolinska Institute. His support wasn't just invaluable; it was the bedrock that helped navigate the tough parts of our research. His bright ideas and hands-on help were like a lifeline,

Acknowledgements

especially when we hit roadblocks in our experiments. What stood out the most was his unwavering availability and the way he cheered me up whenever things didn't go as planned. His guidance was the secret sauce that not only sorted out problems but also made this project come together. But it wasn't just about the work; he taught me something bigger—how to take on challenges with a smile and savor every step of this Ph.D. journey. That's a lesson I'll carry far beyond this thesis.

I express deep gratitude to **Dr. Virginia Stockwell** (Research Plant Pathologist at USDA) and **Dr. Megha Bedekar**, serving as the HOD and Principal Scientist at ICAR Central Institute of Fisheries Education, for their invaluable contribution of bacterial culture, which was pivotal in the successful completion of my project.

Mrs. Madhuri Khanolkar deserves a big shout-out for always having our backs. She went above and beyond, making sure we had everything we needed for our experiments on time. She's been there through thick and thin, not just sorting out supplies but also being a great support system—listening when things got tough and cheering with us during our victories. And a huge thanks to **Mrs. Manasi Naik** and **Mrs. Karuna Dhamane** for being such reliable help with all the academic and administrative staff, sorting out any questions or issues that came up. I extend my sincere appreciation to **Sushma Maushi** and **Sonu** for their invaluable assistance in arranging the necessary glassware. Their efforts were instrumental in ensuring the smooth progress of my experiments.

Just saying "thank you" doesn't capture the depth of gratitude I feel for my Ph.D. mates. Dr. Shriya Sawant has been a constant support, always ready to hear me out whether I was facing challenges, venting frustrations, or celebrating victories. Her advice and unwavering presence have been my rock through tough times. And to **Mr. Mitesh Joshi**, your jokes were the highlight of this journey, bringing laughter when we needed it most. Your reliability made everything smoother. **Ms. Ruchita Joshi's** warm hosting brought us all closer together, making our gatherings memorable. Ms. Jasmeet Kaur Viridi's practical advice and understanding ear were truly appreciated. And to **Mrs. Hiral Khakkar-Hindocha** and **Ms. Meenakshi Shukla**, our talks went way beyond studies, making this whole experience richer. Without our adventures, spontaneous plans, and shared moments – the trips, food escapades, Juhu hangouts, and the decisions we made together – this journey wouldn't have been half as vibrant.

Acknowledgements

My journey through my doctorate wouldn't have been the same without the breaks I took with these amazing friends: **Kanchan, Ashruti, and Preethika**, all part of the KAPA crew. They've been there through thick and thin, their constant support keeping me steady through life's toughest moments. I owe a world of thanks to my rock-solid support system, my #ATM: **Ms. Tejaswini Parmar and Ms. Mamta**. Their knack for simplifying the toughest challenges was a game-changer. A huge shout-out goes to **Mr. Rajesh Soni**, my unwavering supporter. His unshakeable faith in me and his "You've got this" encouragement meant everything, especially when doubts crept in. And I can't forget to express my deepest gratitude to **Mr. Jatin Upadhya, Mr. Hitesh Sonar, Mr. Mukesh Thakur, Mr. Vaibhav Shelar, and Mr. Tapan Kamath** for their unwavering encouragement and support that propelled me toward completing this thesis.

I owe everything to my amazing parents, **Mr. Deepak Kamble and Mrs. Deepika Kamble**. They've been my unwavering support system, guiding me through every step of my journey. Without their constant encouragement, I wouldn't have reached this point. My sister, **Mrs. Ankita Kamble-Wajey**, my brother, **Mr. Shubham Kamble**, and my brother-in-law, **Mr. Ranjeet Wajey**, are my rock-solid crew. They're always there, even if it means listening to my not-so-funny jokes. They're the heart of my life, shaping me into who I am today. Every day, their presence has been my motivation. I can't put into words how much their support—physical, emotional, and mental—has meant to me. They've been there through every high and low, guiding me. They mean the world to me, and I can never thank them enough for everything they've done.

Asmita Kamble

Abstract

Abstract

Enzymes are natural biocatalysts that have wide applications in different domains, such as food and beverages (Nevskaya et al., 2021), animal nutrition (Vashishth et al., 2017), therapeutic (Kaur et al., 2021), bioethanol production (Makolomakwa et al., 2017), and many more. There is a huge demand for enzymes in almost all aspects of life. One such enzyme is phytase.

Phytase, which is myoinositol hexakisphosphate phosphohydrolases, is the enzyme that hydrolyzes phytate (substrate) to release lower myoinositol, micronutrients, and inorganic phosphorous (Vashishth et al., 2017) (Tran et al., 2010). Phytate exhibits anti-nutritional activity and has the potential to cause environmental issues (Vashishth et al., 2017) (Menezes-Blackburn et al., 2013) (Kaur et al., 2021). Phytases are known to eliminate the negative effects of phytate. Hence, phytase has wide applications in the feed and food industry, aquaculture, myo-inositol phosphate production, bioethanol production, and the soil amendment process (Kaur et al., 2021).

The industrial and soil amendment applications of phytases demand highly stable phytases (Jaiwal et al., 2019) (Brinch-Pedersen et al., 2014) (*Industrial Enzymes Market - Global Forecast by 2022*, 2017) (Cang et al., 2004). Commercial phytase application is also limited due to patent protection and expensive downstream processing (Kaur et al., 2021). Hence, there is a continuous demand to search for novel stable phytases with catalytic activity to hydrolyze phytate.

The present study focused on the bioprospecting of novel phytases via different methods and engineering them to improve their characteristics via a protein engineering approach.

Three different bioprospecting methods i.e., conventional, metagenomics, and *in-silico* were employed to isolate and screen novel phytase producers from culturable, non-culturable, and database sources, respectively. In the conventional bioprospecting approach, we used different soil samples and isolated potential phytase producers based on a plate and mini-scale phytase activity assay. These potential phytase producers were identified based on the partial 16S rRNA gene amplification and sequencing. Out of the six isolates, *Klebsiella variicola* PSD soil isolate's phytase gene was further confirmed by its gene amplification and sequencing, and it belongs to the Histidine Acid Phytase. The metagenomics approach involved the isolation of total soil DNA from pig stag soil, followed by 16S rRNA gene metagenomics and analysis by

Abstract

using QIIME2 platform. The results highlighted the *Xanthomonadaceae* family as a potential phytase producer. *In-silico* bioprospecting approach involved exploring the databases, followed by removing redundancy and phylogenetic analysis. A total of 67 different candidates including *Pantoea vagans* (PV) and *Edwardsiella tarda* (ET) were identified as potential phytase producers. Hence, the combination of different bioprospecting methods resulted in the identification of a more diverse set of novel phytases.

Two potential phytases, *P.vagans* phytase (PVP) and *E.tarda* phytase (ETP) from a bioprospecting study were then subjected to *in-silico* characterization to understand features computationally and were further validated experimentally. *In-silico* characterization involved primary sequence analysis, secondary structure, and tertiary structure analysis, followed by molecular dynamics simulation at different temperatures to study protein dynamics, and stability, and identify thermo-labile regions. The analyses indicated that PVP was a better candidate as compared to ETP. To further check this, experimental validation of PVP and ETP was conducted, which included cloning and over-expression studies, followed by enzyme purification and biochemical characterization. The PVP was purified and characterized biochemically; however, the ETP was not stable in overexpression and failed to over-express even after multiple attempts with varying parameters. The optimum temperature and pH of PVP were found to be 40°C and pH 4, respectively. Both *in-silico* characterization and experimental validation indicated that PVP exhibited greater stability than ETP; hence, it was selected for further optimization using rational engineering techniques.

A rational engineering approach was undertaken with the PVP candidate, which included *in-silico* hotspot detection, followed by experimental validation. This analysis helped to identify thermos-labile residues/regions and listed mutations that were predicted to be stable compared to their wild type. A few mutants (M1 and M2) were experimentally validated by comparing them with wild type (WT). The experimental validation included cloning, over-expression, purification, and biochemical and biophysical characterization. Out of the two, M1 was found to be unstable and failed to over-express in the host system, whereas M2 was overexpressed successfully and was further characterized and compared to WT. The optimum pH of WT and M2 was found to be the same at pH 4. However, the optimum temperature of M2 was increased by 5 degrees as compared to WT. The CD data also suggested the proper folding of both WT and M2. The Isothermal Titration Calorimetry (ITC) data revealed that the binding affinity was

Abstract

not affected significantly. The Differential Scanning Fluorimetry (DSF) data revealed that the M2 was more stable than WT. So, the *in-silico* identification of hotspots followed by experimental validation resulted in a phytase with improved stability.

To summarize, the overall study involved different bioprospecting approaches to obtain novel potential phytases from different natural sources and biological databases. Around 67 candidates were identified, out of which, two of them were further studied using *in-silico* characterization and experimental validation. It also helped in shortlisting potential phytase candidates for rational engineering. The rational engineering of potential phytase candidates was performed first by *in-silico* identification of hotspots (prediction of stability by using various tools/software/servers) followed by experimental validation. The study has obtained novel thermostable phytase enzyme candidates that have wide applications in the soil amendment process, feed and food industry, aquaculture, myo-inositol phosphate production, and bioethanol production.

Table of contents

Content No.	Content	Page No.
	List of abbreviations	i-vi
	List of Figures	vii-xv
	List of Tables	xvi-xix
Chapter 1	<u>Introduction</u>	1
1.1	Phytate chemical structure	2
1.2	Problems related to phytate	3
1.3	Phytase	4
1.4	Application of phytase	4
1.5	Classification of phytase	5
1.5.1	Based on the stereospecificity of phytate hydrolysis	5
1.5.2	Based on sources	7
1.5.3	Based on catalytic mechanisms	8
1.5.4	Based on pH	9

Table of contents

1.6	Commercial phytases	9
1.7	Finding novel enzymes	11
1.8	<i>In-silico</i> characterization of novel enzymes	12
1.9	Engineering of novel enzymes	13
1.10	Problems related to the application of phytases	14
Chapter 2	<u>Review of literature</u>	15
2.1	Bioprospecting of novel enzymes	16
2.1.1	Conventional bioprospecting of enzyme	16
2.1.2	Bioprospecting of enzymes through functional metagenomic approach	18
2.1.3	<i>In-silico</i> bioprospecting of novel enzyme	20
2.2	<i>In-silico</i> characterization	35
2.2.1	Analysis of primary sequence and secondary, and tertiary structure	36
2.2.2	Phylogenetic tree	36
2.2.3	Evaluation of Function	36
2.2.4	Protein interaction studies	37

Table of contents

2.2.5	Conserved domain, motifs, domains, peptide, epitope, glycosylation, disulfide bonds, antigenicity, and localization analysis	37
2.2.6	Molecular docking and Molecular Dynamics Simulation analysis	37
2.3	Protein engineering approach	38
2.3.1	Directed evolution	38
2.3.2	Semi-rational strategy	39
2.3.3	Rational engineering strategy	40
Chapter 3	<u>Rationale, Aim, and Objectives</u>	49
3.1	Rationale	50
3.2	Aim	51
3.3	Objectives	51
Chapter 4	<u>Isolation and screening of phytase-producing microbes from natural sources or <i>in-silico</i> bioprospecting</u>	52
4.1	Introduction	53
4.1.1	Conventional Bioprospecting	53
4.1.2	Metagenomics approach	53
4.1.3	<i>In-silico</i> bioprospecting	54

Table of contents

4.2	Conventional Bioprospecting	54
4.2.1	Introduction	54
4.2.2	Materials and Methods	55
4.2.2.1	Materials (Chemicals, kits, reagents, solutions, plasticware)	55
4.2.2.2	Soil sample collection	56
4.2.2.3	Enrichment, screening, and isolation of bacteria	58
4.2.2.4	Qualitative phytase activity	59
4.2.2.5	Molecular identification of bacterial isolates	59
4.2.2.6	Quantitative effect of pH on phytase activity	60
4.2.2.7	Full-length PSD 16S rRNA gene amplification and sequencing	64
4.2.2.8	Phytase gene amplification	66
4.2.3	Results	68
4.2.3.1	Soil sample collection	68
4.2.3.2	Enrichment, screening, and isolation of bacteria	68
4.2.3.3	Qualitative phytase activity	70

Table of contents

4.2.3.4	Molecular identification of bacterial isolates	71
4.2.3.5	Quantitative effect of pH on phytase activity	75
4.2.3.6	Full-length 16 S rRNA gene sequencing	77
4.2.3.7	PSD phytase gene amplification	78
4.3	Metagenomics	80
4.3.1	Introduction	80
4.3.2	Materials and Methods	81
4.3.2.1	Materials (Chemicals, kits, reagents, solutions, plasticware)	81
4.3.2.2	Collection of soil sample	81
4.3.2.3	Processing of Soil Sample	81
4.3.2.4	Total soil gDNA extraction protocol	82
4.3.2.5	Assessing quality and quantity of the extracted total soil gDNA	88
4.3.2.6	Next-generation sequencing using Illumina Miseq	88
4.3.2.7	Hardware and software requirements	88
4.3.2.8	Taxonomic assessment using QIIME2™	89

Table of contents

4.3.3	Results	91
4.3.3.1	Soil DNA extraction	91
4.3.3.2	16S rRNA gene metagenomics	94
4.4	<i>In-silico</i> Bioprospecting	95
4.4.1	Introduction	95
4.4.2	Materials and Methods	95
4.4.2.1	Common steps for all <i>In-silico</i> bioprospecting methods	95
4.4.2.2	NCBI-probe method	96
4.4.2.3	NCBI-keyword method	96
4.4.2.4	ScanProsite method	96
4.4.2.5	UniProtKB method	97
4.4.2.6	Conserved domain database (CDD)	97
4.4.2.7	HMMER	97
4.4.2.8	JGI-metagenome	98
4.4.2.9	Final list of potential candidates	98

Table of contents

4.4.3	Results	98
4.4.3.1	NCBI-probe method	98
4.4.3.2	NCBI-Keyword search	99
4.4.3.3	Scanprosite method	100
4.4.3.4	UniprotKB method	101
4.4.3.5	Conserved domain database	101
4.4.3.6	HMMER	102
4.4.3.7	JGI-metagenome	103
4.4.3.8	Final shortlisted candidates from all databases	104
4.5	Discussion	107
4.6	Summary and Conclusion	110
Chapter 5	<u><i>In-silico</i> characterization and experimental validation of shortlisted phytase</u>	111
5.1	Introduction	112
5.1.1	<i>In-silico</i> characterization	112
5.1.2	Experimental validation	113

Table of contents

5.2	<i>In-silico</i> characterization	113
5.2.1	Introduction	113
5.2.2	Materials and Methods	114
5.2.2.1	Sequence retrieval, phylogenetic analysis, and primary sequence analysis	114
5.2.2.2	Secondary structure analysis	115
5.2.2.3	Tertiary structure analysis and validation	115
5.2.2.4	Functional analysis	116
5.2.2.5	Molecular docking and Molecular dynamics simulation	116
5.2.3	Results	119
5.2.3.1	Sequence retrieval, phylogenetic analysis, and primary sequence analysis	119
5.2.3.2	Secondary Structure Analysis	127
5.2.3.3	Tertiary structure analysis and validation	138
5.2.3.4	Functional analysis	143
5.2.3.5	Molecular docking and Molecular dynamics simulation	148
5.3	Experimental validation	156

Table of contents

5.3.1	Introduction	156
5.3.2	Materials and Methods	156
5.3.2.1	Materials (Chemicals, kits, reagents, solutions, plasticware)	156
5.3.2.2	Confirmation of phytase gene in <i>P.vagans</i> and <i>E.tarda</i>	157
5.3.2.3	Cloning of phytase gene from <i>P.vagans</i> and <i>E.tarda</i>	159
5.3.2.4	Over-expression of phytase in <i>E.coli</i> BL21(DE3) host system and purification of phytase enzyme	166
5.3.2.5	Biochemical characterization of phytase	167
5.3.3	Results	169
5.3.3.1	Confirmation of phytase gene in <i>P.vagans</i> and <i>E.tarda</i>	169
5.3.3.2	Cloning of phytase gene from <i>P.vagans</i> and <i>E.tarda</i>	173
5.3.3.3	Over-expression of phytase gene in <i>E.coli</i> BL21(DE3) host system and purification of phytase enzyme:	178
5.3.3.4	Biochemical and Biophysical characterization of PV phytase	180
5.4	Discussion	182
5.5	Summary and Conclusion	184
Chapter 6	<u>In-silico hotspot identification and experimental validation of shortlisted phytase</u>	185

Table of contents

6.1	Introduction	186
6.1.1	<i>In-silico</i> hotspot identification	186
6.1.2	Experimental validation	186
6.2	<i>In-silico</i> hotspot identification	187
6.2.1	Introduction	187
6.2.2	Materials and Methods	188
6.2.2.1	General Procedure	188
6.2.2.2	Disulfide by Design (D2D)	188
6.2.2.3	Consensus Finder, Consurf, and manual consensus (CCM) method	189
6.2.2.4	Literature	189
6.2.2.5	Hotspot Wizard (HS)	189
6.2.2.6	Fireprot (FP)	190
6.2.2.7	Common mutation between HS and F	190
6.2.2.8	CABSFLEX (Conformational Alphabet-Based Structure Flexible Fitting) + Fireprot combination method	190
6.2.2.9	FOLDX position scan	191

Table of contents

6.2.2.10	MDS + FOLDX Position Scan	191
6.2.3	Results	191
6.2.3.1	Disulfide by Design (D2D)	191
6.2.3.2	Consensus Finder + Consurf + manual consensus (CCM) method	193
6.2.3.3	Literature	193
6.2.3.4	Hotspot wizard	193
6.2.3.5	Fireprot	198
6.2.3.6	Common hotspot wizard and firepot (HS-FP) method	199
6.2.3.7	CABSFLEX (Conformational Alphabet-Based Structure Flexible Fitting) + Fireprot combination method	199
6.2.3.8	FOLDX position scan	200
6.2.3.9	MDS + FOLDX Position Scan	201
6.3	Experimental validation	202
6.3.1	Introduction	202
6.3.2	Materials and Methods	203
6.3.2.1	Materials (Chemicals, kits, reagents, solutions, plasticware)	203

Table of contents

6.3.2.2	Codon optimization, cloning, overexpression, and purification of Wildtype (WT), M1*, and M2*	203
6.3.2.3	Biochemical characterization	204
6.3.2.4	Biophysical characterization	204
6.3.3	Result	205
6.3.3.1	Overexpression and purification of wildtype and shortlisted mutant (M1 and M2) phytases	205
6.3.3.2	Biochemical characterization of wildtype, and mutant phytases	216
6.3.3.3	Biophysical characterization of wildtype, and mutant phytases	220
6.4	Discussion	221
6.5	Summary and Conclusion	224
Chapter 7	Summary and Conclusion	225
Chapter 8	Significance of the study and Future Prospects	227
	Bibliography	229
	Appendix	267
	List of Grant, Conferences, Workshops, Awards	277
	List of Publications	280

List of Abbreviations

Full name	Abbreviations
<i>Pantoea vagans</i>	PV
<i>Edwardsiella tarda</i>	ET
<i>P.vagans</i> phytase	PVP
<i>E.tarda</i> phytase	ETP
Mutant 1	M1
Mutant 2	M2
Wildtype	WT
Circular Dichroism	CD
Isothermal Titration Calorimetry	ITC
Differential Scanning Fluorimetry	DSF
Histidine acid phytase	HAPhy
β -propeller phytase	BPP
Cysteine phytase	CPhy
Purple acid phytase	PAPhy
Phytate	IP6
<i>Yersinia mollaretii</i> phytase	Ymphytase

List of Abbreviations

Root Mean Square Deviation	RMSD
Molecular Dynamic Simulation	MDS
<i>A.niger</i> phytase	Anp
<i>A.fumigatus</i> phytase	Afp
Luria Bertani	LB
Polymerase chain reaction	PCR
Plant Growth Promoting Bacteria	PGPB
DNA Databank of Japan	DDBJ
European Bioinformatics Institute of the European Molecular Biology Laboratory	EMBL-EBI
National Center for Biotechnology Information	NCBI
Joint Genome Institute	JGI
Kyoto Encyclopedia of Genes and Genomes	KEGG
Protein Families	Pfam
Class, architecture, topology and homologues	CATH
Structural Classification of Proteins	SCOP
Homology-derived Secondary Structure of Proteins	HSSP
Biological Magnetic Resonance Data Bank	BioMagResBank
Three-dimensional interacting domains	3DID

List of Abbreviations

The Carbohydrate-Active enZYmes Database	CAZy
The lipase engineering database	LED
Protease substrate specificity webserver	PROSPER
Human Pan Microbe Communities Database	HPMCD
Viral Informatics Resource for Metagenome Exploration	VIROME
Web services for metagenomic analysis	WebMGA
Naïve Bayesian Classification tool	NBC
the Structure-Function Linkage Database	SFLD
Catalytic Families	CatFam
Fungal Peroxidase Database	fPoxDB
Integrated relational Enzyme database	IntEnz
ESTerases and alpha/beta-Hydrolase Enzymes and Relatives	ESTHER Database
A peroxidase database	PeroxiBase
Plant carbohydrate-active enzymes	PlantCAZyme
ORphan ENZYme Activities	ORENZA
Self-Optimized Prediction Method	SOPMA
Computed Atlas of Surface Topography of proteins	CASTp
Protein Structure analysis	PROSA

List of Abbreviations

Qualitative Model Energy Analysis	QMEAN
L-Asparaginases	ASNases
Hypothetical proteins	HPs
Isoelectric point	pI
Grand Average of Hydropathicity	GRAVY
Iterative Threading ASSEmblY Refinement	I-TASSER
Qualitative Model Energy Analysis	QMEAN
Molecular Evolutionary Genetics Analysis	MEGA
Enzyme Function Initiative	EFI
the Search Tool for the Retrieval of Interacting Genes/Proteins	STRING
Multiple Em for Motif Elicitation	MEME
Assisted Model Building and Energy Refinement	AMBER
Chemistry at Harvard Macromolecular Mechanics	CHARMM
Nanoscale Molecular Dynamics software	NAMD
Molecular Mechanics	MM
Quantum mechanics	QM
High-Throughput Screening	HTS
Phenylmethylsulfonyl fluoride	PMSF

List of Abbreviations

Phytase Screening Broth	PSB
Phytase Screening Agar	PSA
Tryptic Soy	TS
Potassium dihydrogen phosphate	KH_2PO_4
Bovine Serum Albumin	BSA
<i>Klebsiella variicola</i>	<i>K.variicola</i>
Multiple Sequence Alignment	MSA
Phenol:Chloroform:Isoamylalcohol	PCI
Phosphate Buffer Saline	PBS
Soil lysis solution	SL
Next generation Sequencing	NGS
16S ribosomal RNA	16S rRNA
European Nucleotide Archive	ENA
Easy Sequencing in PostScript	ESript
<i>Hafnia alvei</i> phytase	HAP
<i>Klebsiella pneumoniae</i> phytase	KPP
Yet Another Scientific Artificial Reality Application	YASARA
Fold and Function Assignment System	FFAS
Qualitative Model Energy Analysis	QMEANDisCo

List of Abbreviations

Protein interaction calculator	PIC
Root Mean Square fluctuation	RMSF
Solvent Accessible Surface Area	SASA
Radius of Gyration	Rg
Single Colony PCR	SCP
Isopropyl β - d-1-thiogalactopyranoside	IPTG
Phospholipase D	PLD

List of Figures

Figure No.	Figure caption	Page No.
1. 1	Phytate structure	2
1. 2	Phytase mechanism of action: Phytase hydrolyzes phytate and releases inositol, divalent metal ions, and phosphate.	4
1. 3	Classification of phytases based on stereospecificity, sources, catalytic mechanisms, and pH.	5
1. 4	Different phytases dephosphorylate phytate in a systematic process of site-specific removal of phosphate groups.	6
4. 1	Overview of the conventional bioprospecting method used in the present study.	55
4. 2	Method of soil sample collection adopted from the National Agricultural Portal of India.	56
4. 3	Enrichment of following soil samples in PSB for 24hrs to support growth of phytase-producing microbes.	69
4. 4	The two-step counterstaining method used to confirm the ability of potential phytase producers (i.e., FD2, FD5T, FD5O, PSD, CD5, and VD5) from different soil samples to degrade phytate and additionally eliminate false positive results.	70
4. 5	Isolation of gDNA from soil isolates and <i>E.coli</i> (used as control) by using the PureLink® Genomic DNA kit.	72
4. 6	16S rRNA gene amplification of soil isolates (FD2, FD5O, FD5T).	73

List of Figures

4. 7	The Maximum Likelihood method (Jukes-Cantor model, 1000 bootstrap) was used to represent the evolutionary history of the soil isolates and closet homologs.	74
4. 8	The Neighbor-Joining method (Jukes-Cantor model, 1000 bootstrap) was used to represent the evolutionary history of the soil isolates and closet homologs.	74
4. 9	Bovine Serum Albumin (BSA) standard curve.	75
4. 10	KH ₂ PO ₄ Standard curve at pH 2.5.	75
4. 11	KH ₂ PO ₄ Standard curve at pH 5.	76
4. 12	The specific activity of phytase was compared among various samples (FD5T, FD2, FD5O, PSD5, VD5, CD5) at both pH 2.5 and pH 5, using wildtype <i>E. coli</i> BL21 (DE3) phytase as the positive control.	76
4. 13	A. Full-length PSD 16S rRNA gene PCR amplicons at different annealing temperatures. B. Purified 16S rRNA gene PCR amplicon used for Sanger's Sequencing.	77
4. 14	A. PSD full-length and mid-length phytase gene amplification at different annealing temperatures(°C).	79
4. 15	BLASTX analysis of full-length PSD phytase gene sequence.	79
4. 16	A. Snippet of the PSD phytase gene electrogram B. Snippet of the alignment constructed in Serial cloner	80
4. 17	Overview of bioprospecting of potential phytase producer via metagenomics approach.	80
4. 18	Overview of total soil gDNA extraction protocol.	82

List of Figures

4. 19	The total soil gDNA extracted from: A. Method 1 to Method 3c; B. Method 4a to Method 6b. C. Method 7; D. Method 8.	92
4. 20	Stacked bar graph depicting the Family-level distribution and Genus-level distribution of bacterial diversity present in pig sty soil sample.	94
4. 21	Overview of <i>in-silico</i> bioprospecting methodology	95
4. 22	The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from the NCBI database based on the probes (well-characterized phytases).	99
4. 23	The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from the NCBI database based on the input 'Keyword'.	100
4. 24	The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from the Scanprosite database, based on the input 'motifs'.	100
4. 25	The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from the UniprotKB database, based on the input 'E.C. number'.	101
4. 26	The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from the Conserved domain database, based on the input 'Keyword and conserved domains'.	102
4. 27	The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers via the HMMER method, based on the input 'Keyword and conserved domains'.	103

List of Figures

4. 28	The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers via the JGI-metagenome method, based on the metagenome sequences deposited in the database.	104
5. 1	Overview of <i>in-silico</i> characterization of PVP and ETP.	113
5. 2	Multiple sequence alignment generated by using ClustalO: PVP against experimentally characterized phytases.	120
5. 3	Multiple sequence alignment generated by using ClustalO: <i>Edwardsiella tarda</i> phytase against the experimentally characterized phytases.	121
5. 4	Clustal O was used to generate multiple sequence alignment of PVP against the well-characterized phytases.	122
5. 5	Clustal O was used to generate multiple sequence alignment of ETP against the well-characterized phytases (retrieved from RCSB PDB) and edited by Easy Sequencing in PostScript (EScript).	123
5. 6	The neighbor-joining phylogenetic analysis of PVP along with reference sequences was performed by using MEGA 7.0 software.	124
5. 7	The neighbor-joining phylogenetic analysis of ETP along with reference sequences was performed by using MEGA 7.0 software.	124
5. 8	The amino acid composition (%) of ETP, PVP, KPP, and HAP.	125
5. 9	Psipred was used to predict the secondary structure of PVP.	128
5. 10	Psipred was used to predict the secondary structure of ETP.	129
5. 11	Polyview2D was used to predict the secondary structure of PVP.	130

List of Figures

5. 12	Polyview2D was used to predict the secondary structure of ETP.	131
5. 13	PDBSum was used to predict the secondary structure of PVP.	132
5. 14	PDBSum was used to predict the secondary structure of ETP.	133
5. 15	Proteus2 was used to predict the secondary structure of PVP.	134
5. 16	Proteus2 was used to predict the secondary structure of ETP.	135
5. 17	Porter was used to predict the secondary structure of PVP.	136
5. 18	Porter was used to predict the secondary structure of ETP.	137
5. 19	Secondary structure analysis of ETP, PVP, KPP, and HAP by using YASARA software.	138
5. 20	The solid ribbon representation of PVP docked with phytate (brown).	141
5. 21	The solid ribbon representation of ETP docked with phytate (brown).	141
5. 22	Superimposition of the modeled tertiary structure of PVP (cyan) and KPP (brown).	142
5. 23	Superimposition of the modeled tertiary structure of ETP (Blue) and HAP (Pink).	143
5. 24	The Castp analysis of PVP structure.	144

List of Figures

5. 25	A. Displayed the sequence of ETP: grey highlighted residues that were involved in catalysis according to Castp analysis. B. The tertiary structure of ETP: Highlighted in Red is the predicted pocket.	145
5. 26	SignalP analysis of PVP.	146
5. 27	SignalP analysis of ETP.	146
5. 28	SCoop analysis of PVP	147
5. 29	SCoop analysis of ETP.	147
5. 30	BioVia Discovery studio used to generate the 2D structure of the PVP phytate complex to identify interactions.	148
5. 31	BioVia Discovery studio used to generate the 2D structure of ETP phytate complex to identify interactions.	149
5. 32	Average RMSD analysis of A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP RMSD values.	152
5. 33	Average RMSF of A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP	153
5. 34	Average Rg of A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP	154
5. 35	Average SASA of: A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP	155
5. 36	Overview of the experimental validation.	156
5. 37	Isolation of gDNA from A. <i>P.vagans</i> , B. from <i>E.tarda</i> by using the PureLink® Genomic DNA kit.	169

List of Figures

5. 38	A. Full-length <i>P.vagans</i> phyk gene PCR amplicons at different annealing temperatures. B. Full-length <i>E.tarda</i> appA gene PCR amplicons at different annealing temperatures.	170
5. 39	Gel purified A. <i>P.vagans</i> phytase gene PCR amplicon , B. <i>E.tarda</i> phytase gene used for Sanger’s sequencing.	171
5. 40	A. Snippet of the <i>P.vagans</i> phytase gene electrogram B. Snippet of the alignment constructed in Serial cloner: Seq_1: Sanger’s sequenced <i>P.vagans</i> full-length phytase gene aligned with the Seq_2: Nucleotide sequence of <i>P.vagans</i> (ENA database) which was used for primer designing. C. Snippet of the <i>E.tarda</i> phytase gene electrogram D. Snippet of the alignment constructed in Serial cloner: Nucleotide sequence of <i>E.tarda</i> from ENA database which was used for primer designing aligned with the Seq_2: Sanger’s sequenced <i>E.tarda</i> full-length phytase gene.	172
5. 41	Full-length phytase gene from: A. <i>P.vagans</i> and, B: <i>E.tarda</i> flanked with NcoI and NotI restriction sites at the different annealing temperatures.	173
5. 42	pET29-b plasmid isolated by using GeneJET Plasmid Miniprep kit, Thermo Fisher.	174
5. 43	Restriction digestion with NcoI and NotI restriction enzymes of the following samples: A. purified <i>P.vagans</i> phyk gene, and isolated pET29b, B. Purified <i>E.tarda</i> appA gene and isolated pET29b.	175
5. 44	Single Colony PCR of: A. phyk-pET29b recombinant construct (from <i>P.vagans</i>) and, B. appA-pET29-b construct (from <i>E.tarda</i>).	176
5. 45	A. Snippet of the phyk-pET29-b (from <i>P.vagans</i>) gene electrogram: B. Snippet of the alignment constructed in Serial cloner.	177
5. 46	PVP purified by using Immobilized Metal affinity chromatography technique (IMAC) via His60SF resins and resolved on 10% SDS-PAGE gel: A. Elution fractions (E1-E8), B. Elution fraction 9, along with Flowthrough (FT), and Wash fraction (Wash).	178

List of Figures

5. 47	Purified PVP after dialysis (Fraction 1 and 2) resolved on 10% SDS-PAGE gel.	179
5. 48	ETP purified by using Immobilized Metal affinity chromatography technique (IMAC) via His60NiSF resins and resolved on 10% SDS-PAGE gel: A. Elution fractions (E1-E8), B. Elution fraction 9, along with Flowthrough (FT), and Wash fraction (Wash).	179
5. 49	Effect of different pH buffer (3, 3.5, 4, 4.5, 5, 5.5, and 6) conditions on PVP activity.	180
5. 50	Effect of different temperature conditions on PVP.	180
5. 51	KH_2PO_4 was used as a standard to analyze the release of inorganic phosphate at different pH conditions (pH3, 3.5, 4, 4.5, 5, 5.5, and 6).	181
6. 1	Overview of in-silico hotspot identification by hotspots by various methods.	187
6. 2	Overview of the experimental validation of WT, M1, and M2.	202
6. 3	A. The wildtype recombinant construct: A. codon optimized wildtype phytase gene cloned in pET29b vector. B. The complete nucleotide sequence of the wildtype recombinant construct (pET29b cloned with the optimized wildtype phytase gene (red highlighted)). C. the codon-optimized wild-type phytase protein sequence.	208
6. 4	A. The M1 recombinant construct: A. codon optimized M1 phytase gene cloned in pET29b vector. B. The complete nucleotide sequence of the M1 recombinant construct (pET29b cloned with the optimized M1 phytase gene (red highlighted)). C. the optimized M1 phytase protein sequence.	212
6. 5	A. The M2 recombinant construct: A. codon optimized M2 phytase gene cloned in pET29b vector. B. The complete nucleotide sequence of the M2 recombinant construct (pET29b cloned with	215

List of Figures

	the optimized M2 phytase gene (red highlighted)). C. the optimized M2 phytase protein sequence.	
6. 6	A. Purified WT and M2 resolved on 10% SDS-PAGE gel. B. Western blot analysis of purified WT and M2 system.	216
6. 7	KH ₂ PO ₄ was utilized as a benchmark to measure the liberation of inorganic phosphate across varying pH levels (pH 2, 3, 4, 4.5, 5, 5.5, 6, 7, and 8).	217
6. 8	Effect of different pH conditions on purified PVP.	218
6. 9	Effect of different temperature conditions on purified PVP activity.	219
6. 10	Effect of substrate concentration on purified PVP activity: A.wildtype PVP, B. M2.	219
6. 11	Circular Dichroism analysis of A. WT, B.M2.	220
6. 12	Differential scanning fluorimetry analysis of WT and M2.	221

List of Tables

Table No.	Table caption	Page no.
1. 1	List of commercial phytases from microbial sources	10
2. 1	Online Databases/tools used for <i>in-silico</i> bioprospecting	21
2. 2	List of a few phytase engineering studies:	43
4. 1	The details of soil samples collected from different regions of Maharashtra and Gujarat.	57
4. 2	Phytase screening broth (PSB) composition.	58
4. 3	PCR reaction mixture to amplify 16S rRNA gene.	60
4. 4	16S rRNA gene PCR amplification conditions.	60
4. 5	Reaction setup for KH_2PO_4 standardization.	61
4. 6	Reaction setup for test samples.	62
4. 7	Reaction setup for BSA standardization.	63
4. 8	The details of different <i>Klebsiella variicola</i> strain 16S rRNA gene sequences used for designing 16S rRNA primers, to amplify full-length 16S rRNA gene of PSD isolate.	65

List of Tables

4. 9	PCR reaction mixture to amplify phytase gene from PSD.	66
4. 10	PSD phytase gene PCR amplification conditions.	67
4. 11	Measurement of the zone of hydrolysis produced by potential soil isolates.	71
4. 12	Concentration and quality estimation of gDNA isolated from different soil isolates and <i>E. coli</i> .	71
4. 13	Blastn analysis of 16S rRNA gene of soil isolates.	73
4. 14	BLASTn analysis of 16S rRNA full-length gene.	78
4. 15	Total soil gDNA quantification by using BioTek Epoch Microplate spectrophotometer.	93
4. 16	The list of final shortlisted candidates.	105
5. 1	The Protparam tool was used for the primary sequence analysis of PVP, KPP, ETP, and HAP.	126
5. 2	PDBsum analysis of PVP and ETP.	127
5. 3	Preliminary screening of modeled structures by using the Structure Assessment tool provided by the EXPASY web portal.	139
5. 4	Tertiary structure validation of shortlisted Swiss model and template structures by using ProSa, ERRAT, and QmeanDisco.	142
5. 5	Residues present in the 1st pocket and predicted to be involved in catalysis according to the Castp analysis of PVP structure.	144

List of Tables

5. 6	Residues present in the 1st pocket and predicted to be involved in catalysis according to the Castp analysis of ETP structure.	145
5. 7	Primers used to amplify the full-length phytase gene of <i>P.vagans</i> and <i>E.tarda</i> .	157
5. 8	PCR reaction mixture to amplify phytase gene.	158
5. 9	Phytase gene PCR amplification conditions.	158
5. 10	Primers used to amplify the phytase gene of <i>P.vagans</i> and <i>E.tarda</i> with flanked restriction sites at both ends.	159
5. 11	PCR reaction mixture to amplify phytase gene with flanked restriction sites at both ends.	160
5. 12	Phytase gene PCR amplification conditions.	161
5. 13	Restriction digestion of insert and vector (pET29b).	162
5. 14	Ligation reaction setup.	163
5. 15	Single colony PCR reaction mixture.	165
5. 16	Single colony PCR reaction conditions.	165
5. 17	Reaction setup for KH_2PO_4 standardization.	168
5. 18	Standard reaction setup for estimating phytase activity of purified phytase.	168

List of Tables

6. 1	Disulfide by Design (D2D) was used to identify hotspots in PVP.	192
6. 2	Consensus Finder, Consurf, and manual consensus (CCM) method was used to identify hotspots in PVP.	194
6. 3	Literature-based <i>in-silico</i> identification of hotspots in PVP.	195
6. 4	Hotspot Wizard (HS) was used for identification of hotspots in PVP.	197
6. 5	Fireprot was used for in-silico identification of hotspots in PVP:	198
6. 6	Common mutations from Hotspot Wizard and Fireprot analysis.	199
6. 7	CABSFLEX was used for in-silico identification of hotspots in PVP.	200
6. 8	FOLDX position scan used for <i>in-silico</i> identification of hotspots in PVP.	201
6. 9	Common hotspots identified by using molecular dynamics simulations (MDS) + FOLDX position scan analysis.	202
6. 10	Binding affinity of WT and M2 by using Isothermal titration calorimetry (ITC) experiment.	220

Chapter 1

Introduction

Introduction

Enzymes, as natural catalysts accelerating chemical reactions, find application across a multitude of industries (Sanchez & Demain, 2011), including food and beverages (Marco Alexander Fraatz, 2014), animal nutrition (Ojha et al., 2018), cosmetics (Singhania et al., 2015), pharmaceuticals (Basso & Serban, 2019), and many more. Enzymes exhibiting improved expression, activity, substrate specificity, and functional stability are particularly suitable for various industrial applications (Wang et al., 2008) (Singhania et al., 2015). One such enzyme is phytase which hydrolyzes phytate (Vashishth et al., 2017).

1.1 Phytate chemical structure:

Phytate, also known as myo-inositol-hexakis-dihydrogenphosphate ($C_6H_{18}O_{24}P_6$), was identified by Hartig between 1855 and 1856. Phytate refers to a poly-anionic, and hexaphosphoric ester of the hexahydric cyclic alcohol meso-inositol molecule known as IP6 (**Figure 1. 1**) (Balwani et al., 2017) (Song, El Sheikha and Hu, 2019) (Gocheva et al., 2023). Phytic acid holds 12 replaceable protons and features 6 reactive sites, contributing to its strong acidity. Phytic acid carries a substantial negative charge at a wide range of pH (Gocheva et al., 2023) (Outchkourov & Petkov, 2019).

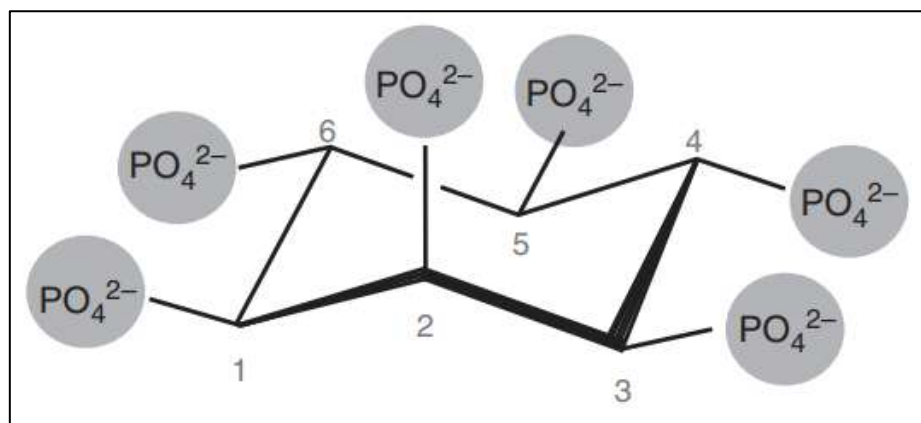


Figure 1. 1 Phytate structure: Bernard Agranoff introduced the turtle-based rule for numbering phytate, where the 2-phosphate axial points upward. Following the D configuration, the carbon atoms are numbered counterclockwise around the ring, as advised by the Nomenclature Committee of the International Union of Biochemistry (Adopted from (Outchkourov & Petkov, 2019)).

1.2 Problems related to phytate:

Phytate represents around 60-70% of the stored phosphorus in germinating cereals, legumes, nuts, and oilseeds (Song, El Sheikha and Hu, 2019). It accounts for 80% of the total organic phosphate pool in soil (Menezes-Blackburn et al. 2013). Around 60-80% of the overall phosphorus-bound phytate is contained within formulated diets for poultry and swine (Outchkourov & Petkov, 2019). It is indigestible in monogastric animals due to the absence of phytase enzymes. This indigested phytate and excess phosphate supplementation in feed can cause environmental pollution through manure (Outchkourov & Petkov, 2019) (Vashishth et al., 2017). It forms complexes in the soil and binds to metal ions which play a vital role in soil fertility and plant growth (Moushree Pal Roy, Subhabrata Datta, 2017) (Outchkourov and Petkov, 2019). It binds with divalent and trivalent cations, like calcium, iron, zinc, and magnesium, as well as trace minerals such as manganese, copper, and molybdenum, reducing their accessibility. This significantly lowers the nutritional quality of the diet (Outchkourov & Petkov, 2019) (Gocheva et al., 2023). To compensate for this, inorganic phosphate and crucial ions like calcium and iron need to be added to feed and food formulations to support animal growth (Lei, 2010) (Goutami Banerjee, Khin Oo, XIyun Zhang, Jie Yang, 2017). It creates complexes with minerals, forming mineral-phytate complexes that remain unabsorbed in the gut and are consequently excreted without being utilized (Outchkourov & Petkov, 2019) (Gujar, 2014). It interacts electrostatically with proteins, leading to the alteration of protein structure, and reduced solubility, activity, and digestibility. This interaction limits access to proteases, resulting in inefficient protein digestion (Vashishth et al., 2017). It binds to enzymes, reduces their activity, and decreases their availability. It also binds to starch and forms starch complexes. It forms lipo-phytic by binding to lipids (Gujar, 2014). The insoluble and indigestible nature of phytate complexes with proteins, amino acids, cations, and minerals makes them challenging for human absorption. This might contribute to malnutrition in Asian nations since 75% of their total calorie consumption comes from cereals (rich in starch and phytate) (Song, El Sheikha and Hu, 2019).

Phytate degradation is important because it can reduce the feed cost which accounts for 60-70% of the cost of livestock farming by reducing the supplementation of inorganic phosphorous as phytate is a rich source of inorganic phosphate (*Industrial Enzymes Market - Global Forecast by 2022*, 2017). Degradation of phytate can also increase the bioavailability of micronutrients in the feed and increase its nutritional value for better growth performance in the poultry industry. Besides the feed industry, phytate degradation is important for increasing

soil fertility, the growth of plants, and enhancing bioethanol production (K. Bhavsar & Khire, 2014) (Makolomakwa et al., 2017) (K. P. Bhavsar, 2012).

1.3 Phytase:

Phytase uses the phytate complex as its substrate and catalyzes the breakdown of O-P bonds within phytate, resulting in the release of products like inositol, intermediate derivatives of inositol, along with phosphate and other micronutrients (**Figure 1. 2**) (Vashishth et al., 2017) (Tran et al., 2010) (Gocheva et al., 2023). Suzuki et al. documented the presence of phytase activity in rice bran in 1907 (Outchkourov & Petkov, 2019). It reduces the phytate content from plant-based food without reducing the amount of minerals (Gocheva et al., 2023)

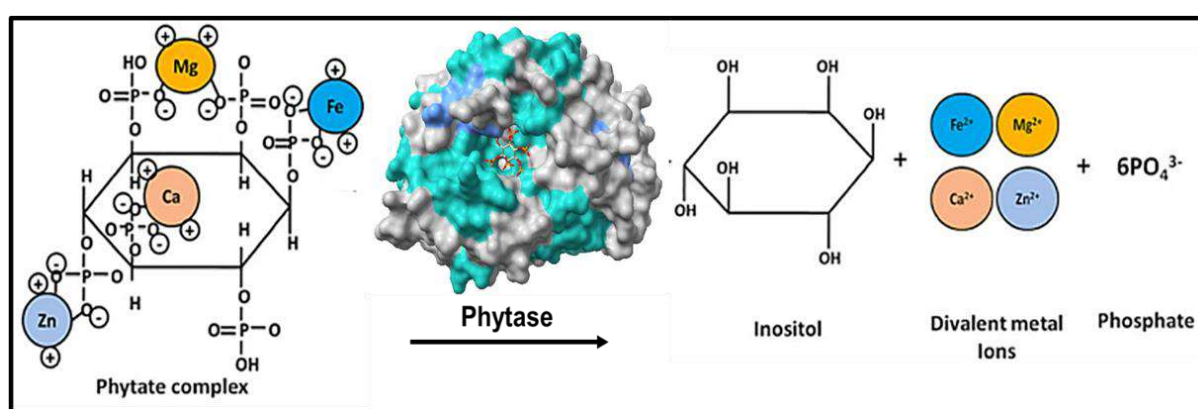


Figure 1. 2 Phytase mechanism of action: Phytase hydrolyzes phytate and releases inositol, divalent metal ions, and phosphate.

1.4 Application of phytase:

Phytase, alongside enzymes such as xylanase, β -glucanase, cellulase, mannanase, and α -galactosidase, constitutes a significant portion of the worldwide feed enzyme market (Outchkourov & Petkov, 2019). Phytases have wide applications in the feed and food industry e.g., the addition of phytase in chickpea flour can lead to enhancement of mineral mobilization such as 20-28% of Zn²⁺, 26-37% of Fe²⁺, and 24-42% Ca²⁺ leading to 75-88% reduction in phytate content (Song et al., 2018). It is also used in bread making process (Gocheva et al., 2023). Phytases are also applied aquaculture industry as well as to produce bioethanol and lower myoinositol phosphates and play a role in soil amendment (K. Bhavsar & Khire, 2014) (Makolomakwa et al., 2017) (K. P. Bhavsar, 2012).

1.5 Classification of phytase:

Phytases are classified based on stereospecificity of phytate hydrolysis (3/1 or 4/6), sources of isolation, optimum pH (acidic/alkaline), and catalytic mechanisms: histidine acid phytases (EC 3.1.3.2), β -propeller (EC 3.1.3.8), cysteine phosphatases and purple acid phosphatases (EC 3.1.3.2) (Balwani et al., 2017) (Menezes-Blackburn et al., 2013) (Gocheva et al., 2023) (Figure 1. 3). Phytases are mainly isolated from plant and microbial sources (Kumar and Sinha, 2018).

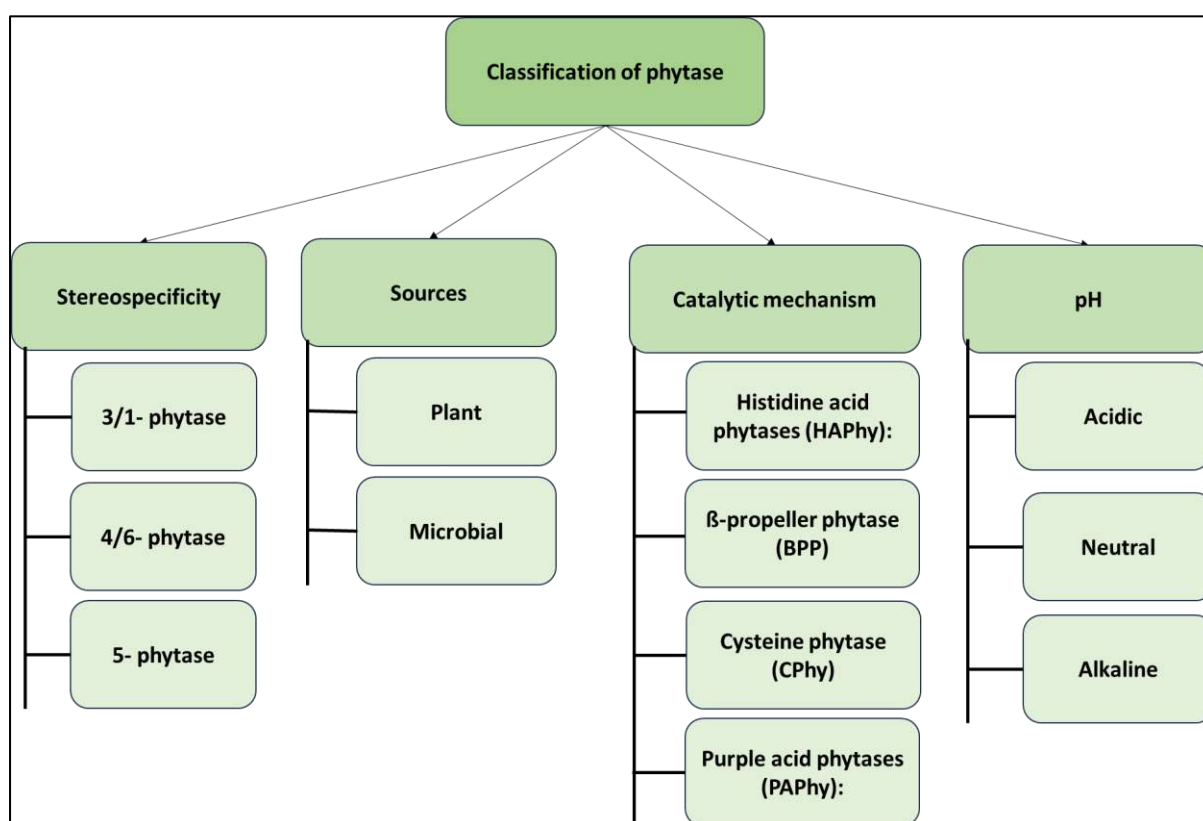


Figure 1. 3 Classification of phytases based on stereospecificity, sources, catalytic mechanisms, and pH.

1.5.1 Based on the stereospecificity of phytate hydrolysis:

Phytate (IP₆) comprises a myoinositol core with six dihydrogen phosphate groups. Phytase enzymes facilitate the dephosphorylation of phytate in a systematic process of site-specific removal of phosphate groups (Figure 1. 4). As per the classification by the International Union of Pure and Applied Chemistry and the International Union of Biochemistry (IUPAC-IUB), phytases are categorized according to their dephosphorylation activities into three types: 3-phytases (EC 3.1.3.8, myoinositol-hexakisphosphate 3-phosphohydrolase), 4/6-phytases (EC

3.1.3.26, myoinositol-hexakisphosphate 4-phosphohydrolase), and 5-phytases (EC 3.1.3.72, myoinositol-hexakisphosphate 5-phosphohydrolase) (Outchkourov & Petkov, 2019) (Gocheva et al., 2023).

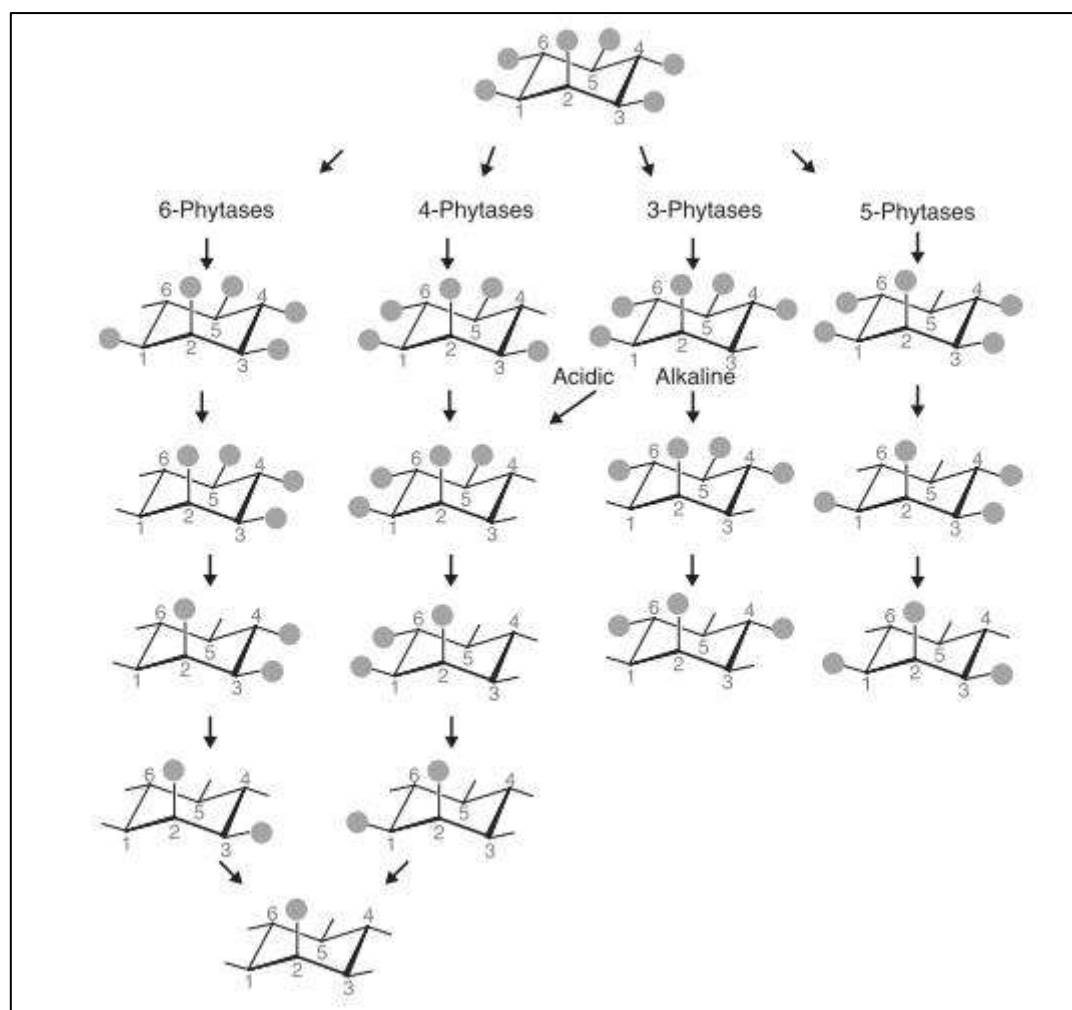


Figure 1. 4 Different phytases dephosphorylate phytate in a systematic process of site-specific removal of phosphate groups (Adopted from (Outchkourov & Petkov, 2019)).

- 3-phytase (Myo-inositol hexakisphosphate 3-phosphohydrolase): This category of phytase triggers the dephosphorylation of phytate at the third position and is derived from microbial sources like *Klebsiella* sp. ASR-1. 3-phytases excel over 6-phytases due to the dephosphorylation of phytate, resulting in IP-1 with 3-phytase and IP-4 along with other lower esters with 6-phytase. Hence, there is a complete degradation of phytate (Balwani et al., 2017) (Gocheva et al., 2023).

- 6-phytase (Myo-inositol hexakisphosphate 6-phosphohydrolase): 6-phytases instigate the dephosphorylation of phytate at the sixth position and are obtained from microorganisms, like the *E.coli* (Balwani et al., 2017) (Gocheva et al., 2023).
- 5-phytase (Myo-inositol hexakisphosphate 5-phosphohydrolase): This variant of phytase starts the dephosphorylation of phytate at the fifth position and is primarily extracted from plant sources found in *Medicago sativa* (Balwani et al., 2017) (Gocheva et al., 2023).

1.5.2 Based on sources:

Phytase, also recognized as myo-inositol-hexakisphosphate-phosphohydrolase, is found in plants, and various microbial sources. Both endogenous and exogenous forms of phytases exist, yet microbial exogenous phytases are extensively utilized commercially due to their superior pH and thermal stability, along with higher specific activity (Song et al., 2018). Microbial-derived exogenous phytases allow for the total breakdown of phytate, whereas endogenous phytases typically reduce phytate content by 73-80% (Song, El Sheikha and Hu, 2019).

- Plant sources: Phytases predominantly found in plants belong mainly to the 6-phytase category. Varieties like wheat, barley, and rye have demonstrated higher activity levels in comparison to maize, millet, and sorghum. Additionally, phytases have been identified and characterized in various plants like rice, rapeseed, and potato, among others. The utilization of plant-derived phytases is constrained by the absence of a cost-efficient and effective production method. The genetically modified rice seeds were engineered to contain a heat-resistant phytase derived from the yeast *Aspergillus fumigatus*, aiming to enhance the availability of iron for human consumption. (Balwani et al., 2017) (Gocheva et al., 2023).
- Microbial phytase: Microbial phytases, are derived from bacteria like *E.coli*, *Citrobacter*, *Enterobacter*, and many more, yeasts such as *Schwanniomyces castellii*, *Hansenula polymorph*, and many more, and various fungal species like *Aspergillus* species are explored to estimate pH, retain their activity between pH 3 and 8, and exhibit resistance to heat. Microbial phytases are favored over plant-derived ones due to their superior activity across a broad pH range and increased thermal stability (Balwani et al., 2017) (Gocheva et al., 2023).

1.5.3 Based on catalytic mechanisms:

Based on the catalytic mechanisms phytases are classified as histidine acid phytases, β -propeller, cysteine phosphatases, and purple acid phosphatases.

➤ Histidine acid phytases (HAPhy) (EC 3.1.3.2)

HAPhy, prevalent in animals, plants, and microorganisms, shares common characteristics in commercial phytases derived from microbes. It showcases a sizable α/β -domain along with a compact α -domain. Its structure features an initial RHGXRXP sequence at the N-terminus and a concluding HD motif at the C-terminus. Its mechanism involves the creation of a covalent phospho-histidine intermediate, where histidine from the RHGXRXP motif initiates a nucleophilic attack on the phosphorus. In the subsequent phase, the aspartic acid within the HD motif functions as a proton donor to the oxygen atom in the phosphomonoester bond (Outchkourov & Petkov, 2019) (Gocheva et al., 2023).

➤ β -propeller phytase (BPP) (EC 3.1.3.8)

BPP, primarily found in *Bacillus subtilis* and *Bacillus amyloliquefaciens*, features a propeller structure comprised of six blades. It displays an optimal pH at the basic level and demonstrates higher resistance to changes in temperature. Its substrate preference leans towards the calcium–phytate complex. In the crystal structure of BPP, the active site tightly binds two phosphates and four calcium ions. The breakdown of phytate by BPP generates two distinct myoinositol trisphosphates: Ins(2,4,6)P₃ and Ins(1,3,5)P₃. To enhance the efficacy of phytases in poultry, HAPhy from *E. coli* can be paired with BPP from *Bacillus*. Despite its capabilities, BPP showcases lower specific activity compared to HAPhy and is currently not utilized at a commercial scale (Outchkourov & Petkov, 2019) (Gocheva et al., 2023).

➤ Cysteine phytase (CPhy):

CPhy originates mainly from *Selenomonas ruminantium*, an anaerobic bacterium found in the rumen. Structurally, this phytase comprises a large and a small domain. The active site is situated along the border of the large domain, forming a shallow pocket to bind the substrate. The structural support for the active site domain comes from the P-loop and the WPD loop. Within the P-loop, there is a conserved motif, HCXXGXXR(T/S). CPhy's catalytic capability involves cleaving the 5-phosphate position of phytate, generating products such as IP₂ and several variations of inositol phosphates: IP (1,2,3,4,6), IP (1,2,3,6), IP (1,2,3), and IP (1,2) (Outchkourov & Petkov, 2019) (Gocheva et al., 2023).

- Purple acid phytases (PAPhy) (EC 3.1.3.2):

PAPhy, a homolog of purple acid phosphatases, belongs to the group of metalloenzymes characterized by a di-nuclear catalytic center of Fe^{3+} , Zn^{2+} , Mg^{2+} , or Mn^{2+} . It's predominantly present in soybean seedlings and various other plant species. However, its phytase activity is relatively low (Outchkourov & Petkov, 2019) (Gocheva et al., 2023).

1.5.4 Based on pH:

- Acidic phytase: Acid phytases catalyze the breakdown of phytate, resulting in the production of IP₂ as the final product. These phytases display a preference for degrading equatorial phosphate groups rather than axial ones. They are active in a pH range of 2-6 (Outchkourov & Petkov, 2019) (Gocheva et al., 2023):
- Neutral phytases: It is isolated from thermotolerant *A. flavus* and *B. subtilis* subsp. *subtilis* JJBS250, and *Bacillus nealsonii* ZJ0702. It is isolated from a few micro-organisms. It is active at pH 7-8 (Outchkourov & Petkov, 2019) (Gocheva et al., 2023):
- Alkaline phytases: Alkaline phytases, primarily sourced from *Bacillus*, target the removal of phosphate groups specifically from the C-2 position. These are found in plants such as *Typha latifolia* and *Lilium longixorum* and micro-organisms such as *Bacillus amyloliquefaciens*, and *B. laevolacticus* (Outchkourov & Petkov, 2019) (Gocheva et al., 2023):

1.6 Commercial phytases:

Commercial phytase development began in 1962 through the efforts of International Minerals & Chemicals (IMC), where they extracted the phytase from *A.niger* NRRL 3135. Mullaney et al. on the other hand, reported the cloning of the PhyA gene and its overexpression, conducted by a team from TNO in Rijswijk, The Netherlands. Subsequently, the first commercial phytase, Natuphos, emerged from the collaboration between Gist-Brocades in The Netherlands (now DSM) and BASF in Germany, marking the inception of the first generation of phytases. The inception of the second generation of phytases traces back to the efforts of Dassa and colleagues, who focused on the *E.coli* appA phytase. Their work involved the cloning of the *appA* gene in 1990. Subsequently, Greiner et al. conducted a biochemical characterization of the *E.coli* appA phytase, determining it to be a 6-phytase. In 1999, research on the *appA* gene revealed its overproduction in the *Pichia pastoris* yeast system, proving its predominant phytase activity over acid phosphatase. The second generation of phytases was developed by Rodriguez et al. at Cornell University. The OptiPhos phytase was initially commercialized by

Chapter 1

JBS United in the United States and later acquired by Huvepharma, expanding its distribution across Europe and globally. This increase in demand for inorganic phosphorus led to a surge in phytase demand, prompting the marketing of several phytases: Phyzyme XP (Danisco), Finase EC (AB Vista), Quantum, and Quantum Blue (AB Vista). Additionally, the *Citrobacter* 6-phytase, overexpressed in *A.niger*, entered the market as RONOZYME HiPhos (DSM/Novozymes). Moreover, the *Peniophora lycii* 6-phytase was introduced under the names Ronozyme P and Ronozyme NP by DSM/Novozymes (Outchkourov & Petkov, 2019). The details of commercial phytases are listed below in **Table 1. 1**.

Table 1. 1 List of commercial phytases from microbial sources:

Commercial name of phytase	Company	Source of phytase
Natuphos™	BASF	<i>Aspergillus niger</i> NRRL 3135
Ronozyme P	Novozymes/ DSM alliance	<i>Peniophora lycii</i>
Optiphos	Huvepharma	<i>E.coli</i>
PhyzymeXP	Danisco, Brabrand	<i>E.coli</i>
Allyzyme	Alltech	<i>A.niger</i>
NatuPhos E	BASF	Hybrid phytase: three bacterial sources
Ronozyme NP	Novozymes/ DSM alliance	Modified <i>Peniophora lycii</i>
Ronozyme HiPhos	Novozymes/DSM alliance	<i>Citrobacter braakii</i>
Finase EC	AB Vista	<i>Escherichia coli appA</i>
Quantum	AB Vista	Modified <i>Escherichia coli appA</i>
Quantum Blue	AB Vista	Modified <i>Escherichia coli appA</i>
Axtra Phy	Danisco	<i>Buttiauxella spp</i>

1.7 Finding novel enzymes:

Researchers have adopted different methods to explore novel enzymes to meet industrial demands. These methods included the use of natural sources (conventional), existing databases (*in-silico* bioprospecting), and metagenomics methods. In general, the conventional method involves searching for new enzymes in natural sources (Castillo Villamizar et al., 2019), while the computational method utilizes existing databases to identify enzymes (Tan et al., 2016). These novel enzymes are further subjected to protein engineering approaches (directed evolution/ rational/ semi-rational approach) to enhance their properties (Zhang et al., 2019) (K. Kumar et al., 2015) (Sadeghian et al., 2020) (Soh et al., 2017) (Chen et al., 2015).

A few studies that adopted conventional methods to explore novel enzymes are highlighted here: Kumar et.al. isolated phytase producers from environmental samples i.e., Himalayan soils, and identified them as *Acromobacter* sp. PB-01, *Tetra- thiobacter* sp. PB-03 and *Bacillus* sp. PB-13. They helped in the growth and phosphorous uptake of *Brassica juncea* (Kumar et al., 2013). Puppala et.al. characterized a novel acidic and thermostable phytase isolated from *Streptomyces* sp. (NCIM 5533) from Sanjivani islands Maharashtra, India. It also acts as a plant growth-promoting bacteria which was evident as it supports the growth of *Solanum lycopersicum* (Puppala et al., 2019). Amritha and Venkateswaran reported a phytase from *Lactobacillus plantarum* MTCC 1325 that can degrade phytate from sorghum. It also tolerates acid, and bile and retains activity after exposure to a simulated gastrointestinal fluid environment (Amritha and Venkateswaran, 2017). Similarly, enzymes such as novel fibrinolytic enzyme was isolated from *Bacillus atrophaeus* V4 (Varol et al., 2023), and many more were isolated from different natural sources. In general, the conventional bioprospecting approach favors the cultivation of microbes that can be grown in the laboratory.

Soil comprises soil biomass, minerals with diverse sizes, shapes, and chemical characteristics, and organic matter. The largest part of the soil mass is made up of prokaryotes. However, the prokaryotic cells in the soil mass vary between samples; for example, one gram of forest soil contains around 4×10^7 prokaryotic cells, while one gram of cultivated soil and grasslands contains about 2×10^9 prokaryotic cells. There are two approaches to investigating this extensive microbial diversity in soil: cultivation-dependent methods and cultivation-independent methods. Cultivation-based methods can only explore a fraction of the total microbial population, estimated at 0.1-1% in soil. To address this limitation, cultivation-independent methods like metagenomics have been developed (Daniel, 2005). In the soil

metagenomics process, the initial phase involves extracting total genomic DNA from the soil, followed by amplification of the 16S rRNA gene to investigate the range of microbial communities present (A. Kamble et al., 2020). These microbial communities can be explored in search of new enzymes like phytases. For instance, Tan et al. discovered novel phytase genes in agricultural soil using a metagenomics-based approach (Tan et al., 2014).

To successfully employ the metagenomics approach, it is crucial to have well-optimized protocols at every step, starting from the isolation of total DNA to library construction. These optimized protocols enable the screening of positive clones that possess the desired property (A. Kamble et al., 2019) (A. D. Kamble & Singh, 2021). *In-silico* bioprospecting method helped to overcome the limitations of conventional and metagenomics approaches.

In-silico bioprospecting approach: Database exploration is another means to uncover new enzymes. For instance, Tan et al. delved into the IMG/M database and identified 11 potential HAP phytase genes from 18 publicly accessible fungus garden metagenomes. Among these, the rPhyXT52 phytase displayed the highest activity (Tan et al., 2016b). In a separate investigation, Tan et al. utilized metagenomic information from an acidic peat-soil microbiome located in northeastern Bavaria, Germany. They employed Pfam00328 (Pfam identifier) to search for new phytases (Tan et al., 2016a). Similarly, novel enzymes such as nitrilases, aldehyde-deformylating oxygenase, and hypothetical proteins were explored through an *in-silico* bioprospecting approach (A. Kamble et al., 2019)

1.8 *In-silico* characterization of novel enzymes:

To get insight into the structure of enzymes, some researchers utilize computational tools, e.g., Pramanik et.al. uses computational tools to characterize *Enterobacter* phytases which were predicted to be thermostable based on its aliphatic indices of below 40, acidic in nature as the isoelectric point was below 7, interaction with water based on lower values of GRAVY, the highest content of alpha-helical compared to other forms, function as histidine acid phosphatases and has conserved residues ‘DG–DP–LG’ (Pramanik et al., 2018). Similarly, V.Kumar et.al. characterized Histidine acid phytase sequences using computational tools and also revealed the motifs that are found in obtained clusters i.e. clusters 1 (PhyA) and 2 (PhyB) has Motif 1 “SPFCDLFTHEEWIQYDYDLQSLGKYYGYGAGNPLGPAQGIGF” and Cluster 3 (AppA) contains motif 9 “KKGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITV-HTQADTSSPDP”. This information can help to study the evolution of HAP phytase and the engineering of phytases to improve their features based on clustering (Kumar et al., 2012).

Similarly, 40 reference Protein tyrosine phosphatase-like cysteine phytases (CPhy) sequences were computationally explored and *In-silico* characterization revealed conserved regions, phylogenetic relationship, physio-chemical features, functional aspects as well common motif ‘TDHKWPTDEMVDYFVQFVKSMFKDTWLHFHCQAGIGRTTTFMI MYDMMKN’ present in all CPhy sequences (KUMAR and AGRAWAL, 2014). Similarly, pyranose dehydrogenase (Verma, 2014) , and hypothetical proteins (Carlos Leonardo Araújo, Iago Blanco, Luciana Souza, Sandeep Tiwari, Lino César Pereira, Preetam Ghosh, Vasco Azevedo, Artur Silva, 2020) and many more enzymes were explored through *in-silico* characterization method.

1.9 Engineering of novel enzymes:

Researchers all over the world have been focusing on different aspects to increase thermostability via protein engineering approach (directed or rational engineering) and computational analysis via software and tools.

The directed evolution approach is time-consuming and requires an extensive screening process. Shivange et al. focus on the reduction of flexibility to enhance thermostability via the adopted KeySIDE approach to improve the thermostability of *Yersinia mollaretii* phytase (Ymphytase) which involved the combination of investigating iterative key residues of the wild type and identifying substitutions using directed evolution method. M6 mutant (T77K, Q154H, G187S, and K289Q) improved the residual activity as compared to the wild type. The mutations T77K, G187S, and K289E/Q reduced the flexibility of the loops near helices which overall increased thermostability (Shivange, Roccatano and Schwaneberg, 2016).

Researchers also adopt a rational engineering approach to improve the characteristics of enzymes e.g., Fei et.al. focused on protein flexibility as well as protein surface analysis and enhancement of salt bridges as a strategy for rational engineering of *Escherichia coli* AppA phytase. They analyzed protein flexibility with the help of a Root mean square deviation (RMSD) graph obtained via Molecular Dynamic Simulation (MDS). Surface thermal unstable residues were targeted based on their RMSF values which were above 2 Å and are above 4 Å away from the AppA functional sites. They demonstrated that the salt bridges and the α/β -domain of *E. coli* phytase are of utmost importance. They introduce mutation Q307D which enhanced thermostability compared to wild type (Fei et al., 2013). Han et.al., adopted a consensus sequence rational protein engineering approach to engineer *A.niger* phytase (Anp) by structural comparison with *A.fumigatus* phytase (Afp) structure with the help of molecular

dynamics simulation. In AnP, the segments A35-P42, R163-Q168, and T248-K254 exhibited the highest main chain deviation, and thus were the targets for engineering. By hydrogen bond analysis of mutant candidate, it was found that S39 and S42 in segment-1, T165, N166, and R167 in segment-2, R248, D251, A252, and Q254 in segment-3 have the greatest contribution to the enhanced thermostability (Han et al., 2018). Similarly, enzymes such as carbonic anhydrase, (Parra-Cruz et al., 2018), lytic polysaccharide monoxygenases (Zhou et al., 2022), and haloalkane dehalogenase (Satpathy et al., 2015) are rationally engineered to make it thermostable.

1.10 Problems related to the application of phytases:

The utilization of phytase in the feed and food industry, bioethanol and myo-inositol phosphates production process are restricted due to its reduced effectiveness during high-temperature processing stages (Rebello et al., 2017) (V. Kumar & Sinha, 2018) (Makolomakwa et al., 2017). The application of phytases is also limited in the soil amendment process as it depends upon various factors such as soil environment (pH, temperature, presence of metal ions), bioavailability of phytate, or presence of inhibitors in soil (Menezes-Blackburn et al., 2013). Hence there is a constant need to explore novel stable phytases through conventional and computational approaches (Tan et al., 2016) (Castillo Villamizar et al., 2019).

The present study is important because there is a demand for stable phytase in the feed and fuel industry since the introduction of phytase has the potential to reduce environmental pollution, natural resource management, increase bioavailability of micronutrients, and increase bioethanol production, reduce feed cost. Few commercial phytases are available but they exhibit reduced activity as temperature increases.

In this context, the present study aimed at finding novel phytases with improved thermostability using *in-silico* bio-prospecting and rational engineering approaches, because it has environmental, economic, and health benefits.

Chapter 2

Review of Literature

Review of Literature

2.1 Bioprospecting of novel enzymes:

There is a huge demand for enzymes in almost every aspect of life in today's world. The reasons or key factors that drive the enzyme demand and growth of the industrial enzyme market are i) the growing global population that consumes food and beverages, ii) rising healthcare costs, iii) personalized medicine (increased diagnostic testing and genome sequencing), iv) more insured people undergoing routine medical care) growth in the bioenergy sector (Li et al., 2012). The bioprospecting of novel enzymes can be performed using different approaches like conventional, metagenomics, and *in-silico* methods.

2.1.1 Conventional bioprospecting of enzymes:

In conventional bioprospecting, the initial enzymes are surveying natural reservoirs, cultivating and isolating microorganisms that produce target enzymes, or adopting a culture-free approach. This is followed by identifying potential candidates or genes and undergoing experimental verification (A. D. Kamble & Singh, 2022). Extremophiles, organisms that thrive in extreme conditions such as high temperature, saline environments, or extreme acidity/alkalinity, possess robust biomolecules like proteins and enzymes that demonstrate greater stability compared to their counterparts found in mesophiles (Sarmiento et al., 2015) (Boteva & Kambourova, 2018). Researchers have isolated stable enzymes from hot springs in Tunisia (Thebti et al., 2016), Sangameshwar (R. M. Kumar & Raja, 2019), Taptapani Hot Spring (R. K. Sahoo et al., 2017), Eryuan Niujie Hot spring (Yan et al., 2017), and the hot spring in Xiamen, China (C. Zhao et al., 2017).

Powar and Jagannathan isolated phytase from *B.subtilis* in 1982 (Powar & Jagannathan, 1982). Kumar et al. extracted phytase-producing micro-organisms such as *Acromobacter* sp. PB-01, *Tetrathiobacter* sp. PB-03, and *Bacillus* sp. PB-13 from Himalayan soils, which helped the growth of *Brassica juncea* (V. Kumar et al., 2013). Aseri et al. explored fungi capable of degrading phytin and glycerolphosphate from arid and semi-arid Indian soils, with varying activity among different fungi (Aseri et al., 2009). Puppala et al. characterized a novel acidic and thermostable phytase isolated from *Streptomyces* sp. (NCIM 5533) on the Sanjivani islands in Maharashtra, India, displaying plant growth-promoting properties by supporting the growth of *Solanum lycopersicum* (Puppala et al., 2019). Enhancing phytase production from *Citrobacter koseri* was achieved using agro-industrial residues like wheat bran and oil meals as substrates in both submerged fermentation and solid-state fermentation (Tripathi et al.,

2016). Reports on phytase from *Enterobacter cloacae* strain PSB-45 in compost and *Serratia* sp. PSB-15 in degraded wood sources indicated K_m and V_{max} values of 1.25 mM and 0.140 $\mu\text{M}/\text{min}$, and 0.48 mM and 0.157 $\mu\text{M}/\text{min}$, respectively (Kalsi et al., 2016). Similarly, Roy and Datta isolated a Ca^{2+} -dependent, extracellular β -propeller phytase from *Bacillus aryabhatai* RS1 in the rhizosphere (Moushree Pal Roy, Subhabrata Datta, 2017). Sajidan et.al. researched soil samples collected from volcanic regions in Central Java to identify phytase-producing organisms. They discovered that these phytases were associated with the *Bacillus* genus, specifically *B.cereus*, *B.aryabhatai*, and *B.psychrotolerans* (Sajidan et al., 2015). Separately, Jorquera et al. isolated thermostable phytases from Chilean hydrothermal environments, attributing them to the *Bacillus* and *Geobacillus* genera (Jorquera et al., 2017). Furthermore, *Enterobacter* sp. ACSS phytase, after purification, displayed stability across a broad temperature range temperature (Chanderman et al., 2016). Phytase-producing bacteria were isolated from *Bambusa tulda* Roxb.'s rhizosphere and were identified based on their halo-to-colony ratio. The activity of phytase was assessed using wheat bran and Luria Bertani (LB) medium after acetone precipitation, while the bacterial genomes were examined for the β -propeller phytase gene associated with *Bacillus* spp. using polymerase chain reaction (PCR). Additionally, their impact on maize seedling growth in phosphorus-deficient loamy soil was evaluated in pot conditions. Ten bacterial isolates from *B. tulda*'s rhizosphere displayed the ability to decompose phytate, with SRBR-04 showcasing the most notable effect. Two isolates (SRBR-01, SRBR-04) were positive for the *Bacillus* phytase gene. In pot trials using phosphorus-deficient soil, isolate SRBR-07 notably increased biomass, improving shoot height, dry shoot weight, dry root weight, and leaf area compared to the untreated control. The introduction of phytase-producing bacteria into phosphorus-deficient agricultural soils could represent a sustainable approach for managing phosphorus nutrition in *Zea mays* (Gauchan et al., 2023). Bharambe et.al. gathered phytase-producing microorganisms from soil and litter samples obtained from poultry, cattle, goatshed, and gardens. Among these, *A.terreus* fsp-4 exhibited promising potential as a phytase producer (Bharambe & Peshwe, 2023). Ihsein et.al. extracted microorganisms capable of degrading phytate from both the rhizosphere and bulk soil. Molecular examination indicated their affiliation with BPP. In the phosphorus-deficient soil, bacteria carrying the BPP gene were present in both the rhizosphere (R) and bulk (B) soil fractions. Around 25% of the bacteria in the bulk soil and 37.5% in the rhizosphere tested positive for the gene. These findings strongly suggest a higher prevalence of phytase bacteria in the rhizosphere of *P.vulgaris* compared to the bulk soil (Ihsein Rokia Amine-Khodja, Ryn Maougal, 2023). The utilization of plant growth-promoting bacteria (PGPB) producing

phytase, an enzyme capable of breaking down inositol phosphate in soil, stands as a sustainable method for providing available phosphorus (P) to plants. Li et al. collected 73 potential bacterial isolates from the rhizosphere of seven significant grass species in the alpine grassland of the Qinghai-Tibetan Plateau. They employed qualitative and quantitative methods to evaluate the bacteria for their ability to promote plant growth (PGP traits). Results showed that most of these bacteria, originating mainly from *Proteobacteria* and *Actinobacteria*, represented 16 different genera. Notably, *Pseudomonas* species were most prevalent among the isolates and demonstrated notable phytase activity. Additionally, six strains showed positive amplification for the phytase gene (β -propeller phytase, BPP), significantly enhancing various parameters in *Lolium perenne* L. under P-limitation conditions. The expression of the phytase gene (BPP) in the root system was confirmed using qPCR. Furthermore, the PHY101 gene responsible for encoding phytase from *Pseudomonas mandelii* GS10-1 was cloned, sequenced, and expressed in *Escherichia coli*. The phytase PHY101 demonstrated its activity at pH 6 and 40°C. It exhibited as a potential PGPB (Q. Li et al., 2023). Nezhad and colleagues isolated ten bacterial strains and one yeast strain known for producing phytase. Identification involved the analysis of 16S rRNA and the nuclear ribosomal transcribed spacer (ITS). To characterize their phytase production, they conducted qualitative and quantitative tests. Additionally, they amplified the histidine acid phosphatase gene using specific primers for further study. Phytase activity levels varied from 194.21 mU/mL to 381 mU/mL. Sequencing identified *Acinetobacter* as the dominant genus among six of the eleven isolates, while the rest were associated with *Enterobacter*, *Pseudomonas*, *Escherichia*, and *Saccharomyces*. Moreover, a new histidine acid phosphatase gene (PhySc) was found and amplified from the SPA isolate, showing a close resemblance to PHO5 from *Saccharomyces cerevisiae* YJM993 (accession number CP004601.2) (Nezhad et al., 2023).

2.1.2 Bioprospecting of enzymes through functional metagenomic approach:

Earth hosts around $4-6 \times 10^{30}$ prokaryotes, with 95-99.9% of these microorganisms being uncultivable due to the constraints of culture-based techniques. The traditional approach has drawbacks: it's costly and time-consuming. The samples must meet specific quality standards for analysis, and purity is crucial. Candidates are selected through specialized media. Screening might not always yield promising candidates (A. D. Kamble & Singh, 2022) (Boddu et al., 2022).

Pace et.al. pioneered a cultivation-independent method aiming to isolate all genomes from diverse microorganisms within a habitat. Schmidt et al. conducted the initial characterization of a community using metagenomic data based on 16S rRNA genes. This approach accounts for both the cultivable and uncultivable components of the microbial community. The basic procedures in metagenomic analysis include obtaining and breaking down the metagenomic DNA from various environmental samples, creating a metagenomics library, sequencing the genetic material, assembling it, and then classifying it based on taxonomical criteria (A. D. Kamble & Singh, 2022).

In recent times, metagenomics has emerged as an effective method for discovering new enzymes by leveraging extensive genomic data from natural habitats. Bilirubin-oxidizing enzymes (Kimura & Kamagata, 2016), phytases (Castillo Villamizar et al., 2019), glycoside hydrolases (G. Zhang et al., 2016), antibiotic resistance enzymes (Elbehery et al., 2017), novel lignocellulose-degrading enzymes (Thornbury et al., 2018), serine metalloproteases (Shamim et al., 2018), hydrogenases (Adam & Perner, 2018), and polyhydroxyalkanoate synthase (Foong et al., 2018) were discovered using this approach. Liew et.al. isolated metagenomic DNA from hot spring water. They identified cellulose-degrading genera and thereby amplified cellulase genes, expressed, and purified recombinant Endoglucanase (EglDG) from *Dictyoglomus* sp. (Liew et al., 2017). Meng et.al. obtained the cellulase-xylanase protein from the buffalo rumen metagenome (Meng et al., 2023).

Gharechahi et.al. identified several types of bacteria in the rumen with the ability to break down various complex carbohydrates. Notably, *Verrucomicrobia* and *Spirochaetes* were found to possess enzymes specifically suited for digesting intricate polysaccharides such as xyloglucans, peptidoglycans, and pectin (Gharechahi et al., 2023). Rathour et.al. identified FMN-dependent NADH azoreductase from the bioreactor biofilm metagenome (Rathour et al., 2023).

Alkaline phytase was extracted from red rice, and composted castor bean cake. The study identified PhytRC001, a novel phytase from an uncultured microbial community. This novel phytase exhibited the potential to be applied in various industries (Farias et al., 2018). Tan et.al. used natural resources such as soil to isolate two potential phytases. One of them belongs to the histidine acid phytases type and the other one was a novel phytase. Both phytases can be applied in industries due to their characteristics (Tan et al., 2014). Kuang et al. explored Baiyangdian Lake to understand the microbial diversity and their associated metabolic

pathways. The dominant groups present in the lake were *Actinobacteria* and *Proteobacteria*. They were found to be involved in various roles, such as phytase activity, regulation of phosphate, and nitrogen and sulfur breakdown and transformations (Kuang et al., 2023).

2.1.3 *In-silico* bioprospecting of novel enzymes:

In-silico bioprospecting entails discovering potential candidates through existing databases and characterizing them using various tools, servers, or software through computational methods. *In-silico* bioprospecting played a role in discovering new enzymes, antibiotics, and therapeutic compounds. Researchers explore databases using methods such as homology comparison, searching for conserved motifs, employing consensus-guided approaches, or utilizing keyword searches. Filters like percent identity, query coverage, and e-value are applied to refine and narrow down the results. Trusted online repositories associated with the International Nucleotide Sequence Database include the DNA Databank of Japan (DDBJ; <http://www.ddbj.nig.ac.jp/>), the European Bioinformatics Institute of the European Molecular Biology Laboratory (EMBL-EBI; <http://www.ebi.ac.uk/ena/>), the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>), and the Joint Genome Institute (JGI; <https://img.jgi.doe.gov/>) (A. D. Kamble & Singh, 2022) (Perez Rojo et al., 2023). Online databases/tools used for *in-silico* bioprospecting are mentioned in **Table 2. 1** below:

Table 2. 1 Online Databases/tools used for *in-silico* bioprospecting: (Adopted from: (A. D. Kamble & Singh, 2022)).

Sr. No	Database/ Details	Weblink/ References
1	Uniprot: It is the protein database. The data is collected from Swiss-Prot, TrEMBL, and PIR-PSD databases.	https://www.uniprot.org/ (Magrane & Consortium, 2011)
2	Conserved Domains database: The protein functional units are annotated based on conservation patterns.	https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml (Lu et al., 2020)
3	Protein database: The protein sequences are distributed in GenBank, RefSeq, TPA, SwissProt, RIP, and PDB databases.	www.ncbi.nlm.nih.gov/home/proteins.shtml (Pruitt et al., 2012) (Berman et al., 2007)
4	Protein cluster: It includes the RefSeq protein sequence encoded by the chlorophyll genome. It also includes the prokaryotic genome.	https://www.ncbi.nlm.nih.gov/proteinclusters (Klimke et al., 2009)
5	Structure: The protein sequence data is derived from the PDB database. The information on sequences is linked to the sequence database, NCBI taxonomy database, and bibliographic information.	https://www.ncbi.nlm.nih.gov/structure (Madej et al., 2012)

Chapter 2

6	Protein Data Bank (PDB): The database consists of X-ray diffraction or NMR structure data of protein and also consists of gene sequences. The 3D structure of a protein is readily available	https://www.rcsb.org/ (Berman et al., 2007) (Burley et al., 2021)
7	InterPro: It helps to predict protein family, domain, and binding site. It uses CATH, CDD, PRINTS, and Pfam database for the prediction.	https://www.ebi.ac.uk/interpro/ (Blum et al., 2021)
8	Brenda: It is a collection of information related to enzymatic reaction, structure, function, kinetic parameters, stability, and sequence data.	https://www.brenda-ews6Y6s.org/ (Schomburg et al., 2004)
9	BRENDA, KEGG, MetaCyc, and SABIO-RK (BKMS-react): It includes information about the enzymatic reactions, metabolic pathways as well as experimental conditions	http://bkms-react.tu-bs.de/index.php (Lang et al., 2011)
10	A Database of Enzyme Catalytic Mechanisms (EzCatDb): This includes information about enzymatic reactions, cofactors, metabolites and active site domain and catalysis	http://ezcatdb.cbrc.jp/EzCatDB/ (Nagano, 2005)
11	Mechanism and Catalytic Site Atlas (M-CSA): It includes information about the catalytic residues, cofactors, and reaction mechanism	https://www.ebi.ac.uk/thornton-srv/m-csa/ (Ribeiro et al., 2018)

Chapter 2

12	Thermodynamic database Database for Proteins and Mutants (ProTherm): This contains information related to the change in thermal stability upon mutation	https://www.iitm.ac.in/bioinfo/ProTherm/ (Gromiha et al., 2000)
13	ProtaBank: It includes mutated protein sequence information	https://www.protabank.org/ (C. Y. Wang et al., 2018)
14	Kyoto Encyclopedia of Genes and Genomes (KEGG): This database is a collection of genomic, chemical, and system function information.	https://www.genome.jp/kegg/ (Yi et al., 2020)
15	Protein Families (Pfam): This is a protein family database. The multiple sequence alignment and hidden Markov model help to cluster these families.	http://pfam.xfam.org/ (El-Gebali et al., 2019)
16	PRINTS: This is a protein motif database	http://130.88.97.239/PRINTS/index.php (Attwood et al., 1994)
17	CATH (Class, architecture, topology, and homologs): This is a protein database. It consists of protein superfamily information such as Class, architecture, topology, and homolog sequences	https://www.cathdb.info/ (Knudsen & Wiuf, 2010) (Knudsen & Wiuf, 2010) (Dawson et al., 2017)
18	Structural Classification of Proteins (SCOP): This helps in protein structure classification	http://scop.mrc-lmb.cam.ac.uk/ (Andreeva et al., 2020)

Chapter 2

19	Homology-derived Secondary Structure of Proteins (HSSP): This includes protein secondary structure database	https://swift.cmbi.umcn.nl/gv/facilities/ (Touw et al., 2015)
20	SWISS-3DIMAGE: This helps to gain information about the 3D structure of proteins	http://www.pdg.cnb.uam.es/cursos/Leon_2003/pages/visualizacion/programas_manuales/spdbv_userguide/us.expasy.org/sw3d/index.html (Sussman, 1995) (Sussman, 1995)
21	Biological Magnetic Resonance Data Bank (BioMagResBank): This has the collection of NMR database of proteins, amino acids, and nucleotides	https://bmr.io/ (Ulrich et al., 2008)
22	SWISS-MODEL Repository: This database consists of a 3D structure	https://swissmodel.expasy.org/repository (Bienert et al., 2017) (Bienert et al., 2017)
23	process: This is a protein classification database	https://proteininformationresource.org/pirwww/dbinfo/iprclass.shtml (C. H. Wu et al., 2004)
24	TIGRFAM: This is a protein family database	http://tigrfams.jcvi.org/cgi-bin/index.cgi (Haft et al., 2013)

Chapter 2

25	OWL: It is a protein sequence database. The information is derived from SWISS-PROT, PIR, GenBank, and NRL-3D primary sequence database	http://130.88.97.239/OWL/ (Bleasby et al., 1994)
26	The database of three-dimensional interacting domains (3DID): This is a database that involves interaction information of protein with known 3D structure.	https://3did.irbbarcelona.org/ (Mosca et al., 2014)
27	DOMINE: This is a protein domain interaction database	https://manticore.niehs.nih.gov/cgi-bin/Domine (Yellaboina et al., 2011)
28	Binding MOAD (the Mother of All Databases): It is a collection of protein-ligand crystal structure information	https://bindingmoad.org/ (R. D. Smith et al., 2019)
29	Phospho.ELM: This is a Protein phosphorylation site database	http://phospho.elm.eu.org/ (Dinkel et al., 2011)
30	SuperSite: The database is a collection of drug-protein binding site information along with its mechanism of action and conservation of binding site information.	https://www.hsls.pitt.edu/obrc/index.php?page=URL1237565601 (Bauer et al., 2009)
31	STITCH: This consists of protein-compound interaction network information	http://stitch.embl.de/ (Kuhn et al., 2008)

Chapter 2

32	<p>Reactome: This includes detailed information about the protein molecules involved in various biochemical pathways. It also provides a network chart of biochemical processes. The information is collected from UniPort, KEGG, OMIM, etc.</p>	<p>https://reactome.org/ (Fabregat et al., 2017)</p>
33	<p>UniHI: This gives information about the protein-protein interaction</p>	<p>http://www.unihi.org/ (Kalathur et al., 2014)</p>
34	<p>Bionemo: It is a collection of data related to proteins and genes involved in biodegradation metabolism. It also includes information related to structural, and biochemical information about the proteins and genes involved in biodegradation pathways</p>	<p>https://www.hsls.pitt.edu/obrc/index.php?page=URL1233778230 (Carbajosa et al., 2009)</p>
35	<p>The proteolysis map (PMAP): Proteolysis pathway database</p>	<p>https://www.hsls.pitt.edu/obrc/index.php?page=URL1233764714 (Igarashi et al., 2009)</p>
36	<p>CAZy (the Carbohydrate-Active enZYmes Database): The carbohydrates degradation enzyme database</p>	<p>http://www.cazy.org/ (Cantarel et al., 2009)</p>
37	<p>The lipase engineering database (LED): The lipase engineering database</p>	<p>http://www.led.uni-stuttgart.de/ (Fischer & Pleiss, 2003)</p>

38	MEROPS: This is a peptidase database	https://www.ebi.ac.uk/merops/ (Rawlings et al., 2018)
39	Protease substrate specificity webserver (PROSPER): It predicts the protease catalytic substrate and cleavage sites based on the sequence comparison with the protease superfamilies.	https://prosper.erc.monash.edu.au/home.html (Song et al., 2012)
40	SABIO-RK: This includes information about the biochemical and kinetic parameters.	http://sabio.h-its.org/ (Wittig et al., 2006)
41	EBI Metagenomics: It is an automated pipeline for Storage of metagenomic data, and functional and metabolic analysis.	https://www.ebi.ac.uk/metagenomics/ (Mitchell et al., 2018) (Mitchell et al., 2018)
42	Human Pan Microbe Communities Database (HPMCD): It consists of manually curated and searchable human gastrointestinal microbial data	http://www.hpmcd.org/ (Forster et al., 2016)
43	KEGG MGENES Database: It consists of gene data obtained from large-scale environmental sequencing studies	https://www.genome.jp/mgenes/ (Yi et al., 2020)
44	Metagenomes Online: It consists of curated protein data obtained from environmental metagenomic studies	http://metagenomesonline.org/ (Eric Wommack et al., 2012)

Chapter 2

		(Nasko et al., 2018) (Nasko et al., 2018)
45	Viral Informatics Resource for Metagenome Exploration (VIROME): The analysis of viral metagenomic sequences	http://virome.dbi.udel.edu/ (Eric Wommack et al., 2012)
46	Piphillin: It helps in the prediction of gene composition. This is done with the help of 16S rRNA OTU from human clinical samples metagenomic dataset.	http://piphillin.secondgenome.com/ (Narayan et al., 2020)
47	Web services for metagenomic analysis (WebMGA): These are a collection of customizable web servers. This helps in fast metagenomic analysis.	http://weizhong-lab.ucsd.edu/webMGA/ (S. Wu et al., 2011)
48	MG-RAST: It is a collection of: repository+ annotation + analysis tools for metagenomic analysis	https://www.mg-rast.org/ (Keegan et al., 2016) (Keegan et al., 2016)
49	Naïve Bayesian Classification tool (NBC): This helps in taxonomic classification of metagenomic data	http://nbc.ece.drexel.edu/ (Rosen et al., 2011)
50	CoMet: It uses protein domain signatures for comparative analysis of metagenomic data	http://comet2.gobics.de/ (Lingner et al., 2011)

Chapter 2

51	MetaBioME: This helps in searching commercially important enzymes in the metagenomic dataset	http://metasystems.riken.jp/metabiome/ (V. K. Sharma et al., 2009)
52	16S Classifier: It uses 16s rRNA hypervariable regions for taxonomic classification.	http://metagenomics.iiserb.ac.in/16Sclassifier/application.php (Chaudhary et al., 2015) (Chaudhary et al., 2015)
53	ENZYME: It is the collection of information related to characterized enzymes and their EC number	https://enzyme.expasy.org/ (Bairoch, 2000)
54	ExplorEnz: The IUBMB enzyme nomenclature is curated and disseminated.	https://www.enzyme-database.org/ (McDonald et al., 2009)
55	the Structure-Function Linkage Database (SFLD): It helps in Hierarchical classification of enzymes as well as predicting structure-function relationship	http://sfld.rbvi.ucsf.edu/archive/django/index.html (Eyal Akiva et.al., 2014)
56	Catalytic Families (CatFam): The enzyme sequence profile is used for understanding protein catalytic functions.	http://www.bhsai.org/download/s/catfam.tar.gz (Yu et al., 2008)
57	MetaCyc: The experimental data of metabolic pathways and enzymes are stored in this database.	https://metacyc.org/ (Karp et al., 2002)

58	EAWAG-BBD: It gives information about the biocatalytic reactions and biodegradation pathways	http://eawag-bbd.ethz.ch/ (Ellis & Wackett, 2012)
59	Fungal Peroxidase Database (fPoxDB): It is a database of fungus peroxidases. It helps in the comparative and evolutionary genomics of peroxidases.	http://peroxidase.riceblast.snu.ac.kr/index.php?a=view (Choi et al., 2014)
60	Enzyme Structure Database: It is a collection of known enzyme structures and this data is submitted to PDB.	https://www.ebi.ac.uk/thornton-srv/databases/enzymes/ (Laskowski et al., 2018)
61	Integrated relational Enzyme database (IntEnz): It helps in enzyme nomenclature.	https://www.ebi.ac.uk/intenz/ (Fleischmann et al., 2004)
62	REBASE: It is a collection of information related to DNA restriction and modification systems.	http://rebase.neb.com/rebase/rebase.html (Roberts et al., 2015)
63	ESTerases and alpha/beta-Hydrolase Enzymes and Relatives (ESTHER Database): It helps in analyzing the proteins and protein domains of hydrolyses.	http://bioweb.supagro.inra.fr/ESTHER/general?what=index (Lenfant et al., 2013)
64	a peroxidase database (PeroxiBase): It helps in identifying putative functions and transcription regulation of peroxidase	https://www.uniprot.org/databases/DB-0072 (Passardi et al., 2007)
65	KinBase: It is a collection of information related to protein Kinases	https://www.hsls.pitt.edu/obrc/index.php?page=URL10548429 58

Chapter 2

		(Caenepeel et al., 2004)
66	ArthropodaCyc: It helps in predicting and identifying enzymes and their pathways that are unique to a given organism and or group of organisms	http://arthropodacyc.cycadsys.org/ (Baa-Puyoulet et al., 2016)
67	a database for plant carbohydrate-active enzymes (PlantCAZyme): It helps in providing sequence and data related to CAZymes	http://bcb.unl.edu/plantcazyme/ (Ekstrom et al., 2014)
68	a database of ORphan ENZyme Activities (ORENZA): It helps in predicting enzymatic activities of enzyme which does not have any representative sequence in the database.	http://www.orenza.universite-paris-saclay.fr/ (Lespinet & Labedan, 2006)
69	MEME: Functional motifs	https://meme-suite.org/meme/ (Bailey et al., 2015)
70	ProtoNet: The orthologous sequences were clustered	http://www.protonet.cs.huji.ac.il/ (Rappoport et al., 2012)
71	TargetP/ChloroP: Subcellular localization	http://www.cbs.dtu.dk/services/TargetP/ http://www.cbs.dtu.dk/services/ChloroP/ (Emanuelsson et al., 2000)

Chapter 2

		(Heinegård & Paulsson, 1987)
72	TMHMM server: Predicting membrane protein	http://www.cbs.dtu.dk/services/TMHMM/ (Krogh et al., 2001)
73	SignalP: Detecting Signal peptide	http://www.cbs.dtu.dk/services/SignalP/ (Almagro Armenteros et al., 2019)
74	ProtParam tool: Predicting the molecular weight, theoretical pI, total number of positive and negative residues, extinction coefficient, aliphatic index, instability index, and grand average of hydropathicity (GRAVY)	https://web.expasy.org/protparam/ (Gasteiger et al., 2005)
75	the self-optimized prediction method (SOPMA): The percentage of α -helix, β -pleated sheet, and coils were calculated	https://vlab.amrita.edu/?sub=3&brch=275&sim=1454&cnt=1 (Geourjon & Deléage, 1995)
76	PsiPred: The structural details such as disordered region of protein, disulfide bond, internal repeats, pest region, ubiquitination site, n-glycosylation site, and intron-exon organization can be identified.	http://bioinf.cs.ucl.ac.uk/psipred/ (McGuffin et al., 2000)

Chapter 2

78	Modeller: Structure modeling	https://salilab.org/modeller/ (Eswar et al., 2006)
79	Computed Atlas of Surface Topography of Proteins (CASTp): Predicting the conserved active catalytic residues in selected protein model	http://sts.bioe.uic.edu/castp/index.html (Tian et al., 2018)
80	PROCHECK: Help in understanding the stereochemistry quality of the protein structure	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/ (ROMAN A. LASKOWSKI, MALCOLM W. MACARTHUR, DAVID S. MOSS, 1993)
81	Protein Structure Analysis (PROSA): It helps in the refinement of structure, structure prediction, and modeling	https://prosa.services.came.sbg.ac.at/prosa.php (Wiederstein & Sippl, 2007)
82	VERIFY3D: It is used for the evaluation of the structure	https://servicesn.mbi.ucla.edu/verify3d/?job=9678&p=verify3d (Eisenberg et al., 1997)
83	ERRAT: It is used to verify the reliability of the modeled structure.	https://servicesn.mbi.ucla.edu/ERRAT/ (Colovos & Yeates, 1993)
84	Qualitative Model Energy Analysis (QMEAN): It is used to check the quality and analyze the stereochemistry of the modeled structure.	https://swissmodel.expasy.org/qmean/ (Benkert et al., 2007)

Bunternngsook et al. (2021) conducted *in-silico* bioprospecting to identify potential lytic polysaccharide monoxygenases (LPMOs) in a thermophilic microbial metagenomics dataset sourced from soil near a bagasse collection site. They utilized the Blastp tool against the Carbohydrate-Active enZymes (CAZy) database (Bunternngsook et al., 2021). Likewise, González-Torres et al. utilized the NCBI nr database to search for putative L-Asparaginases (ASNases) (González-Torres et al., 2020). Araújo et al. retrieved 172 hypothetical proteins (HPs) from 32 *Corynebacterium pseudotuberculosis* biovar ovis genomes using the NCBI database and the online RAST platform (Rapid Annotations using Subsystems Technology) (Carlos Leonardo Araújo et.al., 2020). In 2017, Zhang et al. extracted active or purified enzymes from around 6000 clones within the metagenomics database (ZHANG Jian-zhi et.al., 2020). Góngora-Castillo et.al. investigated the occurrence of new S8A proteases across different phyla, including *Proteobacteria*, *Actinobacteria*, *Planctomycetes*, *Bacteroidetes*, and *Cyanobacteria* (Góngora-Castillo et al., 2020). Ariaeenejad et al. discovered novel alkali-thermostable endo-b-1,4-glucanase via raw sheep rumen metagenomics data (Ariaeenejad et al., 2020). Additionally, Foong et al. discovered the polyhydroxyalkanoate (PHA) synthase gene via Balik Pulau (BP) mangrove soil metagenome data in the MG-RAST database to locate (Foong et al., 2018).

Homology-based *in-silico* bioprospecting:

A known candidate is utilized to explore the existing database to extract new enzyme sequences, either in protein or DNA form. The database must consist of complete sequences with conserved domains. The sequences which were > 90% identical to the known phytases are excluded from the list to obtain novel enzymes with conserved domains and distinct from the closet homologous sequences (A. Kamble et al., 2019)(ZHANG Jian-zhi, et.al., 2020) (Y. Zhang et al., 2020).

Several studies helped to obtain novel enzymes through a homology-based *in-silico* bioprospecting approach. E.g. Gupta et.al., Rolf et.al., and Sadeghian et.al. used the NCBI database and Blastp analysis to obtain a list of hypothetical proteins from *Triticum aestivum*, wheat, and novel cyclic GMPeAMP synthase (cGAS), thermostable multicopper oxidases respectively (Gupta et al., 2018) (Rolf et al., 2020), (Sadeghian et al., 2020). Similarly, Yata et.al. used computational tools to explore Urease from *Cucumis sativus* and *Fragaria vesca* by comparing it to the reference Pigeon pea Urease (Yata et al., 2021). The novel β -esterases were discovered through a peptide pattern recognition tool (Voß et al., 2020). Two new endo-b-N-

acetylglucosaminidase isoenzymes were investigated based on the structural and functional aspects (Şahutoğlu et al., 2020). The novel α -amylases were mined by using the sheep rumen metagenomics data (Motahar et al., 2020). Vaquero et.al. used MEME software to discover the conserved motifs in the *Pseudozyma antarctica* lipase B family (Vaquero et al., 2015). IMG/MER hot spring database was used to explore the different metagenomic ado genes from *Synechococcus elongatus* PCC7942 (Shakeel et al., 2018). The novel β -glucosidase was explored by using the microbial metagenome of Lake Poraquê (Toyama et al., 2018).

Conserved domain-based *in-silico* bioprospecting:

The novel candidates were identified by using the conserved domain. HP sequences utilized the NCBI protein and UniProt databases. The Conserved Domains Database (CDD)-search suite was used to obtain information regarding conserved domains. The information obtained from the CDD was confirmed by using Protein families (Pfam) and InterProScan (Guha et al., 2020). Tan et.al. explored 18 fungus garden metagenomes deposited in the IMG/M database to obtain a list of novel histidine acid phosphatase based on conserved domains, shortlisting a total of 11 putative sequences (Tan et al., 2015). Thornbury et.al. utilized NGS generated microbiome dataset along with a putative domain to obtain a list of novel candidates (Thornbury et al., 2018). Similarly, camel rumen, sheep rumen, cold marine sediment metagenomic sequence data was used to explore various enzymes such as thermostable cellulases and thermostable, halotolerant along with glucose-tolerant novel β -glucosidase, and novel monooxygenases (Maleki et al., 2020) (Ariaeenejad et al., 2020) (Musumeci et al., 2017).

Homology + conserved domain approach:

The combination of methods i.e., homology and conserved domain, was used to explore novel enzymes based on parameters such as e-value, conserved domains, and conserved motifs. E.g., Sahoo et.al. initially obtained a list of HPs from 26 microbial genomes by using genome data and computational tools like Blastp (S. Sahoo et al., 2020).

2.2 *In-silico* characterization:

Before any candidate is experimentally verified and validated, an *in-silico* characterization is helpful to gather all the properties relevant and specific to the protein. This includes structural, evolutionary, and interaction studies using various computational tools and software. A compiled list of these tools used in *in-silico* characterization is given in **Table 2. 1**.

2.2.1 Analysis of primary sequence and secondary, and tertiary structure:

ExPasy-ProtParam and PEPSTATS tools were used to extract information regarding isoelectric point (pI), total amino acid composition, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) (Bhagwat et al., 2021) (N. Sharma et al., 2017) (Guha et al., 2020) (A. D. E. V Sharma, 2020) (A. A. Smith & Caruso, 2013).

The secondary structure analysis was evaluated using tools/servers/software such as PSI-blast-based secondary structure PREDiction (PSIPRED), ESPript, the Self-Optimized Prediction Method with Alignment (SOPMA), and SIB MyHits (Bhagwat et al., 2021) (Hemmati & Mehrabadi, 2020) (Motahar et al., 2020).

I-TASSER (Iterative Threading ASSEmby Refinement), HHpred web server, MODELLER, Phyre2 server, and Swiss-model were used for model tertiary structure (González-Torres et al., 2020) (Şahutoğlu et al., 2020) (Hemmati & Mehrabadi, 2020) (Shakeel et al., 2018) (Ariaeenejad et al., 2020) (Maleki et al., 2020) (Bunterngsook et al., 2021) (Şahutoğlu et al., 2020). PROCHECK, Qualitative Model Energy Analysis (QMEAN), and Verify 3D were used to evaluate the tertiary structure (Hemmati & Mehrabadi, 2020).

2.2.2 Phylogenetic tree:

Phylogenetic analysis was constructed by using different servers/software such as Molecular Evolutionary Genetics Analysis (MEGA) (Bunterngsook et al., 2021), EMBL-EBI (Yata et al., 2021), PhyloPythiaS (Tan et al., 2015), and the MAFFT server (Voß et al., 2020) (Maulana et al., 2023).

2.2.3 Evaluation of Function:

Function evaluation was carried out by analyzing protein interaction, glycosylation, disulfide bonds, conserved domains, motifs, catalytic mechanisms, molecular dynamics simulation (MD), and molecular docking. Databases like BRENDA, Enzyme Function Initiative (EFI), and ExPASy-ENZYME helped to explore different enzyme mechanisms involving examining molecular structures, activation-free energies, binding processes, chemical bond cleavage, and molecular interactions (ZHANG Jian-zhi, et.al., 2020) (Y. Zhang et al., 2020).

ProFun server and SCooP server were used to evaluate the functional traits and melting temperature of the stress-responsive acid phosphatase gene TaPase, and thermo/organotolerant BOD (A. D. E. V Sharma, 2020) (Sadeghian et al., 2020).

2.2.4 Protein interaction studies:

The protein interaction network was evaluated by using the TMHMM Server and the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) tool. E.g. These tools/servers/software was used to identify transmembrane motifs, interaction network and functional protein association in various enzymes/protein such as the insect trypsin enzymes (Hemmati & Mehrabadi, 2020) and novel abiotic stress proteins in *T.aestivum* (Gupta et al., 2018), in hypothetical proteins from the core genome of *C.pseudotuberculosis* biovar ovis (Carlos Leonardo Araújo, et.al, 2020).

2.2.5 Conserved domain, motifs, domains, peptide, epitope, glycosylation, disulfide bonds, antigenicity, and localization analysis:

Several studies evaluated the presence of conserved domain, motifs, peptides, epitope, glycosylation, disulfide bonds, antigenicity, and localization e.g. Endo-b-1,4-glucanase (Ariaeenejad et al., 2018), α -amylase (Motahar et al., 2020), and *Beauveria bassiana* chitinase (Bhagwat et al., 2021) enzymes were subjected to the Multiple Em for Motif Elicitation (MEME) program, Conserved Domain Database, SIB MyHits Motif scan and Pfam for evaluation of the conserved domain. Similarly, Pfam, CATH, SVM-Prot, CDART, SMART, ProtoNet, Self-Optimized Prediction Method with Alignment (SOPMA), SCRATCH, NetNGlyc, and GeneWise were used to understand the functional domains and families of hypothetical proteins (HPs) (Gupta et al., 2018)(S. Sahoo et al., 2020). Araújo et al. (2020) conducted functional annotations of 172 HPs from *C. pseudotuberculosis* biovar ovis genomes using the GO FEAT tool (Carlos Leonardo Araújo, Iago Blanco, Luciana Souza, Sandeep Tiwari, Lino César Pereira, Preetam Ghosh, Vasco Azevedo, Artur Silva, 2020).

Tan et.al. used SignalP to identify signal peptides in novel mesophilic phytase and novel AA9-type LPMO (Tan et al., 2015). PlantmPLoc, MultiLoc2, SherLoc2, WoLF Psort, TargetP, Psortdb and CELLO was used to find localization of HPs and chitinases (Bhagwat et al., 2021) (Gupta et al., 2018) (Guha et al., 2020) (Carlos Leonardo Araújo, et.al., 2020).

2.2.6 Molecular docking and Molecular Dynamics Simulation analysis:

AutoDock Vina (Bhagwat et al., 2021) (Yata et al., 2021), and SwissDock (Şahutoğlu et al., 2020) were used to evaluate the binding affinity of enzymes such as novel N-acetylglucosaminidase enzymes (Şahutoğlu et al., 2020), and potential as inhibitors for *M.tuberculosis* MurB (Nirwan et al., 2020).

GROningen MACHine for Chemical Simulations (GROMACS), AMBER (Assisted Model Building and Energy Refinement), CHARMM (Chemistry at Harvard Macromolecular Mechanics), Nanoscale Molecular Dynamics (NAMD) software, molecular mechanics (MM), and quantum mechanics (QM) approaches were used in different studies to evaluate dynamic behavior of proteins (ZHANG Jian-zhi, et.al., 2020) (Zhang et al., 2020).

Molecular dynamic simulation helped to identify thermally sensitive regions in Np-ADO and its mutants, Cts-ADO (Shakeel et al., 2018). There are number of studies which used MD simulation to study the dynamic behavior of enzymes such as novel N-acetylglucosaminidase (Şahutoğlu et al., 2020), Mycobacterium tuberculosis (Mtb)-MurB oxidoreductase enzymes (Nirwan et al., 2020) and D-lactate dehydrogenase (ZHANG Jian-zhi, et.al., 2020) (Y. Zhang et al., 2020) and *C.sativus* Urease (Yata et al., 2021).

Oyewusi et.al. investigated the binding of the substrate, mainly haloacids, and haloacetates, to the dehalogenase enzyme via a combination of docking and simulations. This study helped to gain information regarding potential reaction mechanisms of dehalogenases (Oyewusi et al., 2023).

2.3 Protein engineering approach:

Enzyme engineering, coupled with computational tools, plays a pivotal role in augmenting enzyme properties like thermostability, catalytic efficiency, and expanding their pH range. Efforts to enhance thermostability often revolve around targeted modifications, focusing on augmenting hydrogen bonding, introducing disulfide bridges, optimizing glycosylation, reducing loop flexibility, fortifying ionic and hydrophobic interactions, augmenting polar surface areas, and increasing the number of salt bridges. Amino acid composition is also critical; higher proportions of Arg, Tyr, and Ser, alongside increased helical content, are associated with heightened thermostability (S. Kumar et al., 2000). Researchers globally concentrate on diverse strategies, such as protein engineering approaches (directed or rational) and computational analyses, to boost enzyme characteristics. The following table presents such studies related to phytase engineering (**Table 2. 2**).

2.3.1 Directed evolution:

Directed evolution mimics natural evolutionary processes by creating DNA libraries from the gene encoding the enzyme of interest. Random mutations generate enzyme variants with improved properties, requiring High-Throughput Screening (HTS) to select the desired variants. This method has shown success in the lab, enhancing enzyme activity, specificity,

thermostability, and enantioselectivity without needing extensive structural or biochemical knowledge. Two distinct methodologies, namely random mutagenesis and HTS, were employed in the pursuit of identifying potential enzyme mutations through directed evolution. Random mutagenesis, a commonly utilized technique in this domain, initiates DNA replication under less favorable conditions, resulting in mutated populations that may exhibit enhanced capabilities compared to the original enzyme. DNA shuffling is one of the Directed evolution techniques that involved similar gene fragmentation, followed by PCR-based amplification that helped to obtain novel genetic combinations. However, the number of combinations generated poses challenges in shortlisting the desired variants. The HTS can help to overcome issues related to DNA shuffling. The properties and yield of the target enzyme play an important role in identifying mutants through directed evolution. There are several ways to screen mutant libraries, such as *in-vitro* and *in-vivo* selection, multi-enzyme setups, screening of plaque, and computational approaches (Victorino da Silva Amatto et al., 2022) (J. Zhao et al., 2021). E.g., the thermostability of *Yersinia mollaretii*, *Aspergillus niger* N25 strain, and *Escherichia coli* phytase was improved by using a directed evolution approach (Amol V. Shivange, Hans Wolfgang Hoeffken, Stefan Haefner, 2016) (Tang et al., 2018). Similarly, Xing et.al. improved thermostability of *Escherichia coli* phytase (Xing et al., 2023). The mutants K46E and K65E/K97M/S209G exhibited enhanced thermostability compared to the wild type. The mutation enabled the formation of hydrogen bonds (Kim & Lei, 2008). However, the application of directed evolution is limited because it is time ineffective and demands extensive screening processes (J. Zhao et al., 2021).

2.3.2 Semi-rational strategy:

The semi-rational approach is the combination of directed evolution and rational engineering methodologies. This method used sequence, and structure-based data along with computational methods to generate mutants, to reduce mutants in the screening process, and reliance on high-yield methods. The semi-rational engineering involves the evaluation of protein sequence, analysis of structure, random mutagenesis, and computational assessments (Victorino da Silva Amatto et al., 2022). The thermostability of *Citrobacter freundii* phytase and XgtA was enhanced by using a semi-rational engineering approach (Gordeeva et al., 2019) (L. Chen et al., 2022).

2.3.3 Rational engineering strategy:

Rational engineering is based on enzyme structure and functionality. The mutants were generated by sequence comparison and molecular dynamic simulation. This approach reduced the screening process to select potential mutants. Information on protein structure and function was obtained by using sequence-based and structure-based approaches (Victorino da Silva Amatto et al., 2022).

Sequence-based methods involve aligning a well-characterized homologous protein sequence with the target protein to identify potential amino acid alterations that could enhance enzyme properties. Conversely, structure-based techniques use the three-dimensional crystal structure of an enzyme to revamp its active site. In both scenarios, mutations can encompass substitution, insertion, or deletion of residues, as well as replacing amino acids with differing physicochemical properties and sizes (Victorino da Silva Amatto et al., 2022).

Recently, hybrid techniques such as structure-guided consensus, merging structure-based and sequence-based approaches, have emerged to offer insights into protein structure for rational design. Despite advantages over directed evolution, including smaller libraries and simplified screening, rational design encounters limitations, particularly when simultaneous substitutions are necessary. In contrast, directed evolution revolves around constructing libraries and screening mutants to uncover proteins with enhanced properties. The advancement of tools and models for rational design heavily relies on three fundamental aspects of enzyme information: accessibility, quality, and organization. Partial insights into these facets can be gleaned from public databases such as the Protein Data Bank and Universal Protein Resource, repositories housing curated data contributed by researchers (Victorino da Silva Amatto et al., 2022). E.g. improving enzyme thermostability is crucial for their industrial use, but no singular set of rules applies universally across enzymes. This study focuses on enhancing the thermostability of the haloalkane dehalogenase enzyme through computational methods. Leveraging the available crystal structure in the Protein Data Bank (PDB ID: 1EDE), various *in-silico* design strategies such as adjusting disulfide bond geometry, creating new hydrophobic pockets, forming fresh salt bridges, and applying multiple mutations (Satpathy et al., 2015).

Different approaches can be used, such as disulfide bonds addition, and introducing internal non-covalent interactions like hydrogen bonds, hydrophobic interactions, salt bridges, and π - π bonds, oligomerization and interaction between subunits, glycosylation, and truncation (Xiong et al., 2022) (Xu et al., 2019).

The types of interactions such as intramolecular hydrogen bonding, salt bridges, aromatic interactions, and hydrophobic interactions play a crucial role in stabilizing structure. Disulfide bonds have been identified as crucial for protein folding and stabilization, especially in thermophilic enzymes. Proteins can form more stable structures when organized into homodimers through hydrogen bonds, salt bridges, and hydrophobic interactions. The oligomerization of thermophilic enzymes and interactions between subunits significantly contribute to enzyme stability. Subunit dissociation is considered an initial inactivation step for multimeric proteins, and reinforcing non-covalent interactions between subunits is essential to prevent this dissociation. Flexible loops and disordered termini regions in proteins often contribute to instability. Truncation of these unstable regions restricts conformational flexibility, while cyclization of proteins, particularly linking the N- and C-termini, is a promising strategy for stability enhancement (Xu et al., 2019). E.g. Fei et al. enhanced the thermostability of *E.coli* AppA phytase by targeting protein flexibility, protein surface analysis, and increasing salt bridges (Fei et al., 2013). Similarly, Li et al. increased the thermostability of the *E.coli* phytase via temperature factor analysis (B-value) and amino acid surface modification (J. Li et al., 2019). Fakhravar et al. improved the enzymatic activity of thermostable *E.coli* phytase by using molecular dynamics simulation and rational engineering strategies. The generated mutant exhibited enhanced thermal resistance and binding interactions. The substitution with a non-polar amino acid (S392F) might have helped to create hydrophobic interactions, thereby enhancing the overall stability of the structure (Fakhravar & Hesampour, 2018).

Disulfide by Design (Craig & Dombkowski, 2013), Consensus sequence finder (Hua et al., 2020), Consurf (Celniker et al., 2013), and consensus analysis with known phytases, along with insights from existing literature (C. C. Chen & Tw, 2018) (X. Wang et al., 2015) can help to improve characteristics of enzymes. Additionally, the investigation involved tools like Hotspot Wizard (Bendl et al., 2016), Fireprot (Musil et al., 2017), Cabsflex (Jamroz et al., 2013), and FoldX (Schymkowitz et al., 2005).

Disulfide bonds are formed due to the oxidation of two non-adjacent cysteines. The bonds reduced the conformational entropy of the unfolded state (J. Zhao et al., 2021).

Maltotetraose-Forming Amylase (Yinglan Wang, Caiming Li, Xiaofeng Ban, Zhengbiao Gu, Yan Hong, 2022), Type A feruloyl esterase (Yin et al., 2015), Lipase B (Le et al., 2012), 1,3-1,4- β -glucanase (Niu et al., 2016), microbial transglutaminase (Suzuki et al., 2022), 2'-

deoxyribosyltransferases (Cruz et al., 2022), and many more were modified by using various tools such as DSDBASE-MODIP, Disulfide by Design2, SSBOND. These tools/servers were used for disulfide bond engineering.

Disulfide bond engineering was also utilized to improve the stability of phytase. E.g., a novel *Acidobacteria* phytase was subjected to disulfide bond engineering to improve thermostability, and half-life as compared to the wild-type candidate (Tan et al., 2016). Sanchez-Romero et.al. introduced disulfide bonds in *C.braakii* phytase by using molecular dynamics simulations at various temperatures. The introduction of disulfide bonds helped to stabilize the fluctuating region that was suggested in the MD simulation analysis (Sanchez-Romero et al., 2013). The consensus sequence-based approach was used to improve the thermostability of *A.niger* phytase (Han et al., 2018), *E.coli* AppA (T. Wu et al., 2014), and glutamate decarboxylase from *Lactobacillus brevis* (Hua et al., 2020). The combination of FireProt with either Disulfide by Design or FRESCO was used to identify hotspots in flavin-dependent monooxygenase (Pongpamorn et al., 2019), Ketoreductase ChKRED12 (Liu et al., 2021) and serine peptidase (Ashraf et al., 2019), and flavin reductase (Maenpuen et al., 2020). Similarly, Chen et.al. improved the thermostability of α -glycosidase extracted from *Xanthomonas campestris* (XgtA) by using FireProt (L. Chen et al., 2022). HotSpot Wizard was used to modify various enzymes such as lactate dehydrogenase (Luo et al., 2022), and FlAlyA (a PL7 alginate lyase) (X. Zhang et al., 2022). Similarly, FoldX, Rosetta ddg_monomer, and I-Mutant were used to enhance the thermostability of *Candida rugosa* lipases (G. Li et al., 2018) and *Rhizopus chinensis* lipase (Rui et al., 2020).

Table 2. 2 List of a few phytase engineering studies:

Reference paper	<i>E.coli</i> phytase		Mutant	Multiple /Point
	Target residue	Position		
(J. Li et al., 2019)	S	342	T	Multiple
	E	383	A	
	E	384	V	
	R	385	A	
	S	80	A	
	P	41	W	
	V	42	S	
	K	43	L	
	Q	285	D	
	K	286	Y	
(Kim & Lei, 2008)	K	46	E	Point
	K	46	E	Multiple

Chapter 2

	K	97	M		
	S	209	G		
(C. C. Chen & Tw, 2018)	A	143	C	Multiple	
	E	262	C		
	R	259	C	Multiple	
	G	312	C		
	V	205	C	Multiple	
	L	257	C		
	A	264	C	Multiple	
	N	309	C		
	A	143	C	Multiple	
	E	262	C		
	R	259	C		
	G	312	C		
	A	143		C	Multiple

Chapter 2

	E	262	C	
	A	264	C	
	N	309	C	
(X. Wang et al., 2015)	W	46	E	Multiple
	Q	62	W	
	A	73	P	
	K	75	C	
	S	146	E	
	R	159	Y	
	N	204	C	
	Y	255	D	
	Q	258	N	
	Q	349	N	
(Fakhravar & Hesampour, 2018)	S	392	F	Point
(J. Zhang et al., 2016)	K	24	E	Point

Chapter 2

(J. Zhang et al., 2016)	W	46	E	Multiple
	K	24	E	
Reference paper	<i>Yersinia</i> and <i>Citrobacter</i> phytase		Mutant	Multiple/ Point
	Target residue	Position		
(Nezhad et al., 2020)	F	89	S	Multiple
	E	226	H	
(Shivange et al., 2016)	T	77	K	Multiple
	G	187	S	
	Q	154	H	
	K	289	Q	
(Gordeeva et al., 2019)	K	46	M	Multiple
	K	138	E	
(Shivange et al., 2014)	D	52	N	Multiple
	T	77	K	
	K	139	E	

Chapter 2

	G	187	S	
	V	298	M	
(Stephane Blesa, Helene Chautard, Delcourt, Laurent Mesta, 2015)	K	210	S	Multiple
	Y	268	E	
	Q	292	P	
(Sanchez-Romero et al., 2013)	G	52	C	Multiple
	A	99	C	
	N	31	C	Multiple
	T	177	C	
	G	52	C	
	A	99	C	
	G	52	C	
	A	99	C	Multiple
	K	141	C	
	V	199	C	

Chapter 2

	N	31	C	Multiple
	T	177	C	
	G	52	C	
	A	99	C	
	K	141	C	
	V	199	C	

Chapter 3
Rationale,
Aim and Objectives

Rationale, Aim, and Objectives

3.1 Rationale

Phytase hydrolyzes its substrate phytate. Phytate is a major source of phosphate in cereals and legumes and is indigestible in monogastric animals due to very limited phytase activity in the digestive tract. Hence, it exhibits the potential to cause environmental pollution. Phytate also exhibits anti-nutritional properties by chelating micronutrients such as Ca^{2+} , Fe^{2+} , and Zn^{2+} , by forming complexes such as protein-phytate, starch-phytate, and lipid-phytate, and reduces bioavailability, which overall decreases the nutritional value of feed (Luo et al., 2009) (Kalsi et al., 2016) (Jaiwal et al., 2019) (Brinch-Pedersen et al., 2014). Hence, phytase is used as a feed additive in the feed industry to hydrolyze phytate to lower myoinositol derivatives and release phosphate. Thereby, it reduces the supplementation of inorganic phosphate in feed (which is obtained from non-renewable resources and accounts for 9% of the average feed cost), reduces the cost of feed which, accounts for 60-70% of the total cost in livestock farming, as well as aid in natural resource management. It also increases the bioavailability of micronutrients, which helps to maintain the high nutritional standard of feed and reduces environmental pollution by reducing eutrophication and deposition of low phytate content manure in soil (Jaiwal et al., 2019) (Brinch-Pedersen et al., 2014) (Cang et al., 2004) (*Industrial Enzymes Market - Global Forecast by 2022*, 2017).

In addition to its wide application in the feed industry, phytases are also used as bread-making additives (Rao et al., 2009). In the bread-making industry, it helps reduce the fermentation time and improves bread texture (Rao et al., 2009). Phytases find use in aquaculture to counteract the detrimental effects of phytate present in fish feed. Their application contributes to enhancing fish growth and bone mineralization by improving mineral availability. Moreover, they can boost protein digestibility and decrease the overall phosphorus burden in aquatic settings (Rao et al., 2009). Microbial phytases are also utilized in soil amendment, leading to improvements in soil quality and enhancing the growth and quality of plants (Rao et al., 2009). Besides that, phytases are used to synthesize lower myoinositol phosphates by enzymatic method. Enzymatic methods are environmentally friendly as compared to that of chemical synthesis methods. These inositol phosphate derivatives can be used as enzyme stabilizers, enzyme substrates, or inhibitors of enzymes (Rao et al., 2009) (Kaur et al., 2021). It is also proposed to apply it as a painkiller and act as an inhibitor against retroviral infections including HIV. Hence, inositol phosphates that can be synthesized by using phytases have potential

therapeutic applications (Kaur et al., 2021). Hence, in recent years, the use of phytase has extended beyond the feed-food industry and has found applications in various other fields. These include nutraceuticals, soil amendments, and the production of myoinositol phosphates. Commercial phytases like Natuphos® or Allzyme find extensive use in the feed industry, but their production is expensive, which ultimately restricts their application as feed additives (Vohra, and Satyanarayana 2021). Furthermore, apart from the constraints, the production industry of inositol phosphates also requires new phytases that possess the capability to achieve stereospecificity, ensuring precise configuration. This is crucial to offer a sustainable alternative to the chemical synthesis process (Kaur et al., 2021). Phytase applications are also limited in soil amendment due to several factors, including the soil environment (such as pH, temperature, and presence of cations), the presence of inhibitors, and the decreased availability of phytate because it binds tightly to soil components (Menezes-Blackburn et al., 2013).

To address these challenges, we aim to obtain novel phytases through different approaches such as conventional, metagenomics, and *in-silico* bioprospecting; and further improve their properties via a protein engineering approach.

3.2 Aim

To obtain a novel phytase enzyme using bioprospecting and protein engineering approach.

3.3 Objectives

- Isolation and screening of phytase-producing microbes from natural sources or *in-silico* bioprospecting.
- *In-silico* characterization and experimental validation of shortlisted phytase.
- *In-silico* hotspot identification and experimental validation of shortlisted phytase.

Chapter 4

Isolation and screening of phytase-producing microbes from natural sources or *in-silico* bioprospecting

Isolation and screening of phytase-producing microbes from natural sources or in-silico bioprospecting.

4.1 Introduction:

This chapter delves into the exploration of new phytases through three distinct approaches: conventional, metagenomics, and *in-silico*. The conventional approach involved the screening of phytase-producing microorganisms from different natural resources such as soil. The metagenomics approach employed 16S rRNA gene sequencing of total soil DNA to investigate the overall microbial diversity and identify phytase-producing microorganisms. In the *in-silico* bioprospecting approach, researchers utilized various databases to screen microorganisms capable of producing phytase.

4.1.1 Conventional Bioprospecting:

Researchers have been exploring phytase producers from natural resources such as soil. E.g., Kumar et.al. isolated phytase producers from environmental samples i.e., Himalayan soils, and identified *Acromobacter* sp. PB-01, *Tetra- thiobacter* sp. PB-03 and *Bacillus* sp. PB-13 is a potential phytase producer. They helped in the growth and phosphorous uptake of *Brassica juncea* (Kumar et al., 2013). Puppala et.al. characterized a novel acidic and thermostable phytase isolated from *Streptomyces* sp. (NCIM 5533) from Sanjivani I-lands Maharashtra, India. It also acts as a plant growth-promoting bacteria which was evident as it supports the growth of *Solanum lycopersicum* (Puppala et al., 2019). In general, the conventional bioprospecting approach favors the cultivation of microbes that can be grown in the laboratory.

4.1.2 Metagenomics approach:

Soil consists of soil biomass, minerals (different sizes, shapes, and chemical characteristics), and organic matter. The largest portion of the soil mass is composed of prokaryotes. However, the number of prokaryotic cells in the soil mass differs from sample to sample e.g., one gram of forest soil composed of 4×10^7 prokaryotic cells, and one gram of cultivated soil and grasslands consisted of 2×10^9 prokaryotic cells. There are two methods to explore this vast soil microbial diversity: cultivation-dependent based methods and cultivation-independent methods. Cultivation-based methods can only help to explore 0.1-1% of the total microbial population present in soil. To overcome this, cultivation-independent methods such as metagenomics are developed (Daniel, 2005). The first step in the soil metagenomics method is

the isolation of total soil gDNA extraction followed by 16S rRNA gene amplification to explore the diversity of microbial communities (A. Kamble et al., 2020). These microbial communities can be explored to find new enzymes such as phytases e.g., Tan et.al., identified novel phytase genes from an agricultural soil via a metagenomics approach (Tan et al., 2014).

4.1.3 *In-silico* bioprospecting:

Database mining can also help to discover novel phytases e.g., Tan et.al. explored IMG/M database and discovered 11 putative HAP phytase genes from 18 publicly available metagenomes of fungus gardens. Among these rPhyXT52 phytase exhibited the highest activity at 52.5 °C (Tan et al., 2015). In another study, Tan et.al. used metagenomic datasets of an acidic peat-soil microbiome in northeastern Bavaria, Germany by using Pfam00328 (Pfam identifier) to explore novel phytases (Tan et al., 2016).

4.2 Conventional Bioprospecting:

4.2.1 Introduction:

This approach aims to identify and isolate phytase-producing microbes from different soil samples (Nagar et al., 2021) (**Figure 4. 1**).

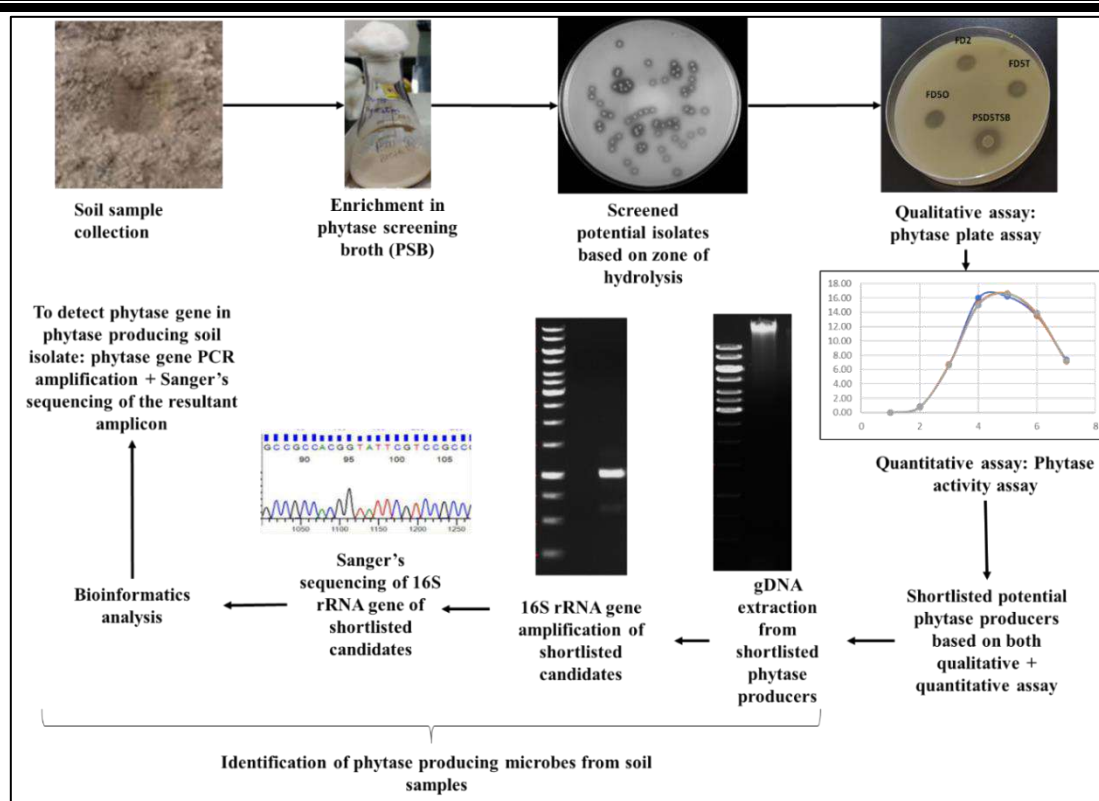


Figure 4. 1 Overview of the conventional bioprospecting method used in the present study.

4.2.2 Materials and Methods:

4.2.2.1 Materials (Chemicals, kits, reagents, solutions, plasticware):

Sodium phytate which was used as substrate was procured from HiMedia Laboratories Pvt Ltd. The Invitrogen PureLink™ Genomic DNA kit (Cat. No. K182001) was acquired from ThermoFisher Scientific, India. Molecular grade agarose powder was purchased from MP Biomedicals, USA. All chemicals were procured from either Sigma-Aldrich, MP Biomedicals, Loba Chemie, or MolyChem. All plasticware utilized in the experiments, including 15ml and 50ml falcon tubes, were obtained from SPL Life Sciences Pvt. Ltd., India. The sterile autoclavable plastic bags were ordered company. All microbiological media such as Luria Bertani and tryptic soy broth and agar were procured from HiMedia Laboratories Pvt Ltd. The 16S rRNA gene primers were acquired from Sigma-Aldrich. The primers that were used to amplify the phytase gene and full-length 16S rRNA gene (Europhins make) were procured from Excel Biosolutions. 2x Emerald Amp 07 PCR Master mix (Takara), DNA ladder (GeneRuler). B-cell lysis buffer (Sigma Aldrich). Phenylmethylsulfonyl fluoride (PMSF), Low melting agarose from GeNei, sieve, Remi incubator shaker, SerialCloner software 2.6.1, NucleoSpin® Gel, and PCR Clean-up kit.

4.2.2.2 Soil sample collection:

Soil samples were collected as per the guidelines given by the National Agricultural Portal of India (http://agritech.tnau.ac.in/agriculture/agri_soil_sampling.html). The field was divided into different sections. The section was cut in a 'V' shape, and five soil samples from each section were collected in a sterile biohazard plastic bag and mixed thoroughly (**Figure 4. 2**).

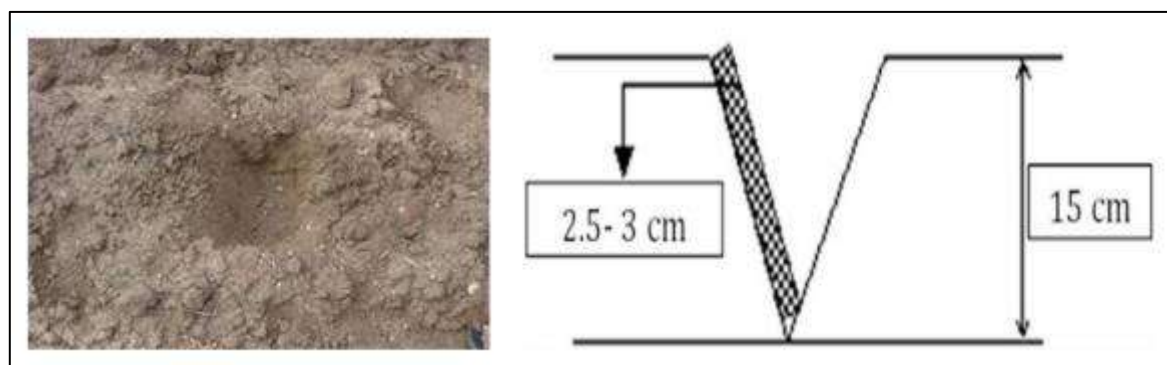


Figure 4. 2 Method of soil sample collection adopted from the National Agricultural Portal of India.

Details such as the location of the farm, the previous crop grown, the present crop, the crop to be grown in the next season, the date of collection, the name of the sampler, and the season were mentioned on the bag. The samples were transported in ice boxes to the lab of Sunandan Divatia School of Science, NMIMS (Deemed to be University), Mumbai. They were dried in a hot air oven for 30 minutes. The soil was spread on blotting paper and powdered by breaking the lumps. These soil samples were then passed through a 2mm sieve to remove litter and lumps, and the processed samples were transferred to sterile falcons. The Falcons were again labeled with information such as the location of the farm, the previous crop grown, the present crop, the crop to be grown in the next season, the date of collection, the name of the sampler, and the season. These processed soil samples were then stored at 4°C.

We selected eleven different sites located in Maharashtra and Gujarat. The sources of soil samples included forest soil (Forest area in Ankleshwar; Gujarat), garden soil (Trikon garden, Mumbai, Maharashtra), and fields with rice, oilseeds (Kursani), udad (pulse), nachini, millet (jowar), vegetation (ladyfinger), cotton, as well as swine enclosures collected from Baruch (Gujarat) and Jawhar district (Maharashtra). The details of soil samples are given in

Table 4. 1.

Chapter 4

Table 4. 1 The details of soil samples collected from different regions of Maharashtra and Gujarat (Nagar, Kamble and Singh, 2021).

Soil sample source	Location	Geographical coordinates	Soil type	Agroecology	Temperature
Forest	Ankleshwar, Gujrat	21.58102° N, 73.04682° E	Forest soil	Mango, guava, and drumstick trees had grown in a 24281 m ² forest land	24 °C
Garden	Mumbai, Maharashtra	19.09935° N, 72.84850° E	Clay soil	Roses, Aloe vera, tulsi, herbs, and shrubs were grown in a 790 m ² area garden	22 °C
Rice, oil seed (kursani), udad (pulse), nachini (millets)	Jawhar, Maharashtra	19.91855° N, 73.23496° E	Hill soil	The crops such rice, oil seed, udad and nachini were grown together in a 4000 m ² field.	26 °C
Swine	Jawhar, Maharashtra	19.91855° N, 73.23496° E	Silt soil	The swine enclosures were maintained for 20 and 4 years, respectively.	26 °C
Millets (jowar), vegetation (lady finger), and cotton	Baruch, Gujrat	21.68931° N, 72.89728° E	Loam soil	These crops were grown on 4047 m ² of land each. The crops were not harvested	24 °C

4.2.2.3 Enrichment, screening, and isolation of bacteria:

Soil samples were enriched in phytase screening broth (PSB) to support the growth of phytate-degrading microbes. The composition of PSB is listed in **Table 4. 2**. Phytase-specific broth is used as an enrichment media in which glucose is a carbohydrate source, ammonium nitrate is the source of nitrogen, and calcium chloride, along with other essential nutrients, helps in enrichment. Sodium phytate is the specific substrate for phytate-degrading bacteria, which hydrolyzes phytate and releases inorganic phosphorous.

One gram of soil samples was inoculated in sterile 50ml PSB. It was then incubated at 37°C and 180 rpm on a rotary incubator shaker, overnight. The enriched broth cultures were serially diluted (till 10^{-5}) and spread-plated on PSA. PSA plates were incubated at room temperature. The plates were observed 24 hours, 48 hours, and 96 hours of incubation.

The phytase-producing microbes were identified based on zones of hydrolysis and were compared to the *E.coli* zone of hydrolysis. The pure cultures of potential phytase-producing microbes were obtained by culturing and isolating them on Tryptic Soy (TS) media. The cultures were incubated at 37°C and 180rpm, overnight.

Table 4. 2 Phytase screening broth (PSB) composition:

Note:

1. D-glucose and Sodium phytate should be filtered sterilized and added after autoclaving.
2. Add 3% agar in PSB to prepare phytase screening agar (PSA).

Components	gm/L of Distilled water and respective (%)
D-glucose	10 (1%)
Sodium phytate (Himedia)	4 (0.4%)
Calcium chloride (CaCl ₂)	2 (0.2%)
Ammonium nitrate (NH ₄ NO ₃)	5 (0.5%)
Potassium chloride (KCl)	0.5 (0.05%)
Magnesium sulfate heptahydrate (MgSO ₄ .7H ₂ O)	0.5 (0.05%)
Ferrous sulfate heptahydrate (FeSO ₄ .7H ₂ O)	0.01 (0.001%)
Manganese sulfate heptahydrate (MnSO ₄ .7H ₂ O)	0.01 (0.001%)

4.2.2.4 Qualitative phytase activity:

The qualitative phytase activity was estimated based on the zone of hydrolysis that was formed due to phytate degradation on PSA by phytase-producing microbes. The method involved spotting about 10µl of soil isolates and *E. coli* BL21 (DE3) culture on PSA. The plates were incubated at room temperature for 96 hours and then stained using a two-step counterstaining method. The two-step counterstaining method involved flooding of PSA plate with cobalt chloride solution (2% (w/v)) for 5 min and then replacing it with 1:1 (v/v) aqueous ammonium molybdate (6.25% (w/v)) and ammonium vanadate solutions (0.42% (w/v)) for 2 to 3 min. This staining method eliminated false-positive results (Bae et al., 1999). The positive phytase-producing microbes exhibited a clear zone of hydrolysis did not regain the turbidity and remained Colorless. The estimation was performed in duplicates.

4.2.2.5 Molecular identification of bacterial isolates:

The molecular identification method involved the isolation of gDNA from bacterial isolates, followed by 16S rRNA amplification and sequencing. The gDNA isolation was performed using the PureLink® Genomic DNA kit (Cat. No. K182001). The purity and quantity of gDNA were estimated on BioTek nanoplates and by using BioTek Gen5 software. The integrity of gDNA was assessed by electrophoresis on 1% agarose gel.

These extracted gDNA from each soil isolate were used as templates to amplify its 16S rRNA gene by using the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and reverse primer Y36 (5'-GAAGGAGGTGTCCADCC-3'). The PCR conditions and reaction mixture conditions are mentioned in **Table 4. 3** and **Table 4. 4** below. The PCR products were electrophoresed in a 0.8% agarose gel with 1X TAE as the buffer medium, at 100 volts against 1 kbp DNA ladder. The PCR product was purified using the NucleoSpin® Gel and PCR Clean-up kit. The resultant amplicons were also subjected to Sanger's sequencing with forward primer i.e., 515F. The 16S rRNA gene sequencing results were analyzed using the NCBI Blastn tool. The nr database was chosen to screen and identify the closest homolog, and other parameters were kept default. The MEGA 7.0 software was used to construct a neighbor-joining and maximum likelihood tree. First, the multiple sequence alignment of the closet homologs and the soil isolates 16S rRNA gene sequences were built using the MUSCLE tool embedded in MEGA7 software. This alignment was used to construct a neighbor-joining phylogenetic tree and a maximum likelihood tree (1000 bootstrap, method: Jukes-Cantor model).

Chapter 4

Table 4. 3 PCR reaction mixture to amplify 16S rRNA gene.

Components	Volume required per reaction
Template*	1 μ l
Forward primer (515F)	1 μ l
Reverse primer (Y36)	1 μ l
2X PCR Emerald mix	10 μ l
Sterile distilled water	7 μ l
Total reaction volume	20 μ l

*Template**: gDNA of each soil isolate.

Note: Non-template control (NTC) was also included in the process. In the case of an NTC reaction mixture includes all the above components except the template. Distilled water (1 μ l) was added instead of the template in the NTC reaction mixture.

Table 4. 4 16S rRNA gene PCR amplification conditions.

Steps	Temperature ($^{\circ}$ C)	Time
Initial denaturation	95	3 min
Denaturation	95	45 sec
Annealing	54	30 sec
Initial extension	72	1 min 30 sec
Final extension	72	10min
Store	4	infinite

4.2.2.6 Quantitative effect of pH on phytase activity:

The soil isolates and *Escherichia coli* culture were inoculated in 10ml LB broth and incubated at 37 $^{\circ}$ C and 180 rpm overnight. 1ml of culture was collected and centrifuged at 10000 rpm,

Chapter 4

25°C for 10 minutes. The supernatant was discarded, and the pellets were stored at -20°C. To the samples, 0.4ml of B-cell lysis buffer and 5µl of 2% PMSF were added. Each sample was freeze-thawed by vortexing for 5 minutes and kept on ice for 2 minutes, for a total time of 15 minutes. This mixture was centrifuged at 10000 rpm, 4°C for 10 minutes, and the supernatant (crude lysate) was collected for enzyme activity analysis at pH 2.5 and pH 5.

The standard to estimate the released inorganic phosphate was constructed by using potassium dihydrogen phosphate (KH₂PO₄) at concentrations ranging from 0.1 mM to 2 mM (Table 4.5). The termination buffer was freshly prepared by adding 5 ml of 100% acetone, 2.5 ml of 5% ammonium molybdate, and 2.5 ml of 5N sulfuric acid.

The determination of statistical significance for differences involved the use of Student's t-tests and two-way ANOVA, both conducted using GraphPad Prism 9.0.0 software. The mean values were compared at a significant level of 5%.

Table 4. 5 Reaction setup for KH₂PO₄ standardization.

KH ₂ PO ₄ concentration (mM)	KH ₂ PO ₄ (ml)	Distilled water (ml)	Phytate* (ml)	Distilled water (ml)	Incubation (Temperature and Time)	Termination buffer (ml)
0.1	0.002	0.048	0.05	0.4	37°C for 30 min	0.5
0.25	0.005	0.045				
0.5	0.010	0.040				
1	0.020	0.030				
1.5	0.030	0.020				
2	0.040	0.010				
50	0.050	-				
Blank	-	0.050				

**Phytate stock solution: 44mM. Phytate working solution: 1:50 dilution made from the 44mM phytate stock in the buffers (based on pH).*

Chapter 4

The modified Heinonen method was used to estimate the release of inorganic phosphate, which was produced due to the phytate degrading activity of soil isolates and phytases at pH 2.5 and 5. The reaction setup for estimating total inorganic phosphate in the test sample was prepared as per the following **Table 4. 6**. Glycine-HCl (0.1M) and sodium acetate (0.1M) were used as buffer systems for pH 2.5 and 5, respectively.

Table 4. 6 Reaction setup for test samples.

Components	Test sample	Substrate blank	Enzyme blank	Combination blank
Substate* (ml)	0.05	0.05 (reaction buffer#)	0.05	0.05 (reaction buffer#)
Enzyme** (ml)	0.05	0.05	0.05 (B-cell lysis buffer)	0.05 (B-cell lysis buffer)
Reaction buffer# (ml)	0.4			
Incubation (Temperature and Time)	37°C for 30 min			
Termination buffer (ml)	0.5			

**Prepared 1:50 diluted phytate (stock 44mM) in the respective buffer*

#Reaction buffer: pH 2.5 Glycine-HCl buffer and pH 5 Sodium Acetate

***Enzymes added in test samples and substrate blanks were the supernatant of each sample*

After adding the termination buffer, the reactions were centrifuged at 14000 rpm, 25°C for 5 minutes. 100µl of the supernatant was added to the 96 well plates and the absorbance was read at a wavelength of 400 nm. The readings were taken in duplicates. The concentration of inorganic phosphate from the sample reaction was calculated from the standard graph of KH_2PO_4 . One unit of phytase activity (U/mL) was defined as the amount of phytase enzyme

Chapter 4

required to liberate 1 mmol of inorganic phosphate per minute by utilizing sodium phytate as the substrate under assay conditions.

Enzyme activity was measured from the equation:

$$\begin{aligned}\text{Enzyme Activity(U)} &= \frac{\text{Product Conc } (\mu\text{M}) \times \text{Reaction Vol (ml)}}{\text{Enzyme Vol (ml)} \times \text{Time (mins)}} \\ &= \frac{\text{Inorganic phosphate Conc } (\mu\text{M}) \times 1 \text{ ml}}{0.05 \text{ ml} \times 30 \text{ mins}}\end{aligned}$$

Protein estimation:

Bovine Serum Albumin (BSA) was used for protein standardization, and the reaction setup was prepared as per **Table 4. 7**.

Table 4. 7 Reaction setup for BSA standardization. Note: Stock solution of BSA: 10mg/ml.

BSA Sample ($\mu\text{g}/\mu\text{L}$)	BSA (μl)	Distilled water (μl)	Total volume (μl)
Blank	-	100	100
10	100	-	
8	80	20	
6	60	40	
4	40	60	
2	20	80	
1	10	90	
0.8	8	92	
0.6	6	94	
0.4	4	96	
0.2	2	98	
0.1	1	99	

For protein estimation, 2 µl of the test/ standard samples were placed on the nanodrop plate and measured for their absorbance.

4.2.2.7 Full-length PSD 16S rRNA gene amplification and sequencing:

The full-length 16S rRNA gene primers of the PSD isolate were constructed using its partial 16S rRNA gene sequence in FASTA format. The partial 16S rRNA gene (Sanger sequencing via 515F primers) in FASTA format was used as the input in the Blastn search. The Blastn search criteria included *Klebsiella variicola* (*K.variicola*) in the inclusion section because, according to the analysis of the partial 16S rRNA gene sequence, *Klebsiella variicola* was the closest homolog. The search results included a list of the closest *K. variicola* strains. The sequence was obtained by clicking on its graphics section and extracting the exact sequence. In total, 31 sequences were shortlisted to design primers. The details are mentioned in **Table 4. 8**. Clustal Omega was used to create a multiple sequence alignment (MSA) of the selected 16S rRNA *K.variicola* gene sequences. Based on MSA, the forward primer sequence (5'ATCATGGCTCAGATTGAACGC3') and reverse primer sequence (5'GGTCCCCCTACGGTTACCTTG3') were designed to amplify the full-length 16S rRNA gene sequence. The PCR reaction mixture and conditions were kept the same as mentioned for the partial 16S rRNA gene amplification of soil isolates, except for the annealing temperature. Here, the annealing temperature was optimized by amplifying the full-length 16S rRNA gene of PSD at different temperatures (54 to 60°C). The amplicons were pulled together into a single Eppendorf tube and purified using the NucleoSpin® Gel and PCR Clean-up kit. The amplicon of size 1504 bp was expected. The purified PCR amplicon was then subjected to Sanger's sequencing using both forward and reverse primers.

Table 4. 8 The details of different *Klebsiella variicola* strain 16S rRNA gene sequences used for designing 16S rRNA primers, to amplify full-length 16S rRNA gene of PSD isolate.

Sr. No.	Strain	Accession I.D.	Position
1	<i>K. variicola</i> strain FDAARGOS	CP050958.1	c2690493-2688944
2	<i>K. variicola</i> strain AHKv	CP047360.1	447751-449300
3	<i>K. variicola</i> strain LEMB11	CP045783.1	3116074-3117623
4	<i>K. variicola</i> strain WCHKP19	CP028555.2	445167-446716
5	<i>K. variicola</i> strain NCTC9178	LR588409.1	1695003-1696529
6	<i>K. variicola</i> strain X39	CP018307.1	c4279692-4278139
7	<i>K. variicola</i> strain E57-7	CP016344.1	1186214-1187767
8	<i>K. variicola</i> strain LMG 23571	CP013985.1	1220309-1221862
9	<i>K. variicola</i> strain KPN1481	CP020847.1	2467130-2468683
10	<i>K. variicola</i> strain GJ3	CP017289.1	211437-212990
11	<i>K. variicola</i> strain RHBSTW	CP056145.1	451902-453451
12	<i>K. variicola</i> At-22	CP001891.1	442637-444164
13	<i>K. variicola</i> strain FDAARGOS_627	CP044050.1	5391785-5393334
14	<i>K. variicola</i> strain 15WZ	CP032354.1	c134599-133050
15	<i>K. variicola</i> strain WCHKV030666	CP027064.2	466410-467959
16	<i>K. variicola</i> strain GJ2	CP017849.1	211437-212990
17	<i>K. variicola</i> strain GJ1	CP017284.1	211437-212990
18	<i>K. variicola</i> strain DSM 15968	CP010523.2	456048-457601
19	<i>K. variicola</i> strain DX120E	CP009274.2	c5346138-5344585
20	<i>K. variicola</i> strain AC CHC	CP066870.1	c2604084-2602545
21	<i>K. variicola</i> strain 13450	CP026013.1	444767-446316
22	<i>K. variicola</i> strain F2R9	CP072130.1	455673-457212
23	<i>K. variicola</i> strain E25	CP067261.1	453752-455291
24	<i>K. variicola</i> strain 13450	CP030173.1	479490-481039
25	<i>K. variicola</i> strain 118	CP048379.1	446643-448192
26	<i>K. variicola</i> strain FH-1	CP054254.1	454393-455942
27	<i>K. variicola</i> strain 8917	CP063403.1	c4332360-4330822
28	<i>K. variicola</i> strain 342	CP000964.1	1223043-1224574
29	<i>K. variicola</i> strain GN02	CP031061.1	454734-456283
30	<i>K. variicola</i> strain KPN029	CP065162.1	447598-449137
31	<i>K. variicola</i> strain KP2757	CP060807.1	c4410622-4409073

4.2.2.8 Phytase gene amplification:

The *K. variicola* phytase protein sequence was retrieved from the UniProtKB database. The criteria used to shortlist the protein sequences: the length of amino acid should be in the range of the known phytases i.e., between 400 to 500, and it should possess the conserved regions i.e., C-terminal conserved region HD and N-terminal conserved region i.e., RHGXRXR. The selected sequence, *Klebsiella variicola* (Accession I.D.: A0A087FHH8), fulfilled both the criteria, and the nucleotide gene sequence was retrieved from the ENA database. This gene sequence was used as a template to design full-length as well as mid-length phytase gene primers, thereby aiding in the confirmation of the PSD soil isolate phytase gene sequence.

The gDNA of PSD was extracted using the PureLink® Genomic DNA kit (Cat. No. K182001) and was used as a template to amplify full-length and mid-length phytase gene sequences. The PCR reaction mixture and conditions are mentioned in **Table 4. 9** and **Table 4. 10** respectively below. The full-length PCR amplicons (approximately 1.4kbp size) were pulled together and resolved on 1% low melting agarose. The amplicon was extracted from the gel and purified using the NucleoSpin® Gel and PCR Clean-up kit. The purified PCR amplicon was subjected to Sanger's sequencing. The sequencing results were analyzed using BLASTX of NCBI tool and Serial Cloner software.

Table 4. 9 PCR reaction mixture to amplify phytase gene from PSD.

Components	Volume/ reaction
Template*	1µl
Forward primer*	1µl
Reverse primer *	1µl
2X PCR Emerald mix	15µl
Sterile distilled water	7µl
Total reaction volume	25µl

Chapter 4

Note: Template*: gDNA of PSD.

Forward primer*:

For full-length phytase gene amplification: 5'- ATG CCT GCA AGA CAT CAG GG -3'

For mid-length phytase gene amplification: 5' CAG ACC GAC AAG TTC GCC GC 3'

Reverse primer *

For full-length phytase gene amplification: 5'- CGG CAG GAC CAT GGC TAC CG -3'

For mid-length phytase gene amplification: 5'- TGG GCC ACC AGC AGC AGC CA -3'

Note: Non-template control (NTC) was also included in the process. In the case of NTC, the reaction mixture included all the above components except the template. Distilled water (1 μ l) was added instead of the template in the NTC reaction mixture.

Table 4. 10 PSD phytase gene PCR amplification conditions

Steps	Temperature (°C)	Time
Initial denaturation	95	3 min
Denaturation	95	45 sec
Annealing	56 to 66 (at intervals of 2°C)	30 sec
Initial extension	72	1 min 30 sec
Final extension	72	10min
Store	4	infinite

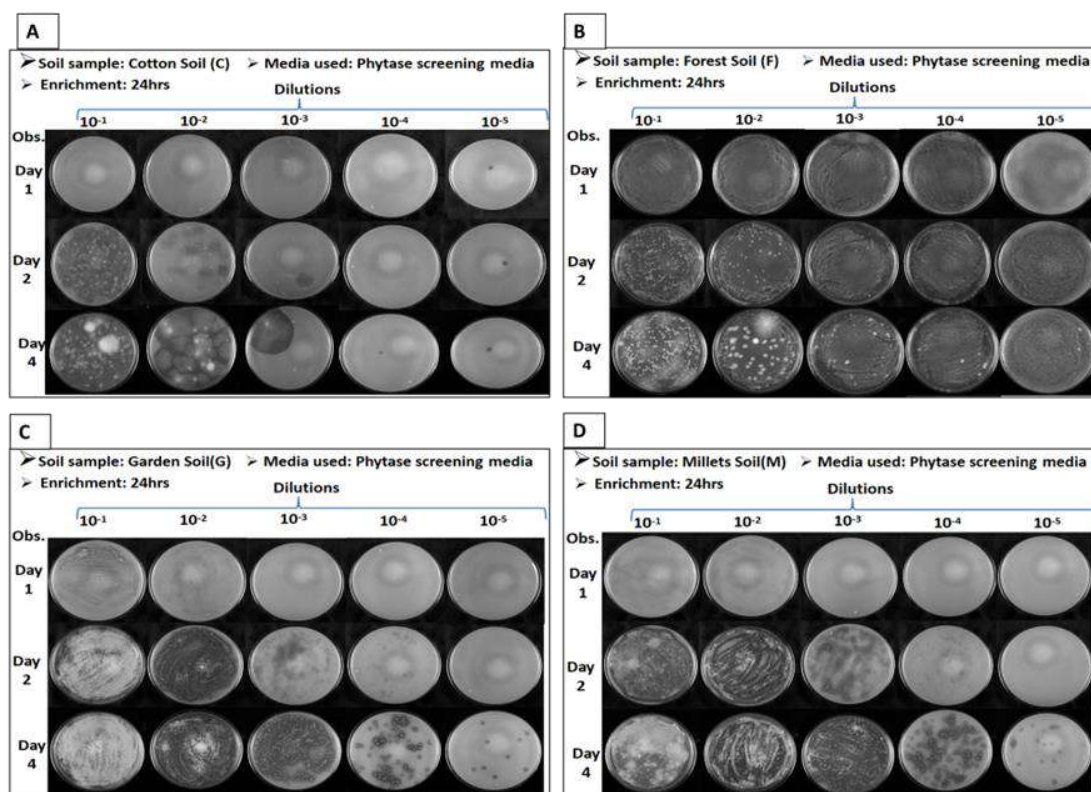
4.2.3 Results:

4.2.3.1 Soil sample collection:

Soil samples were collected from distinct locations in Maharashtra and Gujarat. The forest soil was collected from Ankleshwar (Gujarat); garden soil from Mumbai; paddy areas, oilseeds, millets, and pulses fields, poultry fields, and swine fields in Jawhar, Maharashtra; and vegetation, millets, and cotton fields in Baruch, Gujarat. Eleven soil samples were collected. The sampling procedure followed the prescribed guidelines from the National Agricultural Portal of India. Following collection, the soil samples underwent sieving (2mm) to remove any unwanted debris or clumps, and the processed samples were stored at a temperature of 4°C.

4.2.3.2 Enrichment, screening, and isolation of bacteria:

To enrich phytase-producing microbes, processed soil (1g) was inoculated in PSB. The enriched broths were diluted serially, followed by spread plating on phytase screening agar plates (PSA). The appearance of a hydrolysis zone around the microbes indicated the release of phytase and the subsequent degradation of phytate present in the agar (**Figure 4. 3**). The phytase-producing microbes were isolated and cultured on tryptic soy (TS) to obtain a pure culture.



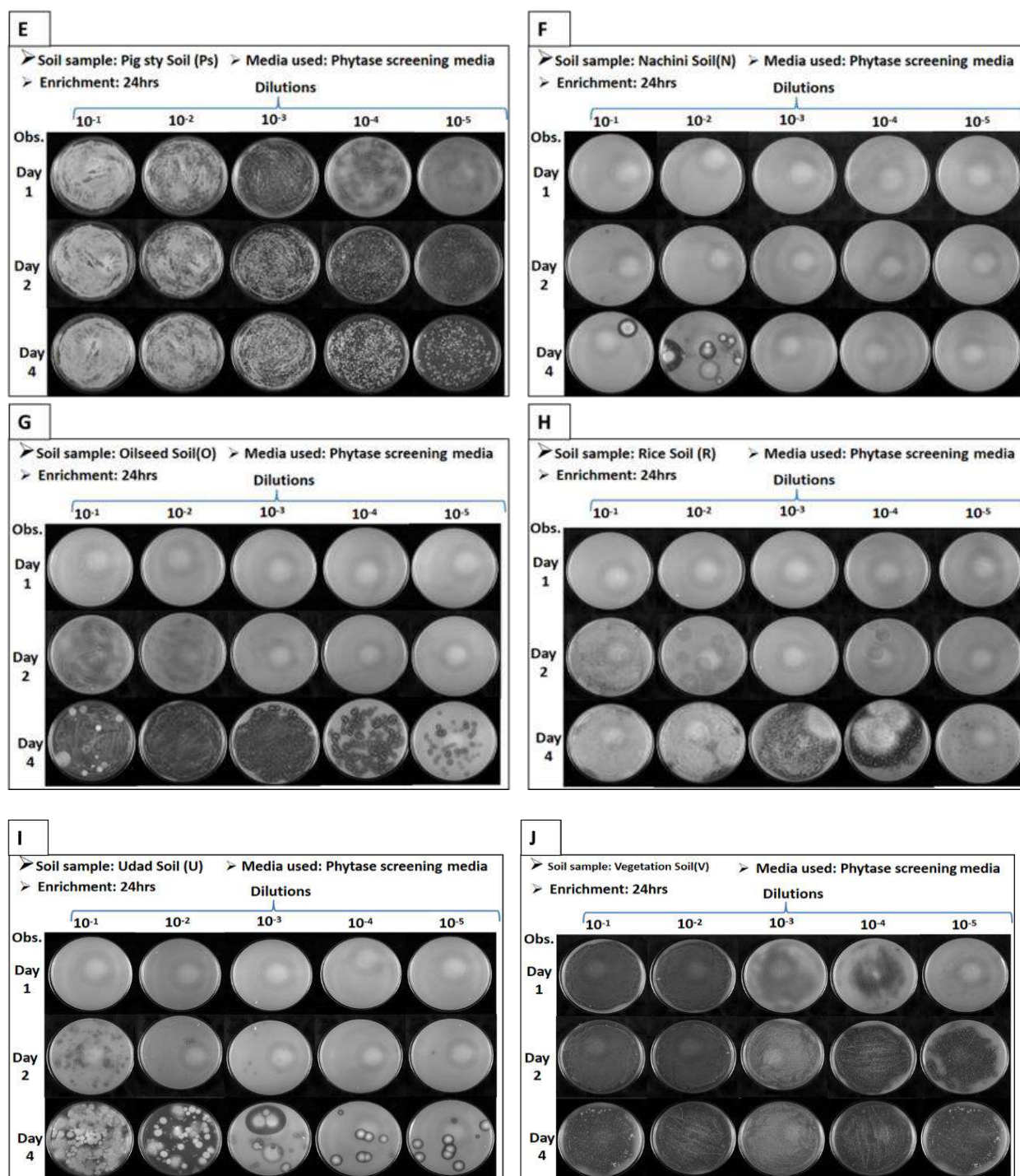


Figure 4. 3 Enrichment of following soil samples in PSB for 24hrs to support growth of phytase-producing microbes. The enriched soil samples were serially diluted and spread-plated on PSA. The PSA plates were observed for 24hrs (day 1), 48hrs (day 2), and day 4 (120hrs) to shortlist candidates based on the zone of hydrolysis formed due to phytate degradation. The following are the soil samples that were enriched in PSB and plated on PSA. A. Cotton soil; B.

Chapter 4

Forest soil; C. Garden soil; D. Millet soil; E.Pig sty soil; F. Nachani, G. Oil-seed; H.Rice; I. Udad; J. Vegetation soil.

A total of 6 potential isolates were obtained through enrichment, followed by preliminary screening via the phytase plate method. The potential soil isolates, namely FD2, FD5T, and FD5O, were screened from forest soil, whereas VD5, CD5, and PSD5TSB were obtained from vegetation, cotton, and pig stag soil samples.

4.2.3.3 Qualitative phytase activity:

The above-maintained isolates and *E.coli* culture were spot plated on PSA and incubated for 96-hr. The two-step counter-staining method was adopted, which included first, the flooding of PSA plates with 2% (w/v) cobalt chloride solution (room temperature, 5 min). Next, the solution was replaced with a 1:1 v/v mixture of aqueous ammonium molybdate 6.25% (w/v) and ammonium vanadate solutions 0.42% (w/v) for 2 to 3 min. The area of the zone of hydrolysis was cleared and failed to regain turbidity, which indicated that these were positive isolates (**Figure 4. 4**). The two-step counterstaining method helped to eliminate false-positive results. The diameter of the zone of hydrolysis produced by each isolate was compared to that of *E. coli* as a reference. The qualitative analysis was performed in duplicate.

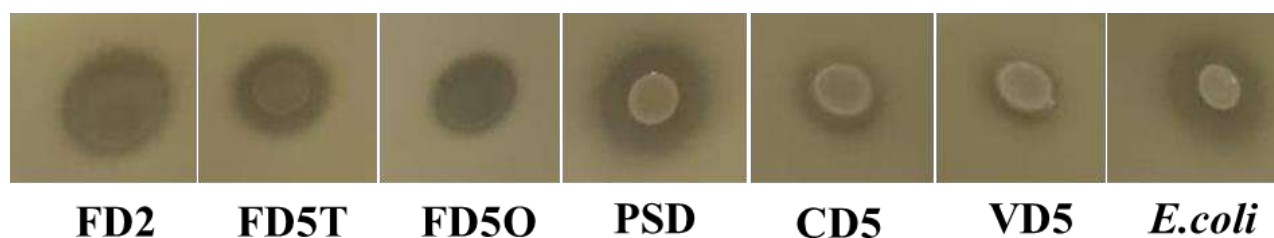


Figure 4. 4 The two-step counterstaining method used to confirm the ability of potential phytase producers (i.e., FD2, FD5T, FD5O, PSD, CD5, and VD5) from different soil samples to degrade phytate and additionally eliminate false positive results. Note: *E. coli* was used as a positive control.

The diameter of the zone of hydrolysis of isolates ranged from 8-14mm, which was close to that of *E. coli* (13mm) (**Table 4. 11**). The zone of hydrolysis of PSD (14mm) was slightly higher than that of *E. coli* (13mm). This suggests that the isolates, including PSD, possessed hydrolytic activity like *E. coli*, indicating their potential ability to break down phytate.

Table 4. 11 Measurement of the zone of hydrolysis produced by potential soil isolates.

Soil isolates	Diameter of zone of hydrolysis (mm)
FD2	11
FD5T	10
FD5O	12
PSD	14
VD5	8
CD5	9
<i>E.coli</i>	13

4.2.3.4 Molecular identification of bacterial isolates:

The PureLink® Genomic DNA kit was utilized to isolate gDNA from pure soil isolates by following the manufacturer's protocol. The isolated gDNA was resolved on a 1% agarose gel to check for its integrity (**Figure 4. 5**). The concentration of gDNA extracted from potential isolates ranged from 9-140 ng μL^{-1} (**Table 4. 12**). The highest concentration of gDNA was extracted from the PSD soil isolate.

Table 4. 12 Concentration and quality estimation of gDNA isolated from different soil isolates and *E. coli*.

Sample	A ₂₆₀ /A ₂₈₀	Concentration of gDNA (ng/ μl)
<i>E. coli</i>	1.9	139.534
FD5T	1.985	19.773
PSD	1.507	140.446
CD5	1.847	85.607
FD5O	1.517	9.053
FD2	1.914	12.349
VD5	1.946	112.757

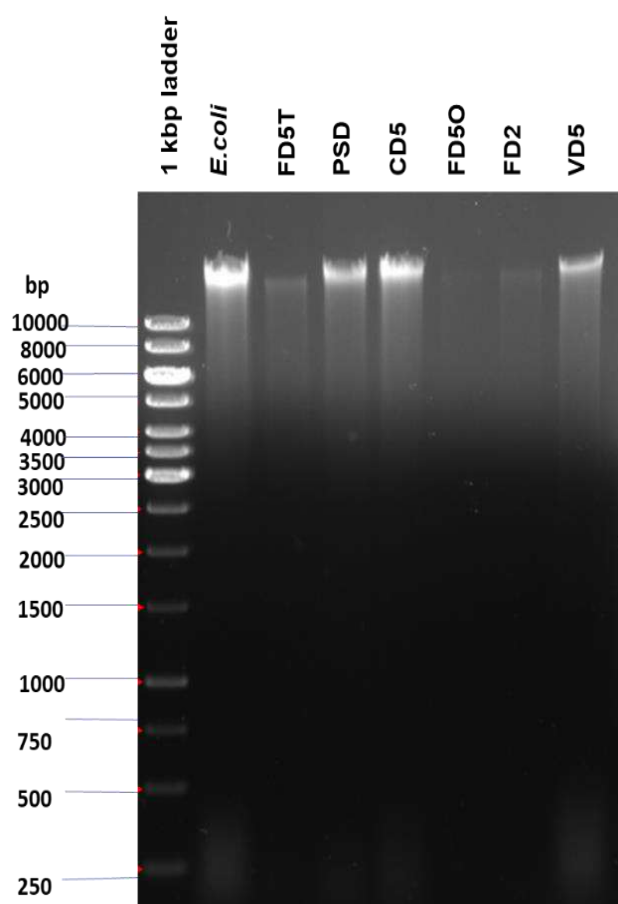


Figure 4. 5 Isolation of gDNA from soil isolates and *E.coli* (used as control) by using the PureLink® Genomic DNA kit.

The isolated gDNA was utilized to amplify the 16S rRNA gene using the 515F and Y36 universal primers. The amplicon of 1kbp was observed on a 1% agarose gel (**Figure 4. 6**) and sequenced via Sanger sequencing. The sequencing data were analyzed via NCBI Blastn to identify the closest homologs of these isolates (**Table 4. 13**). MEGA7 software was used to construct the maximum likelihood (**Figure 4. 7**) and neighbor-joining (**Figure 4. 8**) phylogenetic trees, which represented the relationship between the isolates and their closest homologs. The analysis revealed that isolates VD5 and CD5 were closely related to *Enterobacter cloacae*, FD5T was closely related to *Paenibacillus* sp., whereas FD5O and FD2 were closely related to *Priestia megaterium*, and PSD was closely related to *Klebsiella variicola* (**Table 4. 13**).

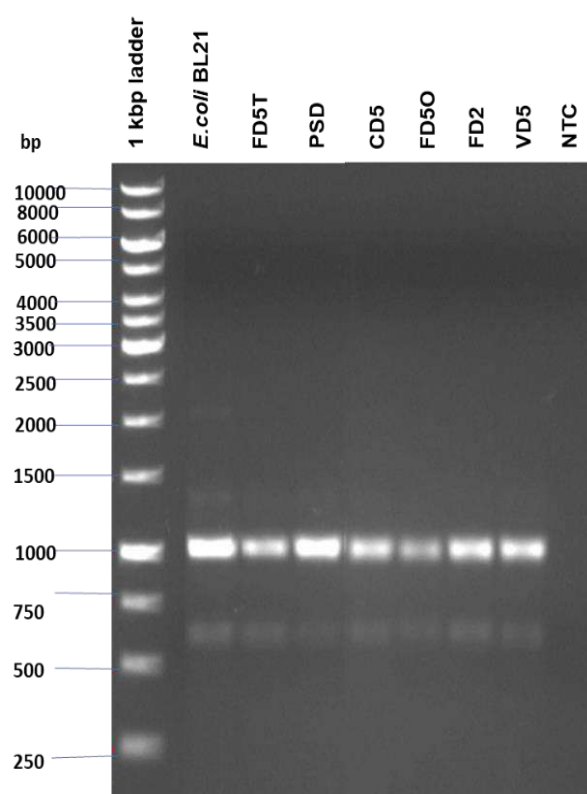


Figure 4. 6 16S rRNA gene amplification of soil isolates (FD2, FD5O, FD5T). NTC: negative control; *E.coli*: positive control.

Table 4. 13 Blastn analysis of 16S rRNA gene of soil isolates

Soil isolates	Close homolog according to Blastn analysis	Percent Identity	Accession
PSD	<i>Klebsiella variicola</i> strain FDAARGOS_628 chromosome	99.40	CP050958.1
VD5	<i>Enterobacter cloacae</i> strain CBG15936 chromosome, complete genome	99.20	CP046116.1
CD5	<i>Enterobacter cloacae</i> strain CBG15936 chromosome, complete genome	99.60	CP046116.1
FD5T	<i>Paenibacillus</i> sp. strain UFPI-57C 16S ribosomal RNA gene, partial sequence	98.37	MK681944.1
FD5O and FD2	<i>Priestia megaterium</i> strain FDU301 chromosome, complete genome	100	CP045272.1

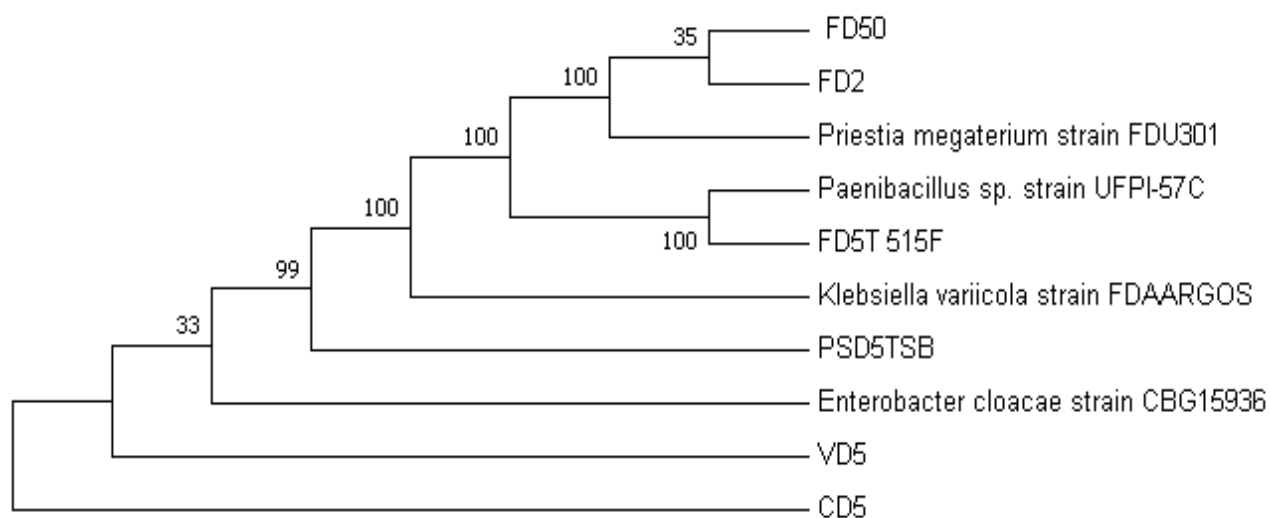


Figure 4. 7 The Maximum Likelihood method (Jukes-Cantor model, 1000 bootstrap) was used to represent the evolutionary history of the soil isolates and closet homologs.

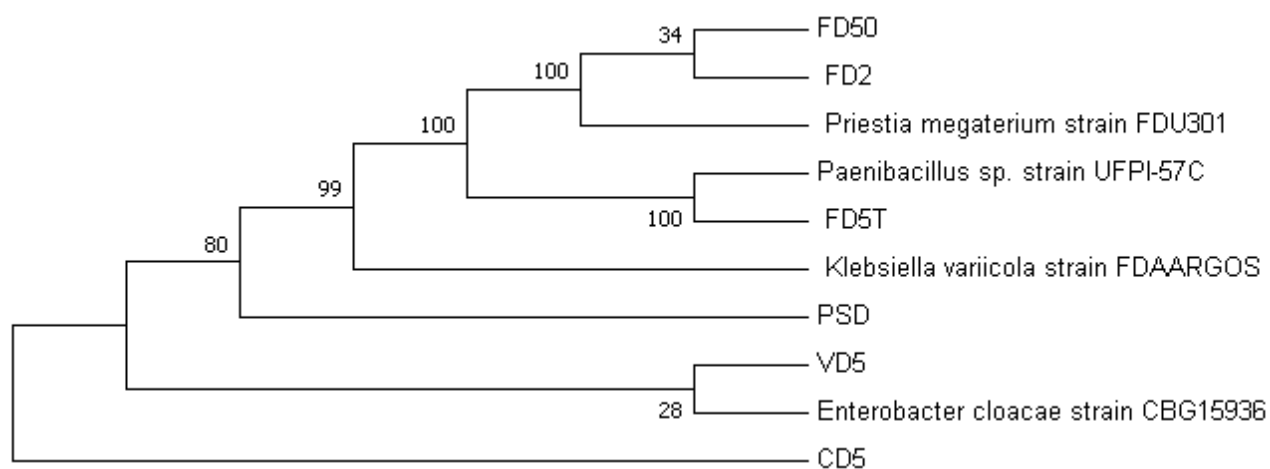


Figure 4. 8 The Neighbor-Joining method (Jukes-Cantor model, 1000 bootstrap) was used to represent the evolutionary history of the soil isolates and closet homologs.

In summary, the PureLink® Genomic DNA kit was used to extract gDNA from soil isolates, and subsequent analysis of the 16S rRNA gene revealed the closest homologs for each isolate. This information was further visualized through a phylogenetic tree, highlighting the relationships between the isolates and their closest known species.

4.2.3.5 Quantitative effect of pH on phytase activity:

The phytase activity of isolates at different pH values was estimated at pH 2.5 and 5. The B-cell lysis buffer and freeze-thaw method were used to prepare cell lysate. Phenylmethylsulfonyl fluoride (PMSF) was added during the lysis process to inhibit protease activity. The prepared lysate was centrifuged to remove debris. The reaction mixture consisted of crude extract and substrate sodium phytate. The mixture was incubated at 37°C for 30 minutes. The released phosphate from the hydrolysis process was measured using the ammonium molybdate method (modified Heinonen method).

Bovine serum albumin (0.2 to 10 $\mu\text{g } \mu\text{L}^{-1}$) was used as a standard to measure total protein (

Figure 4. 9). KH_2PO_4 (0.1 mM to 2 mM) was used as a standard to measure phytase activity (

Figure 4. 10) (**Figure 4. 11**). The phytase activity of soil isolates ranged from 0.012 to 0.043 U/ml (**Figure 4. 12**). The shortlisted soil isolates exhibited optimum activity at pH 5. The findings suggest that these isolates could potentially be used for applications that require phytate degradation, such as in animal feed or soil amendment.

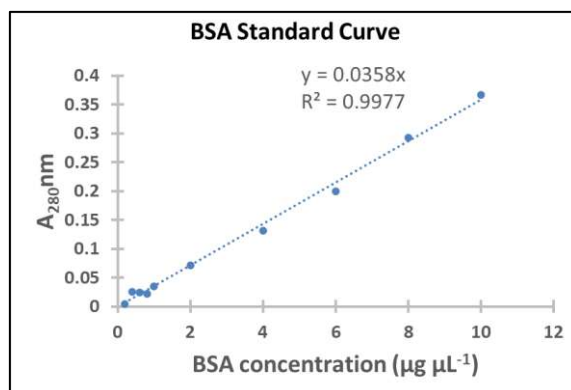


Figure 4. 9 Bovine Serum Albumin (BSA) standard curve.

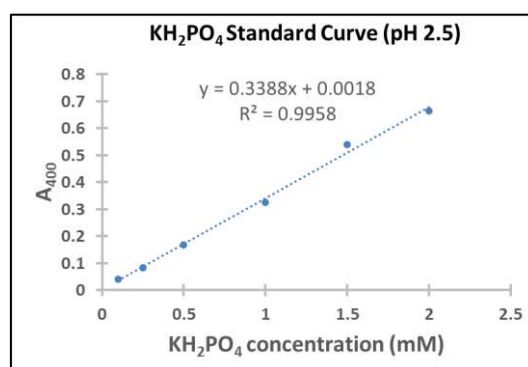


Figure 4. 10 KH_2PO_4 Standard curve at pH 2.5.

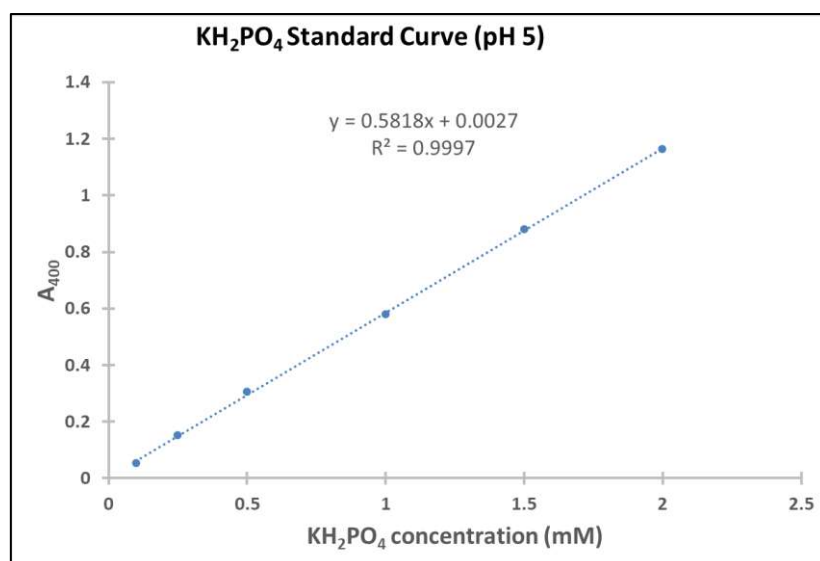


Figure 4.11 KH₂PO₄ Standard curve at pH 5.

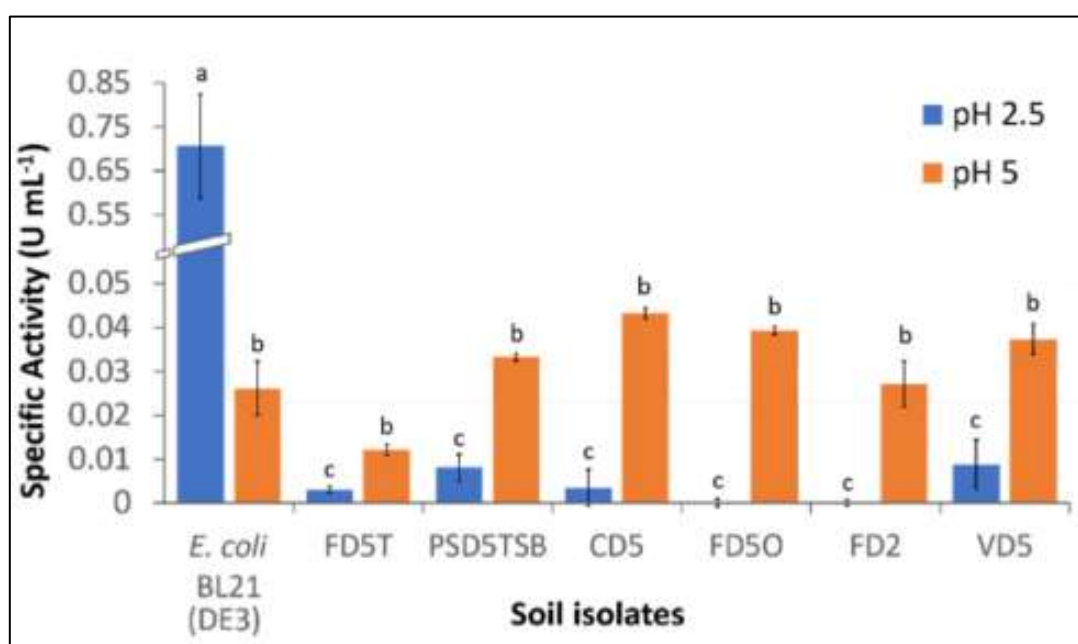


Figure 4.12 The specific activity of phytase was compared among various samples (FD5T, FD2, FD50, PSD5, VD5, CD5) at both pH 2.5 and pH 5, using wildtype *E. coli* BL21 (DE3) phytase as the positive control. The results are presented as mean \pm standard error (U/mL). To determine statistical significance, Student's t-test and two-way ANOVA were applied, setting the significance level at $p \leq 0.05$. The letters are used to represent significant differences in

mean activity between each soil isolate at pH 2.5 and pH 5, as well as differences across the various soil isolates at their respective pH levels.

4.2.3.6 Full-length 16 S rRNA gene sequencing:

PSD was shortlisted based on preliminary screening of phytase-producing microbes from the soil via qualitative (phytase plate assay) and quantitative (phytase activity assay) methods. The full-length 16S rRNA gene of PSD was amplified at different annealing temperatures (**Figure 4. 13**). The amplicons were pooled together and purified (**Figure 4. 13**). This PCR amplicon was sequenced with both forward and reverse primers, which revealed its closest evolutionary relation to *Klebsiella variicola* (**Table 4. 14**).

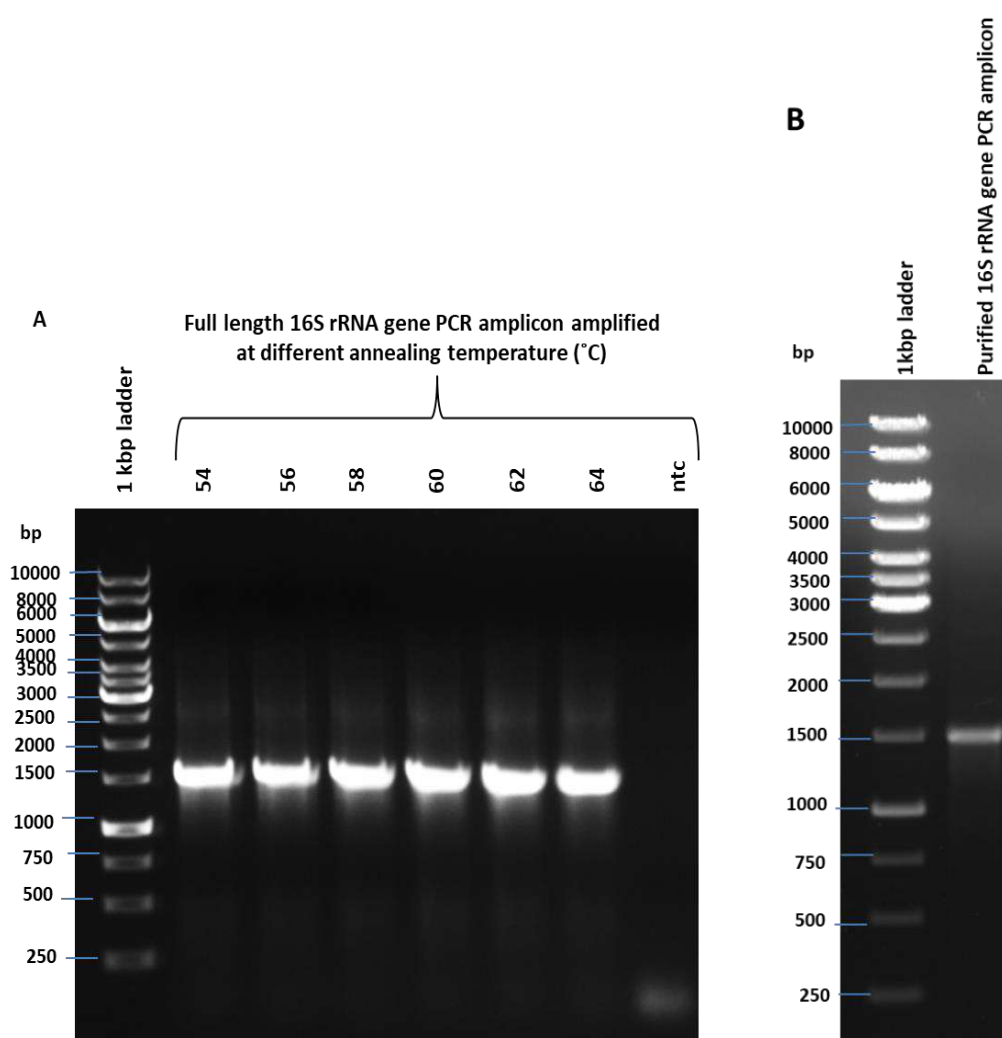


Figure 4. 13 A. Full-length PSD 16S rRNA gene PCR amplicons at different annealing temperatures. B. Purified 16S rRNA gene PCR amplicon used for Sanger’s Sequencing.

Table 4. 14 BLASTn analysis of 16S rRNA full-length gene.

Input for BLASTn	Closet homolog	Query cover	Percent Identity (%)	Accession I.D.
16S rRNA gene sequenced using forward primer	<i>Klebsiella variicola</i> strain FDAARGOS	100%	98.72	<u>CP050958.1</u>
16S rRNA gene sequenced using reverse primer			100	

4.2.3.7 PSD phytase gene amplification:

The phytase sequence of *Klebsiella variicola* PSD was obtained from UniProtKB. Two sets of primers were designed: full-length primers, which targeted the whole phytase gene, and mid-length primers, which targeted the conserved region of the phytase gene. The gel-purified PCR amplicon (approximately 1.4 kbp) was sequenced to confirm the amplified phytase gene sequence (**Figure 4. 14**).

The BLASTX (**Figure 4. 15**) and multiple sequence alignment in serial cloner software (*K.variicola* phytase gene which was used for phytase gene primer sequencing aligned with Sanger's sequenced PSD phytase gene) analysis confirmed the presence of phytase gene in PSD isolate (**Figure 4. 16**). The presence of an N-terminal conserved region (RHGIRPP) in the sequence indicated that it belongs to the Histidine acid phytase group.

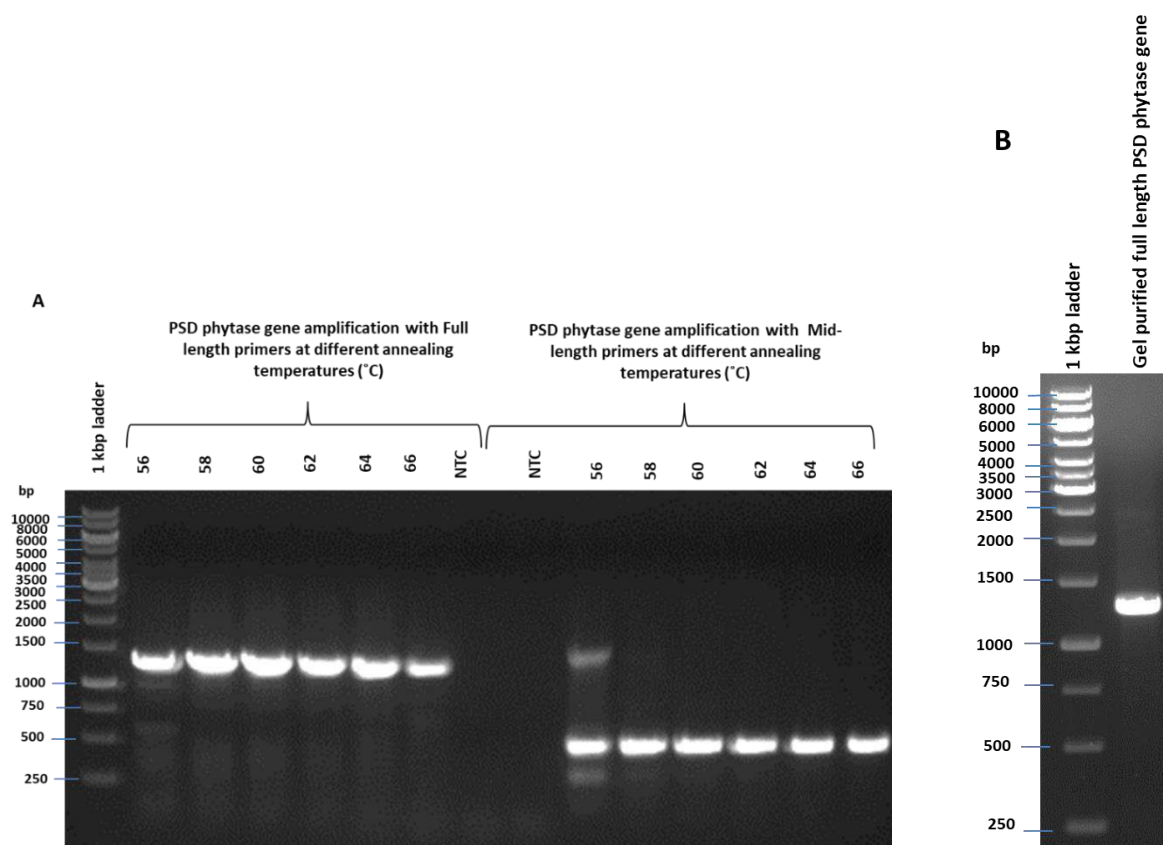


Figure 4. 14 A. PSD full-length and mid-length phytase gene amplification at different annealing temperatures(°C). NTC: non-template control/negative control. B. Gel-purified PSD phytase gene PCR amplicon used for Sanger’s sequencing.

histidine phosphatase family protein [Klebsiella variicola subsp. variicola]
 Sequence ID: [HCB0068922.1](#) Length: 422 Number of Matches: 1
[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 24 to 273 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
472 bits(1215)	2e-164	Compositional matrix adjust.	250/250(100%)	250/250(100%)	0/250(0%)	+3
Query 3	QSAAADWQLEKVVEL	SRHGIRPPT	AGNREAIEAATGRPWEWTTHDGELTGHGYAAVNVN	182		
Sbjct 24	QSAAADWQLEKVVEL	SRHGIRPPT	AGNREAIEAATGRPWEWTTHDGELTGHGYAAVNVN	83		
Query 183	KGREEGQHYRQLGLLQAGCPTAESIYVRASPLQRTRATAQALVDGAFPGCGVAIHVVSVD		KGREEGQHYRQLGLLQAGCPTAESIYVRASPLQRTRATAQALVDGAFPGCGVAIHVVSVD	362		
Sbjct 84	KGREEGQHYRQLGLLQAGCPTAESIYVRASPLQRTRATAQALVDGAFPGCGVAIHVVSVD		KGREEGQHYRQLGLLQAGCPTAESIYVRASPLQRTRATAQALVDGAFPGCGVAIHVVSVD	143		
Query 363	ADPLFQTDKFAATQTDPARQLAAVKEKAGDLAQRRLQALAPAIQLLQAVCQADKPCPIFD		ADPLFQTDKFAATQTDPARQLAAVKEKAGDLAQRRLQALAPAIQLLQAVCQADKPCPIFD	542		
Sbjct 144	ADPLFQTDKFAATQTDPARQLAAVKEKAGDLAQRRLQALAPAIQLLQAVCQADKPCPIFD		ADPLFQTDKFAATQTDPARQLAAVKEKAGDLAQRRLQALAPAIQLLQAVCQADKPCPIFD	203		
Query 543	TPWQVEQSKSGKTTISGLSVMANMVETLRLGWSENLPLSQLAWGKITQARQITALLPLLT		TPWQVEQSKSGKTTISGLSVMANMVETLRLGWSENLPLSQLAWGKITQARQITALLPLLT	722		
Sbjct 204	TPWQVEQSKSGKTTISGLSVMANMVETLRLGWSENLPLSQLAWGKITQARQITALLPLLT		TPWQVEQSKSGKTTISGLSVMANMVETLRLGWSENLPLSQLAWGKITQARQITALLPLLT	263		
Query 723	ENYDLSNDVL	752				
Sbjct 264	ENYDLSNDVL	273				

Figure 4. 15 BLASTX analysis of full-length PSD phytase gene sequence.

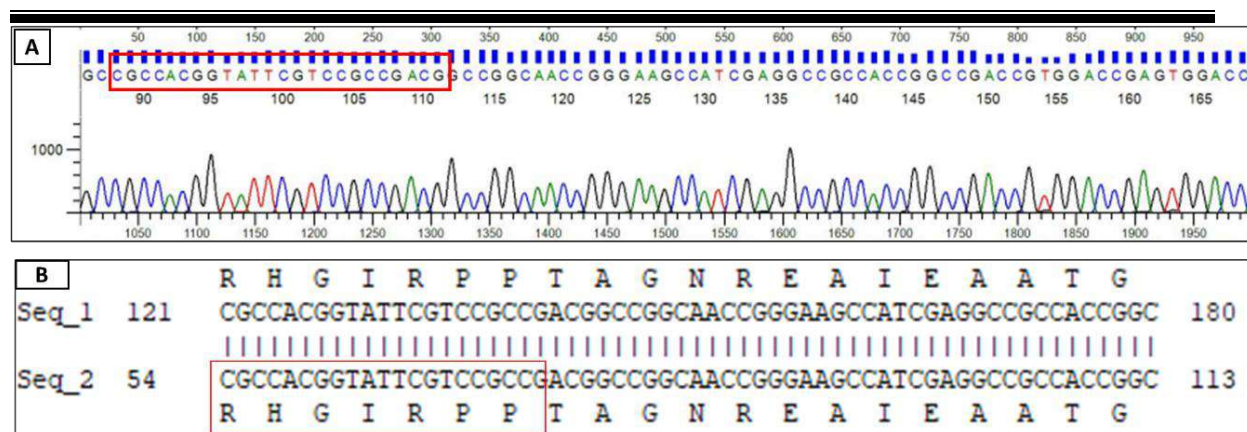


Figure 4. 16 A. Snippet of the PSD phytase gene electrogram: The red highlighted nucleotide sequence codes for the conserved N-terminal ‘RHGIRPP’ protein sequence. The presence of this sequence confirms the presence of the phytase gene in PSD. B. Snippet of the alignment constructed in Serial cloner: Seq_1: Nucleotide sequence of *Klebsiella variicola* which was used for primer designing aligned with the Seq_2: Sanger’s sequenced PSD full-length phytase gene. The PSD full-length phytase gene exhibited the presence of a conserved N-terminal ‘RHGIRPP’ protein sequence (Highlighted in red).

4.3 Metagenomics:

4.3.1 Introduction:

This approach aims to identify bacterial diversity in soil samples via 16S rRNA metagenomics and identify new phytase producers (**Figure 4. 17**).

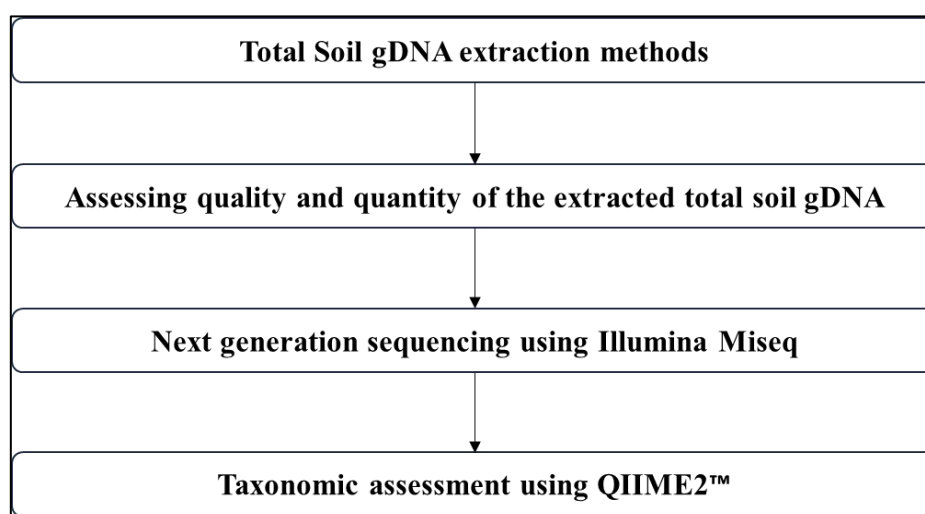


Figure 4. 17 Overview of bioprospecting of potential phytase producer via metagenomics approach.

4.3.2 Materials and Methods:

4.3.2.1 Materials (Chemicals, kits, reagents, solutions, plasticware):

Garden soil (Maharashtra, Mumbai), sterile Falcon (50 and 15ml), motor and pestle, 2mm Sieve, all chemicals were procured from either Sigma-Aldrich, MP Biomedicals, Loba Chemie, MolyChem., Invitrogen PureLink™ Genomic DNA kit (Cat. No. K182001), sterile glass beads (2.5mm size), HiPura soil DNA extraction kit, DNeasy Powersoil kit, gel electrophoresis, Biotek Epc microplate spectrophotometer, MiBiome therapeutic India PVT Ltd Illumina sequencing, HP laptop windows 10 64 bit (intel core i5 72000 CPU @ windows host, Oracle Virtual Box (v. 6.1.22), QIIME2 platform, Hexatech shaker.

4.3.2.2 Collection of soil sample:

The aim was to collect topsoil samples as they were rich in humus, i.e., nutrients that supported the growth of microflora and plant growth. To optimize the protocol of soil DNA extraction, garden soil was selected. The garden soil sample (at a depth of 3cm) was collected in a labeled sterile 50ml Falcon tube with the help of a sterile spatula. Three spots (3m apart) from the same field were selected. All samples were thoroughly mixed and sieved with the help of a 0.22mm mesh. The sieved soil sample, which was free of lumps and litter, was stored at -20°C.

4.3.2.3 Processing of Soil Sample:

The sieved soil sample was divided into two parts. One part was subjected to the liquid nitrogen grinding process, and the other was directly taken for the total soil gDNA extraction protocol. To process the soil sample with liquid nitrogen, the motor and pestle were first cleaned with 100% ethanol. 10g of the soil sample plus liquid nitrogen (added as per needed) was ground together in the cleaned motor and pestle. The grinding process continued until the soil turned into a fine powder. This finely grounded soil sample was labeled as S1. The S1 soil sample was used for all the extraction protocols mentioned below.

The other part of the soil sample (10g) was also ground in the motor and pestle to obtain a finely grounded soil sample (S2). The S2 soil sample was used for the total soil gDNA extraction protocols mentioned below. Please note that in this direct method, liquid nitrogen was not added to the soil sample while processing it.

4.3.2.4 Total soil gDNA extraction protocol:

The overview of the total soil gDNA extraction protocol is represented in **Figure 4. 18**. A total of 8 different extraction protocols were optimized to obtain a high yield and high quality of total soil gDNA from S1 and S2 soil samples.

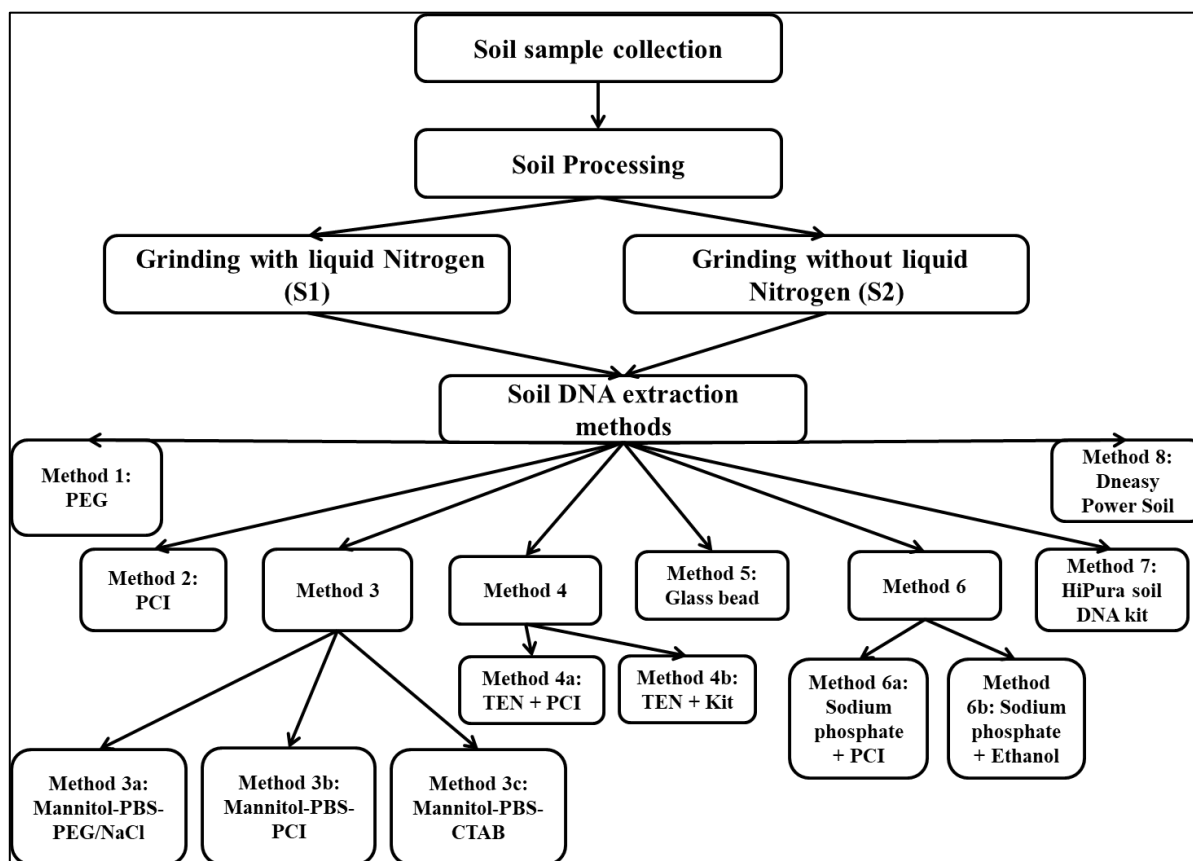


Figure 4. 18 Overview of total soil gDNA extraction protocol.

Method1: PEG method:

In this method, the soil sample (1 gm) was mixed with 10 ml of DNA extraction buffer 1 (120 mM Na_2HPO_4 + 5% SDS (w/v) + 0.1 g PVPP) in a sterile 50 ml falcon tube. The mixture was exposed to heat treatment at 65°C for 1 hour in a water bath. After that, the mixture was centrifuged at 7000 x g for 10 minutes at 4°C. The supernatant was collected and used for the next step. Next, ½ volume of 50% PEG (6000) and 1 volume of 0.6M NaCl were added to the supernatant. This mixture was gently mixed about 3-4 times by slowly inverting the tubes. Then, 1 volume of Chloroform: isoamyl alcohol (24:1) was added to the homogenous mixture. It was centrifuged at 13,000 x g for 10 minutes at 4°C, and the supernatant was collected. To the supernatant, 1/10th volume of 3M sodium acetate (pH 5.2) followed by 2 volumes of 100%

Chapter 4

chilled ethanol were added. The resultant mixture was gently mixed by inverting the tubes about 3-4 times and incubated at -20°C (overnight). After the overnight incubation, the mixture was centrifuged at 13,000 x g for 10 minutes at 4°C. The supernatant was discarded, and the pellet was air-dried for 15-20 minutes followed by resuspension of the pellet in 100 µl of T₁₀E₁ buffer.

Method 2: Phenol:Chloroform: Isoamyl alcohol (PCI) method

In this method, the initial soil sample processing steps, i.e., mixing the soil sample with extraction buffer followed by heat treatment and collection of supernatants, remained the same as in the first method, i.e., the PEG method. The collected supernatant was subjected to an equal volume of Phenol:Chloroform: Isoamyl alcohol (25:24:1). The resultant mixture was gently mixed by inverting the tubes 3-4 times, followed by centrifugation (13,000 x g for 10 minutes at 4 °C). The aqueous layer was collected, followed by the addition of 2 volumes of chilled ethanol and 1/10th volume of 3M sodium acetate (pH 5.2). The mixture was gently mixed and incubated overnight. The overnight mixture was centrifuged (13,000 x g for 10 minutes at 4 °C), and the supernatant was discarded. The pellet was air-dried to remove ethanol and then resuspended with 100µl of T₁₀E₁ buffer.

Method 3:

This method consisted of three sub-methods (3a,3b, and 3c). There are a few common steps that were applied for all three sub-methods:

1 g of soil sample was suspended in 5 mL of 120 mM Phosphate Buffer Saline (PBS, pH 7.4). The mixture was mixed thoroughly on a Hexatech shaker (150 rpm for 10 min at 4 °C) followed by centrifugation at 7,000 x g for 10 min at 4 °C. The supernatant was discarded, and the pellet was washed with PBS. It was again centrifuged at 7,000 x g for 10 min at 4 °C. The pellet was resuspended in 10 mL of DNA extraction buffer 2 (0.1 M Tris-HCl (pH 8.0) + 0.2 M EDTA (pH 8.0) + 10% SDS + 0.2 M Mannitol+ 1 M NaCl + 2% CTAB), followed by incubation at 65 °C for 1 h. This soil suspension was then used for methods 3a, 3b, and 3c.

Method 3a: Mannitol-PBS-PEG/NaCl

The soil suspension was centrifuged at 8,000 x g for 10 minutes at 4 °C, and the supernatant was collected. PEG (50%) in ½ volume, plus 1 volume of 0.6 M NaCl, was added to the supernatant. This mixture was incubated at 4°C overnight and then centrifuged at 13,000 x g for 10 minutes at 4°C. The supernatant was discarded, and the pellet was resuspended in a 3

Chapter 4

ml TE buffer. The resuspended pellet was then processed as per method 2, i.e., the PCI method, right from the addition of the PCI reagent to the resuspension of the pellet in the elution buffer.

Method 3b: Mannitol-PBS-PCI:

In this method, the soil suspension was first centrifuged at 7000 x g for 10 minutes at 4°C. The supernatant was collected and processed as per the protocol mentioned in Method 2, i.e., the PCI method (right from the addition of PCI reagent to the resuspension of the pellet in the elution buffer).

Method 3c: Mannitol-PBS-CTAB:

The soil suspension was centrifuged at 7,000 x g for 10 minutes at 4 °C. The supernatant was collected, and 50 µl of 1 M NaCl plus 50 µl of 10% CTAB (prepared in 0.6 M NaCl) were added to it. This mixture was incubated for 15 minutes at 4°C. The resultant mixture was then processed following the steps given in method 2, i.e., starting from the addition of PCI to the resuspension of the pellet in the elution buffer.

Method 4:

This method consisted of two sub-methods (4a and 4b). There are a few common steps that were applied for these two sub-methods:

0.5 gm of soil was suspended in 1 ml of the TEN buffer (100 mM Tris-HCl+ 50 mM EDTA+ 500 mM NaCl) and vortexed for 1 min. The mixture was centrifuged at 10,000 x g for 10 min at RT. The supernatant was decanted, and the pellet was washed with 1 ml of TEN buffer, followed by centrifugation at 10,000 x g for 10 min at RT. The supernatant was discarded, and the pellet was resuspended in 1 ml of TEN (supplied with 0.2 mg lysozyme). The mixture was incubated for 1 h at 37°C, followed by 10 min on ice and then 20 min at 65°C. After serial incubations at different conditions, 100 µl of 20% SDS was added to the mixture. It was then vortexed for 1 min and incubated for 30 min at RT. The supernatant was collected by centrifugation at 10,000 x g for 10 min at RT. To this supernatant, 500 µl of 3 M sodium acetate was added. This mixture was incubated at 65°C for 5 min, followed by incubation on ice for 20 min. This mixture was centrifuged at a maximum of 14,000 x g for 30 min at 4 °C, and the supernatant was collected. This supernatant was divided into two parts. One part was treated as per the 4a protocol, and the other was treated as per the 4b protocol.

Chapter 4

Method 4a: TEN + PCI

The supernatant was processed as per Method 2, i.e., the PCI method, right from the addition of the PCI reagent until the addition of the elution buffer.

Method 4b: TEN + Kit

In this method, the supernatant was loaded onto the Invitrogen PureLink Genomic DNA mini kit spin column. It was then centrifuged at 16,000 x g for 2 minutes at room temperature (RT). The flowthrough was discarded. Next, 500 µl of wash buffer 1 was added to the column, followed by centrifugation at 16,000 x g for 1 minute at RT. Again, the flowthrough was discarded, and 500 µl of wash buffer 2 was added. The column was centrifuged at 16,000 x g for 1 minute at RT. The flowthrough was discarded, and a dry spin was given by centrifugation at 18,407 x g for 3 minutes at RT. Subsequently, the spin column was placed in a fresh Eppendorf tube for the elution step. 50 µl of elution buffer was added to the spin column and centrifuged at 18,000 x g for 2 minutes at RT. The extracted gDNA was then stored at -20 °C.

Method 5: Glass bead method

1 mg of soil sample was suspended in 400 µl of 50 mM Tris-HCl and 1 mM EDTA. To the mixture, sterile glass beads (2.5 mm size) were added and vortexed for 1 min. The sterile beads were then separated from the mixture, and 50 µl of lysozyme (10 mg/ml) was added to the homogenized soil solution. It was then incubated for 1.5 hours at 37 °C. 250 µl of Guanidine Hydrochloride (4 M) was then added to the mixture and mixed gently for 45 seconds. After mixing, 300 µl of Sodium Lauryl Sarcosine was added and vortexed for 10 min at 37 °C. This was then incubated for 1 hour at 70 °C. After incubation, 300 mg of 0.1 mm zirconia beads were added to the mixture and vortexed for 20 min. To this mixture, 15 mg of PVPP was added and gently vortexed, followed by centrifugation at 14,000 x g at 4 °C for 5 min. The supernatant was collected in a fresh tube, and the pellet was washed with 200 µl of TENP (50 mM Tris-HCl+ 20 mM EDTA+ 100 mM NaCl+ 1% PVPP), followed by centrifugation at 14,000 x g at 4 °C for 5 min. Again, the supernatant was collected and mixed with the earlier collected supernatant. To this supernatant, 2 volumes of 100% ethanol were added and mixed gently by inverting tubes. This was then incubated at room temperature for 5 min, then centrifuged (at 14,000 x g at 4 °C for 5 min). The supernatant was removed, and the pellet was air-dried for 15 min. The pellet was then resuspended with 450 µl of PB supplemented with 50 µl of 3 M sodium acetate and further incubated at 4 °C for 1 hour. To this mixture, 2 µl each of RNase

Chapter 4

and Proteinase K were added and incubated for 30 min at 37 °C. After incubation, 50 µl of sodium acetate (3 M) and 1 ml of ethanol were added. This was then centrifuged at 14,000 x g for 10 min at 4 °C, and the supernatant was decanted while the pellet was washed with 70% ice-cold ethanol. The pellet was air-dried and resuspended in 100 µl of TE buffer.

Method 6:

This method consisted of two steps: 6a and 6b. These steps are common for both sub-methods: First, 1mg of soil sample was suspended in 100 mM sodium phosphate buffer (pH 8) along with 1mg of lysozyme solution. The mixture was incubated at 37 °C for 2 hours, followed by the addition of 1 ml of SDS solution. The mixture was thoroughly mixed and then incubated at room temperature for 5 minutes and later on ice for 2 minutes. It was then incubated at 65 °C for 20 minutes and quickly incubated on ice for 2 minutes.

To this mixture, 1 ml of 3M sodium acetate was added and vortexed vigorously. The mixture was then centrifuged at 6,000 x g for 10 minutes, and the supernatant was collected. This supernatant was divided into two parts and processed as per methods 6a and 6b.

Method 6a: Sodium phosphate + PCI

The collected supernatant was processed as per method 2, i.e., the PCI method (right from PCI addition to the elution).

Method 6b: Sodium phosphate + Ethanol

To the collected supernatant, an equal volume of ethanol was added and centrifuged at 14,000 x g for 10 minutes at 4°C. The supernatant was decanted, and the pellet was air-dried, followed by resuspension in 100 µl of TE buffer.

Method 7: HiPura soil DNA kit

The HiPura soil DNA kit was utilized to extract total soil gDNA from garden soil. We followed the manufacturer's protocol to yield high-quality total soil gDNA. Here, the protocol is explained briefly: This method consists of three major steps: cell lysis, removal of inhibitors, and purification plus elution.

a. Cell lysis step:

A 250mg soil sample was used for the extraction procedure. The lysis of microbial cells present in the soil sample was conducted by combining bead beating and soil lysis solution (SL), along

Chapter 4

with vigorous vortexing for 10 minutes. The mixture was then centrifuged at 13,000 x g for 1 minute, and the supernatant was collected.

b. Inhibitors removal:

Adding the Inhibitor Removal Solution (IRSH) to the collected supernatant, we subjected it to centrifugation at 10,000 x g for 1 minute. The collected supernatant was subjected to the elution process.

c. Purification and elution:

The supernatant was passed through the spin column and purified using binding and wash buffers. The total soil gDNA was eluted by using an elution buffer

Method 8: Dneasy Power Soil

250mg of soil was processed using the DNeasy Powersoil kit by using the manufacturer's protocol. Cell lysis, inhibitor removal, purification, and elution steps were performed using respective buffers to obtain total soil gDNA.

a. Cell lysis buffer:

Cell lysis was performed by using both the mechanical (using bead beating method) and chemical lysis (cell lysis buffer) process. The supernatant was collected by centrifuging the mixture 10,000 x g for 30 seconds.

b. Inhibitor removal:

The inhibitor removal solution was added to the above supernatant. This helped to remove inhibitors present in the sample. *Note: The solution should be transparent.*

c. Purification and elution:

The above mixture was then subjected to purification using the spin column method, along with buffers such as binding and washing buffers. Finally, the total soil gDNA was extracted using an elution buffer.

Storage conditions:

The extracted total soil gDNA from all methods was stored at -20 °C until further use.

4.3.2.5 Assessing quality and quantity of the extracted total soil gDNA:

The integrity of the total soil gDNA extracted from different methods was checked by resolving it on a 0.8% Agarose gel (Tank buffer: 1X TAE). A mixture of 5 µl of the extracted DNA + 2 µl of 6x gel loading dye was loaded on the gel. The ladders used were a 1 kbp ladder and lambda DNA HindIII digest. The gel image was visualized using the Gel Documentation system and Image Lab software.

The BioTek Epoch Microplate spectrophotometer was used to estimate the quantity and quality of the total soil gDNA. 2 µl of each sample and their respective elution buffers were loaded on the Take-3 plate (Note: clean the Take3 plate with distilled water before use). The results were analyzed using Gen 5 Software. The quantity was measured in ng/µl, and the quality was assessed using A260/A280 and A260/A230 ratios.

4.3.2.6 Next-generation sequencing using Illumina Miseq:

The sequencing procedure was performed at MiBiome Therapeutic India Pvt Ltd, Mumbai. The PSD soil sample (isolated in the 4.2.2.5 section) was subjected to NGS sequencing using Illumina Miseq (Illumina, San Diego, CA). The PCR amplification of bacterial 16S rRNA hypervariable region V3-V4 was carried out using primers Forward(5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWG CAG) and Reverse (5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAAT CC). Briefly, 12.5ng DNA was subjected to 16S V3-V4 PCR using the above primers, and the PCR products bead purified and subjected to another round of PCR with dual indices and adapters to generate the libraries. The cleaned libraries were quantitated on a Qubit fluorometer. The library was diluted to 4 nM, spiked with 20% PhiX premade library from Illumina, and loaded on a MiSeq v3 kit. Sequencing was performed for 2X300 cycles.

4.3.2.7 Hardware and software requirements:

The sequencing raw data was obtained, and all the sequence analysis studies described herein were performed at SDSOS, NMIMS University, Mumbai on an HP laptop (Intel® Core™ i5-7200U CPU @ 2.50 GHz, RAM 16 GB) running Windows 10 64-bit Home Basic Operating System. To run QIIME2 on a Windows host, Oracle VM VirtualBox (Version 6.1.22) was installed followed by importing and installing QIIME2™ (Version 2021.2).

4.3.2.8 Taxonomic assessment using QIIME2™:

The demultiplexed paired-end V3-V4 sequenced data was obtained in Casava One Eight Single Lane Per Sample format.

- This data was imported into the QIIME2™ pipeline using the following command line:

```
qiime tools import \  
  --type 'SampleData[PairedEndSequencesWithQuality]' \  
  --input-path file_path \  
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \  
  --output-path demux-paired-end.qza  
  
qiime demux summarize \  
  --i-data demux-paired-end.qza \  
  --o-visualization demux.qzv
```

The (.qza) files are QIIME2™ artifacts, which can be visualised after converting them into a (.qzv) file.

- The forward and reverse reads were joined using “vsearch” using the following command line:

```
qiime vsearch join-pairs \  
  --i-demultiplexed-seqs demux.qza \  
  --o-joined-sequences demux-joined.qza  
  
qiime demux summarize \  
  --i-data demux-joined.qza \  
  --o-visualization demux-joined.qzv
```

- The quality control of sequences was carried out to discard low quality reads using the following command line:

```
qiime quality-filter q-score-joined \  
  --i-demux demux-joined.qza \  
  --o-filtered-sequences demux-joined-filtered.qza \  


---


```

```
--o-filter-stats demux-joined-filter-stats.qza
```

- The next denoising step was carried out using “Deblur” and the sequences were trimmed resulting in 300 nucleotide sequences since best quality reads (Q Phred score>30) were observed up to 300bp using the following command line:

```
qiime deblur denoise-16S \  
--i-demultiplexed-seqs demux-joined-filtered.qza \  
--p-trim-length 300 \  
--p-sample-stats \  
--o-representative-sequences rep-seqs.qza \  
--o-table table.qza \  
--o-stats deblur-stats.qza  
  
qiime feature-table summarize \  
--i-table table.qza \  
--o-visualization table.qzv  
  
--m-sample-metadata-file NGSmetadata-Sheet1.csv  
  
qiime feature-table tabulate-seqs \  
--i-data rep-seqs.qza \  
--o-visualization rep-seqs.qzv
```

- The resulting outputs i.e., table.qza (Feature table (frequency)) and rep-seqs.qza (Feature data sequences) are important outputs, which were used for further down processing. The result of both of these methods will be a FeatureTable[Frequency] QIIME2™ artifact, which contains counts (frequencies) of each unique sequence in each sample in the dataset, and a FeatureData[Sequence] QIIME2™ artifact that maps feature identifiers in the FeatureTable to the sequences they represent.

- The taxonomy assessment of the sequences was carried out using the Naive Bayes classifier trained against the GreenGenes (13-8-99 version) database using the following command line:

```
qiime feature-classifier classify-sklearn \  
  --i-classifier gg-13-8-99-nb-classifier.qza \  
  --i-reads rep-seqs.qza \  
  --o-classification taxonomy.qza  
  
qiime metadata tabulate \  
  --m-input-file taxonomy.qza \  
  --o-visualization taxonomy.qzv
```

- The taxonomic composition of samples with interactive bar plots was viewed using following commands:

```
qiime taxa barplot  
  
  --i-table table-deblur.qza  
  
  --i-taxonomy taxonomy.qza  
  
  --m-metadata-file Final-metadata-NGS.tsv  
  
  --o-visualization taxa-bar-plots.qzv
```

- The data was exported, and family-level and genus-level taxonomic plots were constructed to visualize the sample microbial distribution at the respective levels.

4.3.3 Results:

4.3.3.1 Soil DNA extraction:

Different conventional as well as extraction kits were used to obtain a high yield of total soil DNA. A total of 8 different methods were optimized. The integrity of total soil DNA extracted from different methods was assessed by resolving it on a 0.8% agarose gel (**Figure 4. 19**). The purity and yield of DNA were analyzed using a BioTek Epoch Microplate spectrophotometer. The total soil DNA extracted from different methods ranged from 14.8 to 9217.8 ng/μl (**Table**

Chapter 4

4. 15). The purity was measured at A_{260}/A_{280} and A_{260}/A_{230} , with A_{260}/A_{280} ranging from 1.4 to 3, and A_{260}/A_{230} ranging from 0.023 to 1.789.

The soil DNA kit-based methods, i.e., HiPura soil DNA kit and Dneasy Power soil DNA kit, helped to obtain the best quality of total soil DNA (A_{260}/A_{280} : 1.8) compared to the conventional method (Kamble Asmita & Singh Harinder, 2020) (A. Kamble et al., 2020).

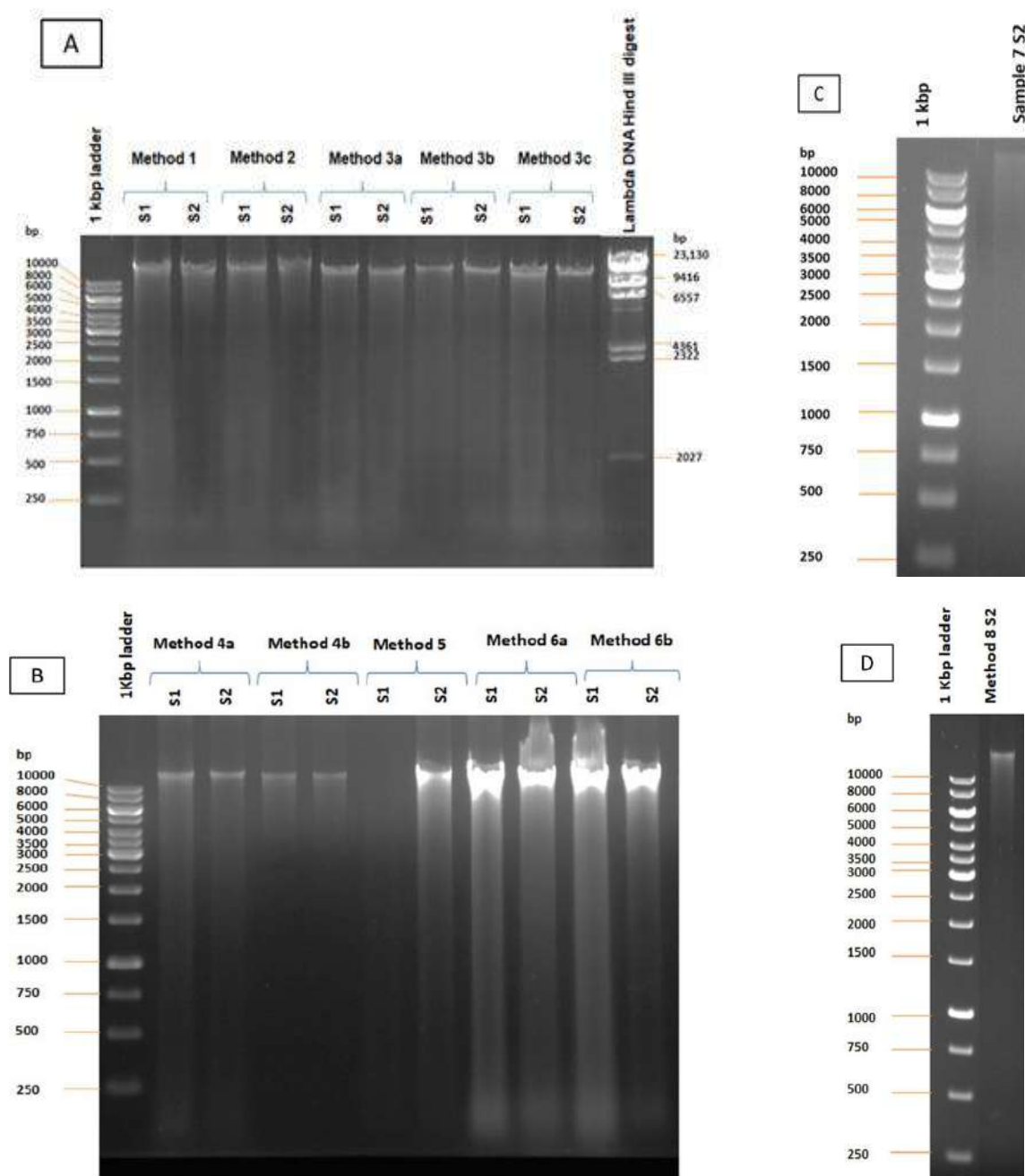


Figure 4. 19 The total soil gDNA extracted from: A. Method 1 to Method 3c; B. Method 4a to Method 6b. C. Method 7; D. Method 8.

Chapter 4

Table 4. 15 Total soil gDNA quantification by using BioTek Epoch Microplate spectrophotometer

Sample	ng/ μ l	A ₂₆₀ /A ₂₈₀	A ₂₆₀ /A ₂₃₀
Method 1 S1	9217.8	1.438	0.534
Method 1 S2	6380.6	1.487	0.482
Method 2 S1	4277.8	1.478	0.577
Method 2 S2	6003.4	1.46	0.616
Method 3a S1	3704.6	1.534	1.174
Method 3a S2	5767.7	1.458	1.416
Method 3b S1	6576.6	1.516	0.711
Method 3b S2	2814	1.515	0.627
Method 3c S1	4298	1.472	0.626
Method 3c S2	4601.7	1.507	0.677
Method 4a S1	856.3	1.19	0.689
Method 4a S2	681.2	1.442	0.745
Method 4b S1	34.1	2.33	0.048
Method 4b S2	14.8	3	0.023
Method 5a S1	104	2.33	0.184
Method 5a S2	563.8	1.468	0.649
Method 6a S1	616.9	1.732	0.691
Method 6a S2	1825.8	1.604	1.433
Method 6b S1	1567.5	1.699	1.427
Method 6b S2	453.1	1.561	1.391
Method 7 S2	11.045	1.851	1.052
Method 8 S2	19.608	1.87	1.789

4.3.3.2 16S rRNA gene metagenomics:

Pig sty soil was used for metagenomics studies. The total soil DNA was extracted using the DNeasy Powersoil kit. The total pig sty soil gDNA was then subjected to 16S rRNA metagenomic sequencing. The QIIME2 platform was used to analyze metagenomics data. The analysis revealed family-level distribution and genus-level distribution (**Figure 4. 20**). Among the family-level distributions, the *Xanthomonadaceae* family seems to be an interesting group, based on literature studies, which can be further explored for novel phytases.

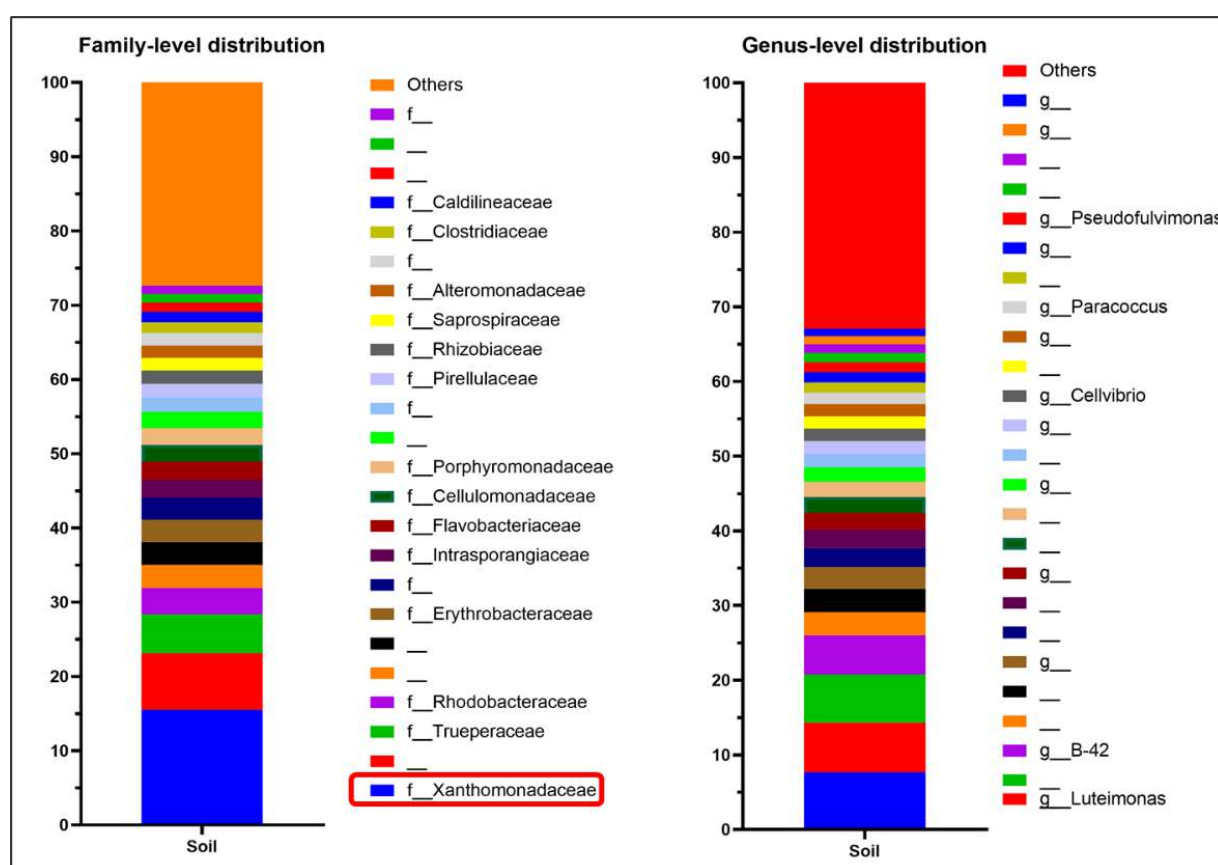


Figure 4. 20 Stacked bar graph depicting the Family-level distribution and Genus-level distribution of bacterial diversity present in pig sty soil sample. The *Xanthomonadaceae* family can be further explored as a potential phytase producer.

4.4 *In-silico* Bioprospecting:

4.4.1 Introduction:

The *in-silico* bioprospecting approach involved three major steps: exploring the database, removing redundancy, and conducting phylogenetic analysis (**Figure 4. 21**). Databases were explored based on certain criteria listed in the 'Materials and Methods section.

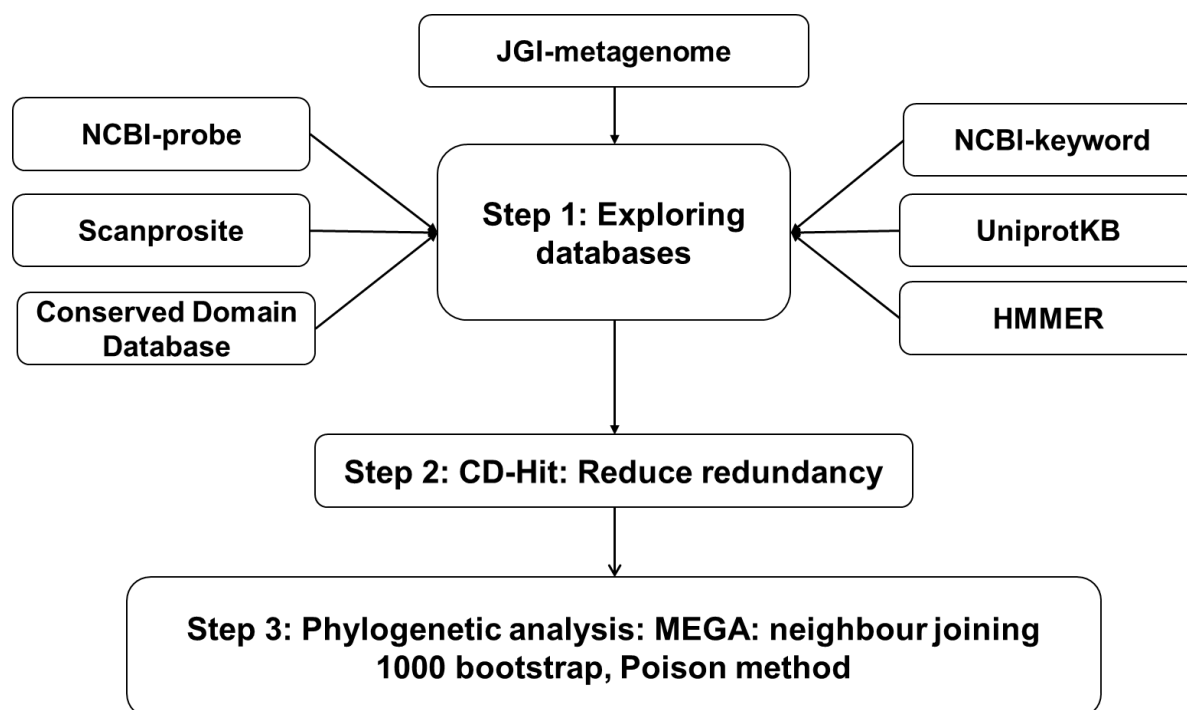


Figure 4. 21 Overview of *in-silico* bioprospecting methodology

4.4.2 Materials and Methods:

Dell Laptop Windows 10 64-bit (Intel core i3), MEGA software v.7, Clustal Omega, Databases: NCBI, ScanProsite, UniProtKB, HHMER, JGI-metagenome.

4.4.2.1 Common steps for all *In-silico* bioprospecting methods

1. The well-characterized phytases were used as probe/ reference sequences. These were the probes/ reference sequences: *Buttiauxella* sp. GC21 (ABX80238.1), *Escherichia coli* (AMH85921.1), *Klebsiella pneumoniae* (BBE68239.1), *Citrobacter braakii* (AAS45884.1), *Hafnia alvei* (EHM46937.1), and *Yersinia pestis* (ABU98780.1). These sequences were excluded from the list which was obtained from each database.
2. The sequences that exhibited 50-80% identical to the reference sequences were selected for further processing.

3. Removal of redundancy by using CD-Hit: All sequences that fulfilled the criteria (mentioned in each method below) were combined in one single file and subjected to CD-Hit (default parameters). This resulted in the list of sequences that were then used for constructing phylogenetic trees.
4. Neighbor-joining phylogenetic tree construction: The list of sequences obtained from CD-Hit, along with the probes/reference sequences, was used as the input for MEGA 7.0 software. The MUSCLE tool, which is embedded in MEGA 7.0 software, was used to form a multiple sequence alignment (MSA) of all CD-Hit sequences, along with the references/probe phytase sequences. The MSA was used to construct a neighbor-joining phylogenetic tree by using MEGA 7.0 software (1000 bootstrap, Poisson method).

4.4.2.2 NCBI-probe method

1. Probe selection: The well-characterized phytase sequences of experimentally characterized phytases were used as probes (details mentioned in 4.4.2.1).
2. BLASTp search: Each probe sequence was used as input for a BLASTp search. We selected the database 'non-redundant protein sequences' (nr) and excluded probes from the organism list. All other parameters were kept as default. From the displayed list, candidates that were 50-80% identical to each probe were selected.
3. CD- Hit and phylogenetic tree construction as mentioned in '4.4.2.1'.

4.4.2.3 NCBI-keyword method

1. Keyword search: the NCBI protein database was explored based on the keyword 'Histidine acid phytase and *Enterobacterales*'. The phytases belonging to the reference genera were excluded from the list. This list was used as input for CD-Hit.
2. CD- Hit and phylogenetic tree construction as mentioned in '4.4.2.1'.

4.4.2.4 ScanProsite method

1. The combination of motifs was added: PS00616 (LerVviVsRHGvRaP) and PS00778 (LIFiAGHDTNLanLsgA).
2. Selection of database: UniProtKB and bacterial taxonomy are selected in the 'filters' section.
3. Output selection: Only complete sequences were retrieved. These sequences were used as input for CD-Hit

4. CD- Hit and phylogenetic tree construction as mentioned in '4.4.2.1'.

4.4.2.5 UniProtKB method

1. The filters were added in the 'Advanced search option' of the UniProtKB database:
Enzyme classification: 3.1.3.8
Taxonomy: 9606 (Human) and 4751 (Fungi) were excluded by selecting 'NOT' from the drop-down list.
Protein name: 3-phytase
The same search criteria were used for 3.1.3.26 i.e., 4-phytase.
2. The list of sequences was further filtered out through the UniProtKB option panel based on length (400-600bp).
3. 3-phytase and 4-phytase sequences were combined and used as input for Clustal Omega.
4. Multiple sequence alignment by using Clustal Omega: The sequences that exhibited the conserved regions, RHGXRXP and HDTN, were selected and used as input.
5. The above list of sequences was used as input for CD-Hit.
6. CD- Hit and phylogenetic tree construction as mentioned in '4.4.2.1'.

4.4.2.6 Conserved domain database (CDD)

1. The CDD was explored based on the keyword: Histidine acid phytase.
2. Two options were explored: PRK10172 and His_Phos_2.
3. Each option had two sub-options: representative and specific protein sequences.
4. We clicked on 'representative' and 'specific protein' (one at a time) which directed us to the NCBI protein database. We selected 'Bacteria' from the 'species' list.
5. Downloaded all sequences which fulfilled the above criteria and were used as input for CD-Hit.
6. CD- Hit and phylogenetic tree construction as mentioned in '4.4.2.1'.

4.4.2.7 HMMER

1. The probes (mentioned in section 4.4.2.1) were used as input sequences.
2. These sequences were used to explore the database: ne_bac70_15. The following parameters were set: MSA enrichment iterations using Hhblits: 1, Evaluate cutoff for reporting: 1e-3, Max target hits: 250, Maximal Sequence Identity (%): 90, Min seq identity of MSA hits with query (%): 30, Minimal coverage with the query (%): 90.

3. The list of sequences that fulfilled the above criteria was used as input for CD-Hit.
4. CD- Hit and phylogenetic tree construction as mentioned in '4.4.2.1'.

4.4.2.8 JGI-metagenome

1. In the JGI-metagenome database: first, the function tab was selected. In the function tab, the keyword: pfam00328 was searched.
2. In addition to the above criteria, the filters used were: Pfam, sequencing status: Finished, Domain: Metagenome, and selected metagenome: 257.
3. Function selected: pfam00328: His_Phos_2- Histidine phosphatase superfamily (branch2).
4. 563 genes were present in pfam00328. It was clicked.
5. This list was filtered by using the Gene-product name: 3-phytase/4-phytase.
6. All sequences were downloaded. This was used as input for CD-Hit.
7. CD- Hit and phylogenetic tree construction as mentioned in '4.4.2.1'.

4.4.2.9 Final list of potential candidates:

1. All shortlisted candidates from all *In-silico* bioprospecting methods were listed down.
2. The candidates who appeared in more than 1 method were selected.
3. The genera of each candidate were searched on the UniprotKB database by using the keyword 'genera name (depends upon the genera of interest) phytase' e.g., *Pantoea* phytase.
4. This list was refined based on the length of the amino acid sequence: 400-600 and the presence of conserved residue 'RHGXRX' and 'HD'.
5. The presence of conserved residues was also confirmed by multiple sequence alignment by using the MUSCLE alignment tool embedded in the MEGA 7.0 software.
6. This MSA was used for phylogenetic construction as mentioned in 4.4.2.1.

4.4.3 Results:

4.4.3.1 NCBI-probe method:

The experimentally characterized phytases from *Buttiauxella* sp. GC21, *Escherichia coli*, *Klebsiella pneumoniae*, *Citrobacter braakii*, *Hafnia alvei*, and *Yersinia pestis* were retrieved from the NCBI database and used as probes, i.e., input for Blastp search. The sequences that were 50-80% identical to the probes were shortlisted. The list of novel candidates was subjected to CD-Hit to remove redundancy, and a phylogenetic tree was constructed along with the

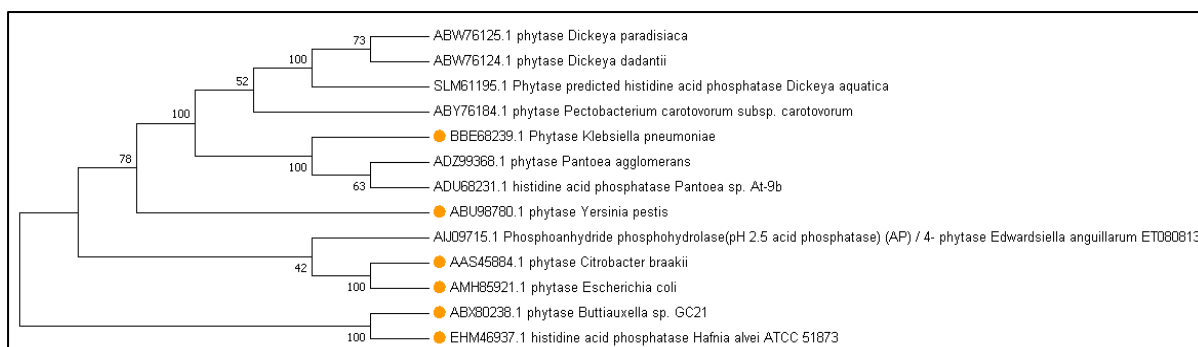


Figure 4. 23 The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from the NCBI database based on the input ‘Keyword’. The colored circle represents the well-characterized phytases.

4.4.3.3 Scanprosite method:

The following motifs were used to scan this database: LerVviVsRHGvRaP (PS00616) and LIFiAGHDTNLanLsgA (PS00778). In addition to this, 'bacteria' was selected in the 'Taxonomy' section. The list of candidates was subjected to CD-Hit. From the list, those candidates that belonged to phytases were selected, and a phylogenetic tree was constructed (Figure 4. 24). The list of potential phytase producers at the genus level, obtained using the Scanprosite method, included *Xanthomonadaceae*, *Sphingomonas*, *Caulobacter*, *Methylobacterium*, *Candidatus*, *Serratia*, *Cronobacter*, and *Kosakonia*.

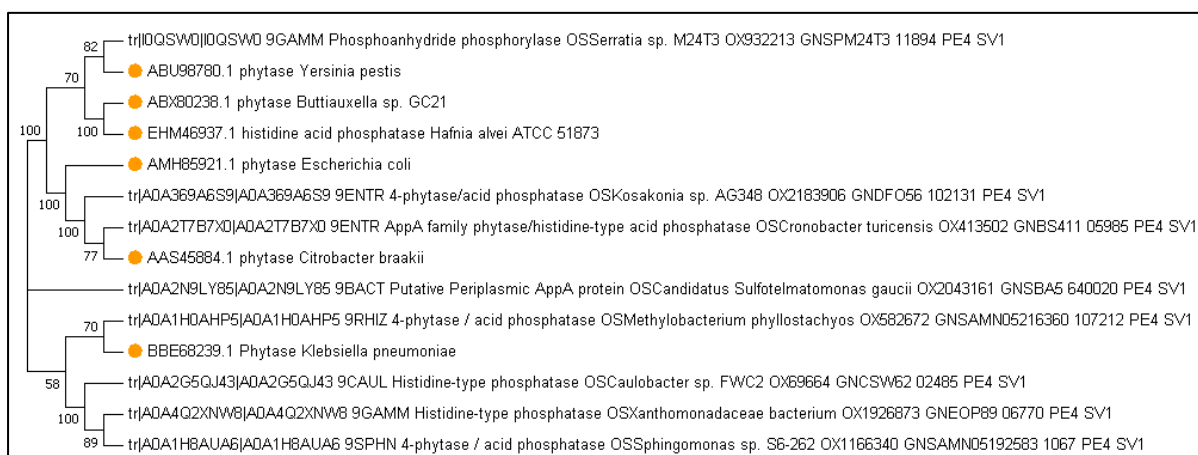


Figure 4. 24 The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from

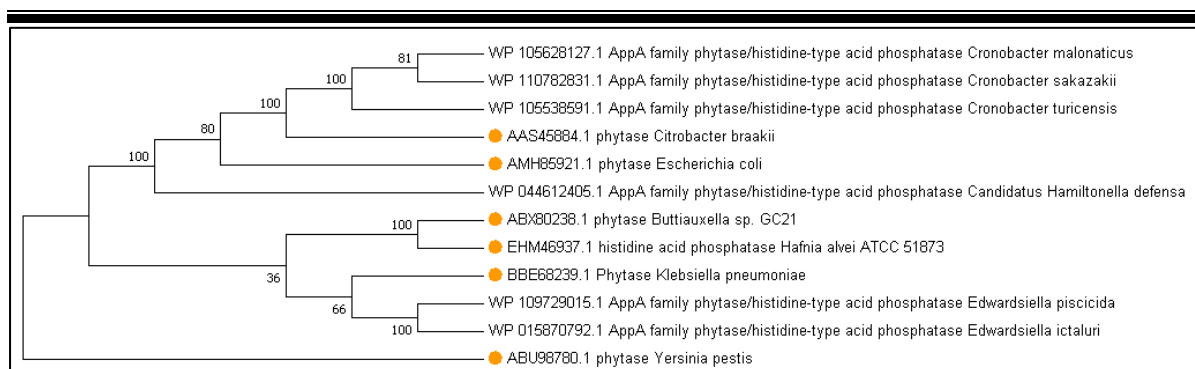


Figure 4. 26 The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers from the Conserved domain database, based on the input ‘Keyword and conserved domains’. The colored circle represents the well-characterized phytases.

4.4.3.6 HMMER:

The sequences that fulfilled the criteria mentioned in HMMER were then subjected to CD-Hit, and a phylogenetic tree was constructed (**Figure 4. 27**). The list of genera obtained via the HMMER method includes *Sphingomonas*, *Novosphingobium*, *Phenylobacterium*, *Luteimonas*, *Brevundimonas*, *Altericroceibacterium*, *Nitrospirillum*, *Caulobacter*, *Rhizomicrobium*, *Granulicella*, *Acidobacterium*, *Verrucomicrobia*, *Sulfotelmatomonas*, *Acidipila*, *Dyella*, *Opitutus*, *Terriglobus*, *Paludibaculum*, *Silvibacterium*, *Terracidiphilus*, *Bryocella*, *Edaphobacter*, *Bilophila*, *Desulfovibrio*, *Azospirillum*, *Microvirga*, *Paraburkholderia*, *Trinickia*, *Pandoraea*, *Aeromonas*, *Deefgea*, *Cronobacter*, *Kosakonia*, *Enterobacter*, *Edwardsiella*, *Plesiomonas*, *Rouxiella*, *Budvicia*, and *Pantoea*.

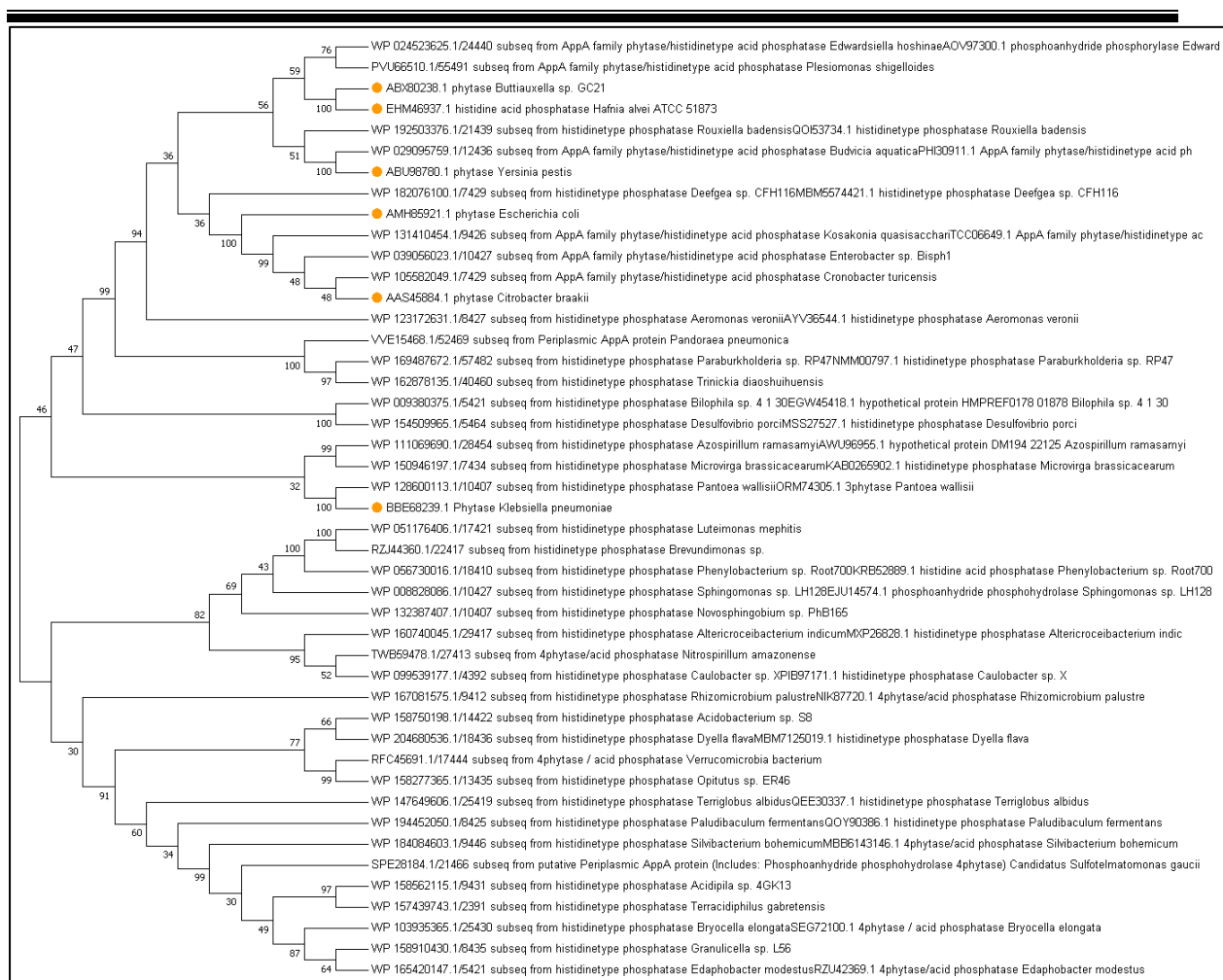


Figure 4. 27 The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers via the HMMER method, based on the input ‘Keyword and conserved domains’. The colored circle represents the well-characterized phytases.

4.4.3.7 JGI-metagenome:

This database was explored using a keyword search: pfam00328. The list was further refined by applying filters such as gene product names (4-phytase and 3-phytase), domain: metagenome, and sequence status: finished. The list of sequences was subjected to CD-Hit, and a phylogenetic tree was constructed (**Figure 4. 28**). The genera obtained by this method were *Erwinia* and *Pseudomonas*.

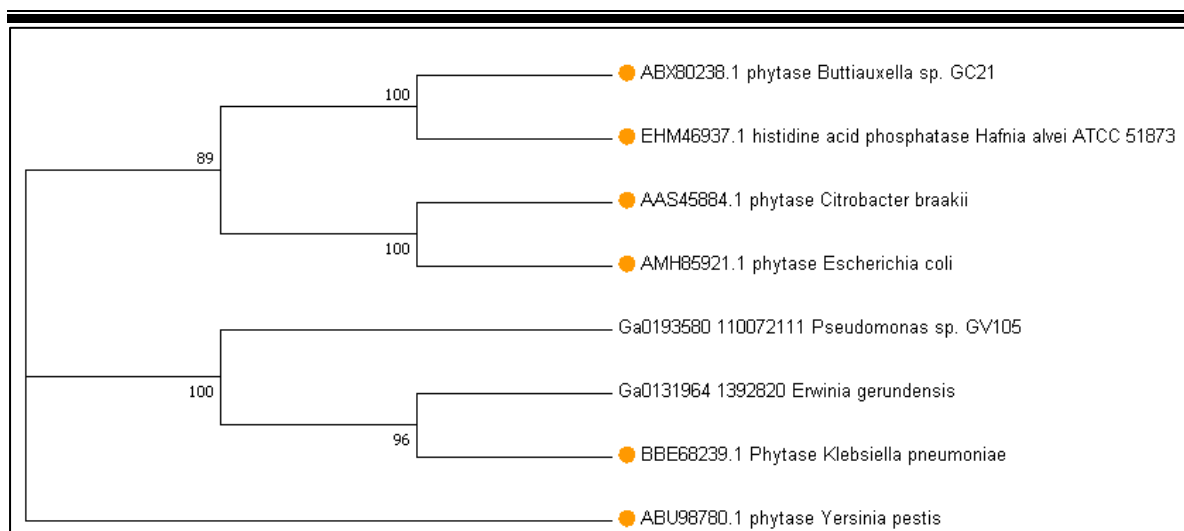


Figure 4. 28 : The Neighbor-Joining method (bootstrap 1000 replicates) was constructed to represent the evolutionary relationship of the bio-prospected potential phytase producers via the JGI-metagenome method, based on the metagenome sequences deposited in the database. The colored circle represents the well-characterized phytases.

4.4.3.8 Final shortlisted candidates from all databases:

In total, 67 candidates were shortlisted from the results of different *in-silico* bioprospecting methods **Table 4. 16.**

Chapter 4

Table 4. 16 The list of final shortlisted candidates:

Sr. No.	Name of the candidate	Accession number	Sr. No.	Name of the candidate	Accession number
1	<i>Budvicia aquatica</i>	A0A2C6DQ33	18	<i>Edwardsiella anguillarum</i>	A0A076LMH5
2	<i>Candidatus Sulfolobus solfataricus</i>	A0A2N9LY85	19	<i>Edwardsiella ictaluri</i>	C5BDT1
3	<i>Candidatus Sulfolobus solfataricus</i>	A0A2N9MZR7	20	<i>Erwinia mallotivora</i>	A0A014NCB0
4	<i>Caulobacter sp. UNC279MFTsu5.1</i>	A0A1I3UXR1	21	<i>Erwinia iniecta</i>	A0A0L7T3C3
5	<i>Caulobacter sp. BK020</i>	A0A4R3GVP7	22	<i>Erwinia sp. Leaf53</i>	A0A0Q4N476
6	<i>Cronobacter turicensis</i>	C9Y2R9	23	<i>Erwinia gerundensis</i>	A0A0U5KY50
7	<i>Cronobacter malonaticus</i>	A0A423XSW5	24	<i>Erwinia sp. OLMDLW33</i>	A0A2G8D6S0
8	<i>Cronobacter sakazakii</i>	A7MF64	25	<i>Erwinia sp. Jub26</i>	A0A3N1WG26
9	<i>Desulfovibrio legallii</i>	A0A1G7J6J8	26	<i>Silvibacterium bohemicum</i>	A0A841JP71
10	<i>Desulfovibrio intestinalis</i>	A0A7W8C2N5	27	<i>Kosakonia sacchari</i>	A0A1G4XQZ1
11	<i>Dickeya solani</i>	A0A2K8VZX8	28	<i>Kosakonia quasisacchari</i>	A0A4R0HFX2
12	<i>Dickeya dadantii</i>	B4XT20	29	<i>Granulicella mallensis</i>	G8P0L9
13	<i>Dyella marensis</i>	A0A1I2GM63	30	<i>Granulicella pectinivorans</i>	A0A1I6MRB5
14	<i>Dyella sp. SG609</i>	A0A846V6T3	31	<i>Obesumbacterium proteus</i>	Q6U677
15	<i>Dyella sp. 333MFSha</i>	A0A1G7L606	32	<i>Nitrospirillum amazonense</i>	A0A560EJH2
16	<i>Edwardsiella tarda</i>	A0A0H3DTQ2	33	<i>Nitrospirillum iridis</i>	A0A7X0AY01
17	<i>Edwardsiella piscicida</i>	A0A034T3G5	34	<i>Pantoea vagans</i>	A0A8X8DPR1

Chapter 4

Sr. No.	Name of the candidate	Accession number	Sr. No.	Name of the candidate	Accession number
35	<i>Pantoea wallisii</i>	A0A1X1DC61	52	<i>Pectobacterium actinidiae</i>	A0A1V2R0J8
36	<i>Pantoea rodasii</i>	A0A0B1RA14	53	<i>Pectobacterium Polaris</i>	A0A3S1G1Y1
37	<i>Pantoea sp. OV426</i>	A0A1I5ALB0	54	<i>Pectobacterium fontis</i>	A0A7V8ILW4
38	<i>Pantoea latae</i>	A0A1V9DQA5	55	<i>Pectobacterium sp. S27</i>	B6RGT0
39	<i>Pantoea conspicua</i>	A0A1X1BYI3	56	<i>Xanthomonas perforans</i>	A0A0G8U018
40	<i>Pantoea cypripedii</i>	A0A1X1EKT6	57	<i>Xanthomonas bromi</i>	A0A1C3NIQ0
41	<i>Pantoea allii</i>	A0A2V2BHL1	58	<i>Xanthomonas translucens</i>	A0A1M4J6D6
42	<i>Pantoea sp. PNA 03-3</i>	A0A2V3Q0A5	59	<i>Xanthomonas alfalfae</i>	A0A1S1XGR7
43	<i>Pantoea brenneri</i>	A0A653W753	60	<i>Xanthomonas axonopodis</i>	A0A1T1PAS6
44	<i>Pantoea stewartii</i>	A0A7U5G0D3	61	<i>Xanthomonas campestris</i>	A0A1T1R8I7
45	<i>Pantoea sp. SM3</i>	A0A837FM66	62	<i>Xanthomonas phaseoli</i>	A0A1V9H240
46	<i>Pantoea dispersa</i>	A0A8E1V8E5	63	<i>Xanthomonas citri</i>	A0A220WLT2
47	<i>Pectobacterium carotovorum subsp. Carotovorum</i>	B4XY38	64	<i>Xanthomonas prunicola</i>	A0A2N3RLE5
48	<i>Pectobacterium Brasiliense</i>	A0A086EQZ8	65	<i>Xanthomonas arboricola</i>	A0A2N9DBU3
49	<i>Pectobacterium betavasculorum</i>	A0A093RU79	66	<i>Xanthomonas vasicola</i>	A0A836ZS20
50	<i>Pectobacterium peruviense</i>	A0A0J5XCV4	67	<i>Xanthomonas hortorum</i>	V7ZGQ4
51	<i>Pectobacterium wasabiae</i>	A0A1D7YZ01			

4.5 Discussion:

Exploring the immense microbial diversity in natural environments, like soil, offers the potential to discover novel enzymes, including phytase (Suleimanova et al., 2015) (Sajidan et al., 2004). Each gram of soil contains an extensive variety of bacterial and archaeal species, ranging from 100,000 to 1,000,000 (Satyanarayana et al., 2017). The conventional method deals with the isolation and screening of culturable phytase producers from natural sources such as soil.

In this study, various soil samples from different regions of Maharashtra and Gujarat were explored to obtain novel phytase producers (Nagar et al., 2021). The potential phytase producers from soil samples were identified based on the zone of hydrolysis. The method of evaluating the hydrolysis zone is frequently employed to conduct an initial screening of phytase producers from different natural sources such as soil (Kalsi et al., 2016) (Dev et al., 2016). Additionally, phytase activity was examined in our study to precisely narrow down the sample count and validate the existence of phytase. The isolates exhibited phytase activity at pH 5, and the results were comparable with the reference, i.e., *E.coli* phytase. We adopted both qualitative plate assay and quantitative assay to isolate and screen six potential phytase producers, i.e., FD5T, FD5O, FD2 from forest soil, VD5 from vegetation soil, CD5 from cotton field, and PSD from Pigstag soil. These shortlisted soil isolates were identified via 16S rRNA gene sequencing using universal primers, and the results were analyzed using NCBI-Blastn. This analysis identified *Enterobacter cloacae* (VD5 and CD5), *Paenibacillus sp.* (FD5T), *Bacillus megaterium* (FD5O and FD2), and *Klebsiella variicola* (PSD). According to the literature, *Enterobacter* (Kalsi et al., 2016), *Klebsiella sp. ASRI* (Sajidan et al., 2004), which was isolated from soil, exhibited maximum activity at pH 7 and 5, respectively, supporting our results. Hence, these techniques provide increased assurance for subsequent analysis on a larger scale, which involves the purification and characterization of phytase using phytase purification and characterization methods. It should be noted that the capability to dissolve phosphate is not limited to a specific genus (Motamedi, 2016). Given the diverse range of soil microbes, conducting preliminary screening helps minimize the additional costs and resources associated with extensive experimentation on many isolates. Therefore, the conventional approach adopted in this presents a mini-scale alternative for identifying highly productive bacterial phytase producers (Nagar et al., 2021).

Chapter 4

We further selected PSD to investigate the presence of the phytase gene in it via PCR amplification and Sanger's sequencing. PSD was chosen as it was close to *Klebsiella*, which belongs to Histidine acid phytases (Sajidan et al., 2004). Histidine acid phytases have wide applications in industry compared to other phytases (Balaban et al., 2018) (Rao et al., 2009). The sequencing analysis confirmed that PSD belongs to the HAP group, as indicated by the presence of the N-terminal conserved region RHGXRXR in its sequence. Moreover, the isolated PSD demonstrated promising potential for further exploration in industrial applications.

The conventional approach can only be utilized to investigate the cultivable microbial population. Only 1% of the microbial communities can be cultured in laboratory conditions, while 99% are still unexplored because they are non-culturable (Robe et al., 2003) (Faria Fatima, Chaudhary Ira, Ali Jasarat, Rastogi Smita, 2011) (A. Kamble & Singh, 2020). To counter these limitations, an alternative option is to employ the metagenomic strategy centered around the 16S ribosomal RNA (16S rRNA) gene. This method allows for the exploration of the microbial community present in soil samples that are challenging to cultivate (Muwawa et al., 2021). In the context of this study, the metagenomic approach was harnessed to investigate the diversity of microorganisms within soil collected from a pig enclosure. The soil's elevated phytate content makes it a potential habitat for microorganisms capable of producing phytase, which is significant due to the incapability of monogastric animals to digest phytate (Menezes-Blackburn et al., 2013).

In the present study, the metagenomic approach included first, the isolation of total soil DNA from Pig stag soil (Kamble Asmita and Singh Harinder 2020), followed by 16S rRNA metagenomic sequencing. The analysis was performed using the QIIME2 platform (A. Kamble et al., 2020). The analysis helped to construct family and genus-level distributions of microbial communities present in the soil sample. The outcome was cross-referenced with existing literature to identify phytase producers, and it was discovered that the *Xanthomonadaceae* family holds potential for further investigation (DiFonzo & Bordia, 2003). *Xanthomonas* phytase can be further explored by designing full-length degenerate phytase gene primers to amplify the phytase gene, cloning, and overexpressing it in an appropriate host system, and ultimately characterizing the purified phytase.

In-silico bioprospecting, a computational approach, is a straightforward and efficient method to discover new candidates from existing databases (Voß et al., 2020). It allows us to overcome

the limitations of traditional approaches. The advent of sequencing technology has dramatically increased the number of sequences available in databases, thereby expanding the pool of enzyme diversity (Stewart, 2012). *In-silico* bioprospecting takes advantage of this diversity by employing the following steps: exploration of databases to find novel candidates, followed by screening, analysis, and shortlisting of these candidates using bioinformatics tools. The exploration of the database can be based on homology, conserved domain, keyword search, and consensus-guided (A. D. Kamble & Singh, 2021)(A. Kamble et al., 2019). For example, Gupta et al. explored the NCBI database using keywords "Hypothetical Protein of *T. aestivum*" and "Hypothetical Proteins of wheat," followed by Blastp search to find homologous sequences that fulfilled the criteria of >40% sequence identity plus an e-value <0.005. This process shortlisted potential candidates which were further studied using the HHpred tool (Gupta et al., 2018). Many other studies had adopted the in-silico bioprospecting approach to explore novel enzymes such as novel cyclic GMP-AMP synthase (cGAS) (Rolf et al., 2020), novel Urease (Yata et al., 2021), novel bilirubin oxidase (Sadeghian et al., 2020).

In the present study, different databases were explored, such as NCBI, Scan-Prosite, UniProtKB, Conserved Domain database, HMMER, and JGI-metagenome, based on homology, keyword search, conserved motif/domain, and consensus-guided approaches. The focus was to explore novel phytases that were 50-80% identical to experimentally well-characterized phytases. The steps involved exploring databases, followed by removing redundancy and studying their evolutionary relationship by constructing a phylogenetic neighbor-joining tree. Finally, we obtained 67 different candidates as potential phytase producers.

To summarize, the utilization of three different approaches—conventional, metagenomics, and *in-silico* bioprospecting—enabled the identification and screening of new phytases. Each method yielded distinct phytase producers, and the combination of all three approaches expanded the potential for discovering additional novel phytases. In this study, *Pantoea vagans* (PV) and *Edwardsiella tarda* (ET) phytases were shortlisted for our next objective i.e. *in-silico* characterization and experimental validation.

4.6 Summary and Conclusion:

In this present study, three different approaches were used to identify potential phytase producers: conventional, metagenomics, and *in-silico* bioprospecting.

Conventional Bioprospecting: Alongside the initial screening technique involving the assessment of the hydrolysis zone, this study also utilized activity determination to accurately identify bacteria capable of producing phytase. By employing these methods on a smaller scale, the number of samples under consideration was reduced, and the presence of phytase was definitively confirmed. This approach instilled greater confidence for subsequent analysis, typically encompassing the broader processes of phytase purification and thorough characterization.

Metagenomics Approach: The research optimized procedures for extracting DNA from soil, and using a technique known as metagenomics. Different methods and soil DNA kits were tested to achieve a high yield of pure DNA from garden soil. The DNA yield ranged from 15 to 9218 ng/ μ l, with the method involving a soil DNA kit producing the highest-quality DNA (A_{260}/A_{280} : 1.8). Through metagenomic analysis involving 16S rRNA gene sequencing of soil DNA from pig sty, the study identified the distribution of bacterial families and genera. Notably, the *Xanthomonadaceae* family emerged as a promising candidate for further investigation into phytase activity.

***In-silico* Bioprospecting:** In essence, *in-silico* bioprospecting methods were instrumental in identifying a total of 67 potential genera capable of producing phytase from various databases. Out of these 67, two candidates—*Pantoea vagans* (PV) and *Edwardsiella tarda* (ET)—were selected for comprehensive computational assessment and subsequent experimental characterization.

Hence, the combination of these three methods enabled us to explore a vast array of novel phytases, and among them, PV and ET were shortlisted for further study.

Chapter 5
In-silico
characterization
and
experimental
validation of
shortlisted phytase

In-silico characterization and experimental validation of shortlisted phytase

5.1 Introduction

This chapter consisted of primary, secondary, and tertiary structure analysis and validation and functional analysis of PVP and ETP. This study also consisted of molecular docking and molecular dynamics simulation studies to gain knowledge on the dynamics of PVP and ETP at different temperatures from 300 to 450K for 30ns. The experimental validation of ETP and PVP included confirmation of the phytase gene, cloning of the phytase gene in the pET29-b vector, overexpression of phytase in *E. coli* BL21(DE3), and biochemical characterization. The analysis will help to understand these two novel phytases i.e., ETP and PVP, and shortlist the candidates for rational engineering to improve their properties such as thermostability.

5.1.1 *In-silico* characterization

The *in-silico* characterization process involved comprehending the primary, secondary, and tertiary structures alongside validation, functional assessment, and dynamic studies at different temperatures. E.g., *In-silico* characterization has been reported for *Pseudomonas aeruginosa* phosphatase (Pramanik et al., 2017), thermostable chitinase II (Iqbal et al., 2015), Cellobiohydrolase (Dadwal et al., 2022) and phytases (Gontia-Mishra et al., 2014). Analyzing the primary sequence allowed predictions of enzyme characteristics and molecular properties. Secondary structure analysis yielded insights into folding, stability, and functional aspects. Evaluating the tertiary structure determined the reliability of the model and provided details about local structural features. Functional analysis revealed details about conserved amino acids, disulfide bonds, and stability such as T_m, offering insight into the enzyme's potential for further investigation. Molecular dynamics simulations provided data on temperature-induced structural changes, stability, and accessibility to solvents. E.g., molecular dynamic simulation (MDS) of chitosanase (Wang et al., 2023), glutamate decarboxylase (Hua et al., 2020), *Aspergillus niger* phytase (K. Kumar et al., 2015), and *E.coli* phytases (Shivange et al., 2010) helped to gain insights into conformational dynamics of various enzymes. All this information contributes to understanding an enzyme's potential for rational engineering purposes.

5.1.2 Experimental validation

Experimental validation helps to understand the characteristics of enzymes for their potential use in various industries. E.g., 1,3(4)- β -D-glucanase (Chen et al., 2020), metalloprotease (Zhu et al., 2023), *Hafnia* phytase (Ariza et al., 2013), *E.coli* phytase (Abeldenov et al., 2017) and *Klebsiella* phytase (Sajidan et al., 2004) are well-characterized enzymes reported in the past.

In the present study, experimental validation involved confirmation of the phytase gene present in PV and ET, cloning and overexpression of phytase, purification, and biochemical characterization. This will help us to shortlist the phytase for rational engineering.

5.2 *In-silico* characterization:

5.2.1 Introduction:

In-silico characterization of shortlisted phytase includes primary sequence analysis, secondary structure analysis, tertiary structure analysis and validation, functional analysis, and molecular docking and molecular dynamics simulations (**Figure 5. 1**).

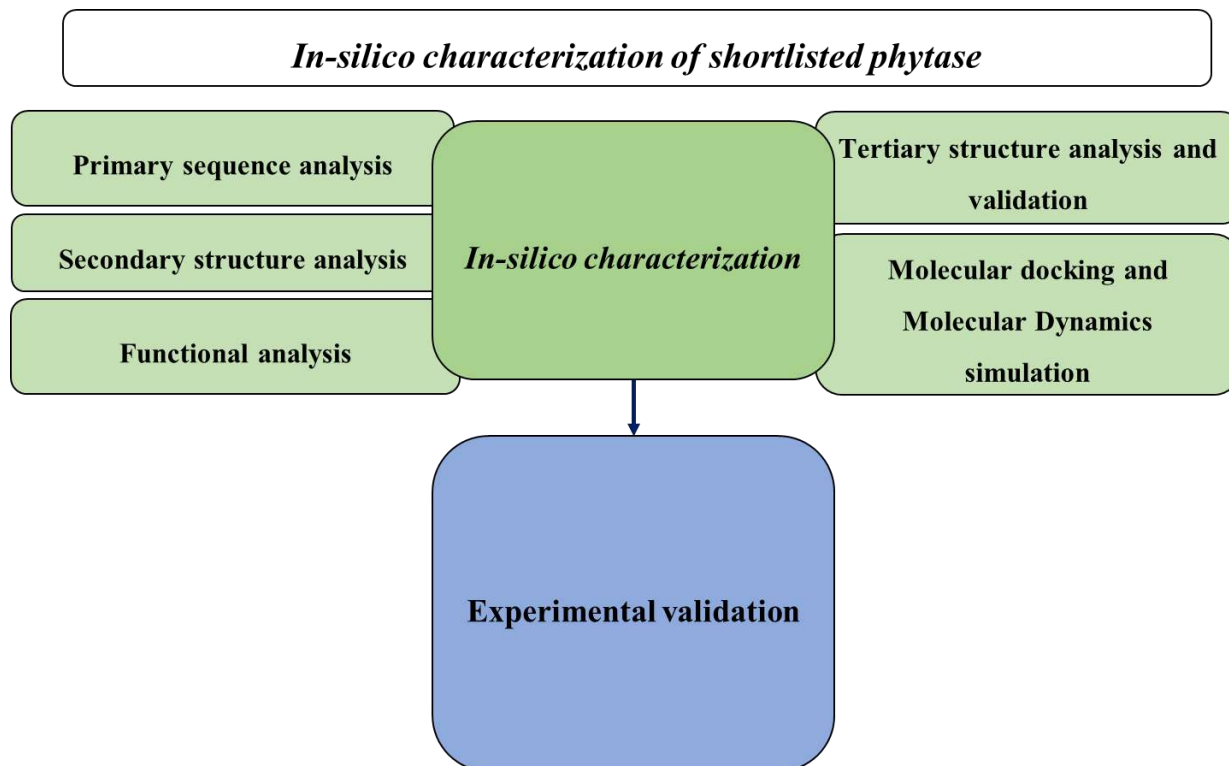


Figure 5. 1 Overview of *in-silico* characterization of PVP and ETP.

5.2.2 Materials and Methods:

5.2.2.1 Sequence retrieval, phylogenetic analysis, and primary sequence analysis:

Sequence retrieval:

The full-length phytase sequence of *Pantoea vagans* phytase (PVP) (E1SF61) and *Edwardsiella tarda* phytase (ETP) (A0A2A7U467) was retrieved from the UniProtKB database (Magrane & Consortium, 2011) (<https://www.uniprot.org/>) in FASTA format. The nucleotide sequence of PVP (Accession I.D.: ADO08555) and ETP (Accession I.D.: PEH73109) was retrieved from ENA i.e., European Nucleotide Archive. ClustalO (Sievers et al., 2011) (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to generate multiple sequence alignment (MSA) of PVP and ETP against experimentally characterized phytases from *E.coli*, *Klebsiella*, *Citrobacter*, *Hafnia*, and *Yersinia*. The MSA generated by ClustalO was edited by using Easy Sequencing in PostScript (ESpript) (<https://escript.ibcp.fr/ESpript/cgi-bin/ESpript.cgi>) (Gouet et al., 1999).

Phylogenetic analysis:

MEGA 7.0 (S. Kumar et al., 2016) (<https://www.megasoftware.net/>) was used to construct the phylogenetic tree. The muscle tool was first used to generate MSA of PVP and ETP against the experimentally characterized phytases from *E.coli*, *Klebsiella*, *Citrobacter*, *Hafnia*, and *Yersinia*. The generated MSA was then subjected to neighbor-joining phylogenetic tree construction with 1000 bootstrap replicates and JTT-matrix-based method/ Poisson correction method.

Primary sequence analysis:

ProtParam tool on ExPasy proteomics server (<https://web.expasy.org/protparam/>) (Gasteiger et al., 2005) and InterProScan (Mulder & Apweiler, 2007) (<https://www.ebi.ac.uk/interpro/search/sequence/>) was used for primary sequence analysis of ETP, PVP along with its closet homology i.e., *Hafnia alvei* phytase (HAP) and *Klebsiella pneumoniae* phytase (KPP). This analysis helped to gather information related to physiochemical properties such as isoelectric point, molecular weight, aliphatic index, amino acid composition, and grand average of hydropathicities (GRAVY). InterProScan helped to gain information related to family, CDD, catalytic core, Pfam I.D., Prosite I.D., and homologous superfamily.

5.2.2.2 Secondary structure analysis:

Secondary structure analysis of PVP and ETP was performed by using various tools such as Psipred (McGuffin et al., 2000) (<http://bioinf.cs.ucl.ac.uk/psipred/>), Polyview2D (Porollo et al., 2004) (<https://polyview.cchmc.org/>), PDBsum (Laskowski et al., 2018) (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>), Proteus (Montgomerie et al., 2008) (<http://www.proteus2.ca/proteus2/>), Yet Another Scientific Artificial Reality Application (YASARA) software (Land & Humble, 2018) (<http://www.yasara.org/>).

These tools helped to gain information related to the secondary structure elements such as helices, sheets, turns, and coils of PVP and ETP. Each of these tools/serves/software operates via inbuilt algorithms and statistical methods e.g., Psipred used neural network algorithm and recognized patterns. Polyview2D helped to generate a secondary structure by using different methods that were embedded in it. PDBsum helped to generate a wiring diagram of the secondary structure. Proteus used machine learning methods along with different secondary structure methods. Similarly, YASARA also used different methods to generate a secondary structure.

5.2.2.3 Tertiary structure analysis and validation:

Tertiary structure analysis was performed by using the following tools/serves/ software:

Funfold (Roche et al., 2011) (<https://ffas.godziklab.org/ffas-cgi/cgi/ffas.pl>), Hhpred (Zimmermann et al., 2018) (<https://toolkit.tuebingen.mpg.de/tools/hhpred>), Phyre (Kelley et al., 2016), Swiss Model (Bienert et al., 2017) (<https://swissmodel.expasy.org/>), Fold and Function Assignment System (FFAS) (Xu et al., 2014) (<https://ffas.godziklab.org/ffas-cgi/cgi/ffas.pl>), Iterative Threading ASSEmbly Refinement (I-TASSER) (Yang et al., 2014) (<https://zhanggroup.org/I-TASSER/>), Intfold (Mcguffin et al., 2019) (<https://www.reading.ac.uk/bioinf/IntFOLD/>) were used to model tertiary structure of PVP and ETP.

The Structure Assessment tool embedded in the EXPASY web portal (<https://swissmodel.expasy.org/assess>) was used for the preliminary screening of PVP and ETP structures generated by various tools/serves/ software. The evaluation was based on the percentage of favored regions in the Ramachandran plot. The structures that exhibited >90% of their amino acids in the Ramachandran favored region were shortlisted.

Chapter 5

SWISS-MODEL structures were shortlisted for further validation via Protein Structure Analysis (ProSA) (<https://prosa.services.came.sbg.ac.at/prosa.php>) (Wiederstein & Sippl, 2007), ERRAT (<https://saves.mbi.ucla.edu/>) (Colovos & Yeates, 1993), Qualitative Model Energy Analysis (QMEANDisCo) (Benkert et al., 2011) (<https://swissmodel.expasy.org/qmean/>).

The modeled structures were visualized using Pymol (<https://pymol.org/2/>), BioVia Discovery studio (<https://discover.3ds.com/discovery-studio-visualizer-download>, Chimera (Pettersen et al., 2004).

5.2.2.4 Functional analysis:

The following tools/servers/software were used for functional analysis:

InterProScan (Mulder & Apweiler, 2007) (<https://www.ebi.ac.uk/interpro/search/sequence/>), Prosite (de Castro et al., 2006) (<https://prosite.expasy.org/>), MotifFinder (https://software.broadinstitute.org/software/igv/motif_finder), SignalP (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>) (Almagro Armenteros et al., 2019), Protein interaction calculator (PIC) (Tina et al., 2007) (<http://pic.mbu.iisc.ernet.in/job.html>), VMD, Castp (Tian et al., 2018) (<http://sts.bioe.uic.edu/castp/index.html?3igg>) and SCooP (Pucci et al., 2017) (<http://babylone.3bio.ulb.ac.be/SCooP/index.php>).

5.2.2.5 Molecular docking and Molecular dynamics simulation:

Molecular docking was performed by using AutoDock Vina (Trott & Olson, 2009) (Faiza, 2020) (<https://autodock.scripps.edu/>). AutoDock Vina was used to perform docking of phytate with PVP and ETP as mentioned below:

1. Model minimization: Modelled structures of ETP and PVP were minimized by using Swiss PDB Viewer (<https://autodock.scripps.edu/>).
2. Ligand preparation: Ligand i.e., phytate was extracted from 1DKQ by using Chimera software. The protein part was deleted and the non-standard residue (i.e., 1-IHP) was selected. The minimization process was conducted by using the steepest descent steps (1000) and conjugate gradient steps. The charges and hydrogens were added, followed by exposing it to the Amber ff14SB force field.
3. .pdbqt files generation: .pdbqt files for phytate, PVP, and ETP were generated by using AutoDock4. The phytate, ETP, and PVP structure was opened in AutoDock4. In the case of ETP and PVP.pdb structures: polar hydrogens were added, water molecules

were deleted, and kollam charges were added (atoms AD4 type). The PVP structure was saved as a .pdbqt file. In case of phytate.pdb: opened in Autodock4 and detected root feature was applied. This file was saved as phytate.pdbqt.

4. Highlighting the conserved residues: The conserved residues involved in the catalysis of phytate by PVP and ETP were highlighted. In-case of PVP the conserved residues were: R38, H39, G40, V41, R42, P43, P44, H303, and D304. In-case of ETP the conserved residues were N and C-terminal conserved residues: R46, H47, G48, V49, R50, A51, P52, H331, D332, T333, and N334.
5. Grid box generation was generated for local docking.
6. Generation of configuration file: The following information regarding receptor, ligand, size of grid box, exhaustiveness, and output was stored in the conf.txt file.

The following is the .conf.txt file: The following is an example of the configuration file.

Note: all parameters are editable depending on the input minimized protein and ligand structure.

```
receptor= minimized_protein.pdbqt
```

```
ligand= minimized_ligand.pdbqt
```

```
center_x= -32.692
```

```
center_y= 47.899
```

```
center_z= -2.086
```

```
size_x= 40
```

```
size_y= 40
```

```
size_z= 40
```

```
out= vina_pvwithmini.pdbqt
```

```
log= lig.txt
```

```
exhaustiveness= 100
```

```
num_modes = 100
```

Chapter 5

This conf.txt file was run on the Windows command prompt to obtain ETP and PVP bound to phytate.

7. Pymol visualization: the complex of ETP and PVP-bound phytate was visualized in Pymol.
8. The 2D structure of PVP and ETP phytate complex to identify interactions was generated by BioVia Discovery studio (<https://discover.3ds.com/discovery-studio-visualizer-download>).

Molecular dynamics simulation (MDS):

The docked structures of PVP and ETP bound phytate were opened in Chimera. The protein structure i.e., PVP and ETP (remove phytate) was saved as a .pdb file and the phytate file (remove PVP and ETP) was saved as a .mol2 file.

GROMACS software (Pronk, 2013) was used to prepare PVP.pdb and ETP.pdb files for simulation. The PVP and ETP topology were generated by using the pdb2gmx command and AMBER99SB-ILDN forcefield. This process resulted in the generation of .gro files of PVP and ETP plus phytate structures. The complex file was built by merging PVP/ETP and phytate coordinates (present in .gro file) and updated the system's topology files such as .itp and, prm, topol files.

The system was then surrounded by a cubic box. In the case of PVP the box size was 0.3nm and in the case of ETP it was 0.35nm. Ions (Na⁺ and Cl⁻) were added to the system. It was minimized by using the steepest descent minimization method, maximum forces were < 10.0 kJ/mol and maximum 50000 nsteps.

The two-step equilibration method was applied. Phytate undergoes restraint by generating position restraint topology followed by temperature coupling groups. NVT and NPT equilibration was performed by using a modified Berendsen thermostat and Parrinello-Rahman pressure coupling for 1000 ps each.

The MD production of the system was carried out using leapfrog integrator, 1.2 nm short-range van der Waals cutoff, periodic boundary conditions, and long-range electrostatics via Particle Mesh Ewald (PME) for 30ns at 300, 350, 400, and 450K.

The MD simulation analysis involved: Root Mean Square Deviation (RMSD), Root Mean Square fluctuation (RMSF), Solvent Accessible Surface Area (SASA) of protein, and Radius of Gyration (Rg).

5.2.3 Results:

5.2.3.1 Sequence retrieval, phylogenetic analysis, and primary sequence analysis:

The full-length phytase sequence of PVP (E1SF61) and ETP (A0A2A7U467) was retrieved from the UniProtKB database (Magrane & Consortium, 2011) (<https://www.uniprot.org/>) in FASTA format. PVP and ETP consisted of 418 and 443 amino acids respectively. Multiple sequence alignment was performed using Clustal Omega (**Figure 5. 2**) (**Figure 5. 3**) and further edited by ESPript. This analysis revealed that the closest homolog of PVP and ETP were *Klebsiella pneumoniae* phytase (KPP) (67.43% percent identity) and *Hafnia alvei* phytase (HAP) (56.76% percent identity) respectively. The Clustal Omega analysis edited via ESPript (**Figure 5. 4**) (**Figure 5. 5**) also revealed that both PVP and ETP belonged to the Histidine acid phytase group as it consisted of N-terminal (RHGXRX) and C-terminal (HD) conserved amino acids residues.

Chapter 5

E1SF61	MTLPTL----CRCALILGSL--WLLSPATQAADYQLEKVVLSRHGVRPPTPGNRKEMEA	54
2WNH	-----MD--IGINSDPPPRDWQLEKVVLSRHGIRPPTAGNREAIEA	40
1DKL	-----QSEPELKLESVVIVSRHGVRAPTKAT-QLMQD	31
3ZHC	--MSTFIIRLLIFSLLCGSFSI----HAEEQNGMKLERVVIVSRHGVRAPTKFT-PIMKN	53
4ARO	-----SDTAPAGFQLEKVVILSRHGVRAPTKMT-QTMRD	33
4ARV	MTVAKKYLRLSVLTVLVSSFTLSAAPLAAQSTGYTLERVVILSRHGVRSPTKQT-QLMND	59
	*** ** :****:* ** . :.	
E1SF61	ASQQPWTQWTTADGELTGHGYSVVNKGWEGEHYRQLGLLGT-GCPDAAQVYVRASPLQ	113
2WNH	ATGRPWTEWTTDHDGELTGHGYAAVVNKGREGQHYRQLGLLQA-GCPTAESIYVRASPLQ	99
1DKL	VTPDAWPTWPVKLGWLTPRGGELIAYLGHYQRQLVADGLLAKKGCPSGQVAIIADVDE	91
3ZHC	VTPDQWPQWDVPLGWLTPRGGELVSELGQYQRLWFTSKGLLNNQTCPSPGQVAIIADTDQ	113
4ARO	VTPHQWPEWPKLGYITPRGEHLISLGGFYRERFQQGGLLPKDNCPDPAVYVWADVDDQ	93
4ARV	VTPDKWPQWPVKAGYLTPRGAGLVTLMGGFYGDYFRSYGLLPA-GCPADESIYVQADVDDQ	118
	.. * * . * : * : * : * * * * * : : * . :	
E1SF61	RTRATAAALTDGAFPGCGVTVHHVAG--DVDPLFQGEKLTVTRTDPAQELAAKQQKAGD-	170
2WNH	RTRATAQALVDGAFPGCGVAIHYANG--DADPLFQTDKFAATQTDPARQLAAVKEKAGD-	156
1DKL	RTRKTGEAFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGS	151
3ZHC	RTRKTGEAFLAGLAPKCQIQVHYQKDEEKNDPLFNPVKMGKCSFNTLQVCNAILERAGGN	173
4ARO	RTRKTGEAFLAGLAPQCDLAIHHQNTQQADPLFHPVKAGICSMDSQVHAAVEKQAGTP	153
4ARV	RTRLTGQAFLDGIAPDCGLKVHYQADLKKIDPLFHTVEAGVCKLDPEKTHAVEKRLGGP	178
	*** * . * : * * * : : * . . ***** : : : . * . : * :	
E1SF61	LAHLQQQLQPAIQQLKA-----AVCPP----ATDCPLFEA-PWTFRQTRNG-NSVYVG	217
2WNH	LAQRRQALAPTIQLLKQ-----AVCQA----DKPCPIFDT-PWRVEQSKSG-KTTISG	203
1DKL	IADFTGHRQTAFRELERVLNFPQSNLCLKREKQDECSLTQALPSELKVSAD--NVSLTG	209
3ZHC	IELYTQRYQSSFRLENVLNFSQSETCKTTEK-STKCTLPEALPSELKCTPD--NVSLPG	230
4ARO	IETLNQRYQASLALMSSVLDLDFPKSPYQCOH-NIGKLCDFSQAMP SRLAINDDGNKVALEG	212
4ARV	LNELSQRyakPFALMGEVLNFSASPYCNSLQKQKACDFATFAANEIEVNKEGTVKVSLSG	238
	: : : * : * * : * : * : * : * : * : * : * : * : * : * :	
E1SF61	LSVMASM-VETLRLGYSENLPFDQLAWGHITTAQITSLPLLTANYDLSNDVLYLAQRR	276
2WNH	LSVMANM-VETLRLGWSENLPLSQLAWGKIAQASQITALLPLLTENYDLSNDVLYTAQKR	262
1DKL	AVSLASMLTEIFLLQ--QAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSR	267
3ZHC	AWLSSTLTEIFLLQ--EAQGMPPQVAVGRITGEKEWRDLLSLHNAQFDLLQRTPEVARSR	288
4ARO	AVGLASTLAEIFLLE--HAQGMPPVAVWGNIHTEQQWNSLLKLHNAQFDLMSRTPYIAKHN	270
4ARV	PLALSSTLGEIFLLQ--NSQAMPDVAVNRLSGEENWISLLSLHNAQFDLMAKTPYIARHK	296
	::. * : * . : . . * . : : * * * . : : * . * : . :	
E1SF61	GSILLNTMLEAIA-----ADSSP-GRWLVLVAHDTNIAMVRTLMDFNWQLPGYSRG	326
2WNH	GSVLLNAMLDGVK-----PEASPNVRWLLLVAHDTNIAMVRTLMNFSWQLPGYSRG	313
1DKL	ATPLLDLIKTAALTPHPP-QKQAYGVTLPSTVLFIAGHDTNLANLGGALELNWTLPGQP-D	325
3ZHC	ATPLLDMIDTALLTNGT-TENRYGIKLPVSLFLIAGHDTNLANLSGALDLNWSLPGQP-D	346
4ARO	GTPLLQTIahalgsnit-SRPLPDISPDNKILFIAGHDANIANSGLMGTWTLPGQP-D	328
4ARV	GTPLLQQIDTALVLRDADQQTLPSPQTKLLFLGGHDTNIANIAGMLGANWQLPQQQP-D	355
	.. ** : : . : * : * * * * * : * : * * * * * : * * * * * :	
E1SF61	NIPPGSSLVLERWRDTRSGERFLRLYFQAQSLDGIRQLQPIDDK-HPLLRQEWHPDCRV	385
2WNH	NIPPGSSLVLERWRDAKSGERYLRVYFQAQGLDLDLRLQTPDAQ-HPMLRQEWHPGCRQ	372
1DKL	NTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQQMRDKTPLSLN-TPPGEVKLTLGCEE	384
3ZHC	NTPPGGELVFERWKRSDNTDQVQVSVFYQTLRDMRDIQPLSLE-KPAGKVDLKLIIACEE	405
4ARO	NTPPGGALVFERWVDNA-GKPYVSVNMVYQTLAQLHDQAPLTLQ-HPAGSVRLNIPGCS	386
4ARV	NTPPGGLVFEWQNPDNHQRYVAVKMFYQTMEQLRNADKLDLKNPARIPIAIEGCEN	415
	* * * . * : * * : : : * : : : : * : * : * :	
E1SF61	TAVGLLCPYQSTLTQLRKNLNSAVLPVSVILP----- 418	418
2WNH	TDVGTLCFQAAITALGQRIDRPSAPAVAMVLPKLAALAEHHHHHH 418	418
1DKL	RNAQGMCSLAGFTQIVNEARI PACSL----- 410	410
3ZHC	KNSQGMCSLKSFSRLIKEIRVPECAVTE----- 433	433
4ARO	QTPDGYCPLSTFSRLVSHSVEPACQLP----- 413	413
4ARV	EGDNKLCQLETfQKKVAQVIEPSCHI----- 441	441
	* : : . . .	

Figure 5. 2 Multiple sequence alignment generated by using ClustalO: PVP (Uniprot accession number: E1SF61), against experimentally characterized phytases from *E.coli* (PDB I.D.: 1DKL), *Klebsiella* (PDB I.D.: 2WNH), *Citrobacter* (PDB I.D.: 3ZHC), *Hafnia* (PDB I.D.: 4ARO), *Yersinia* (PDB I.D.: 4ARV).

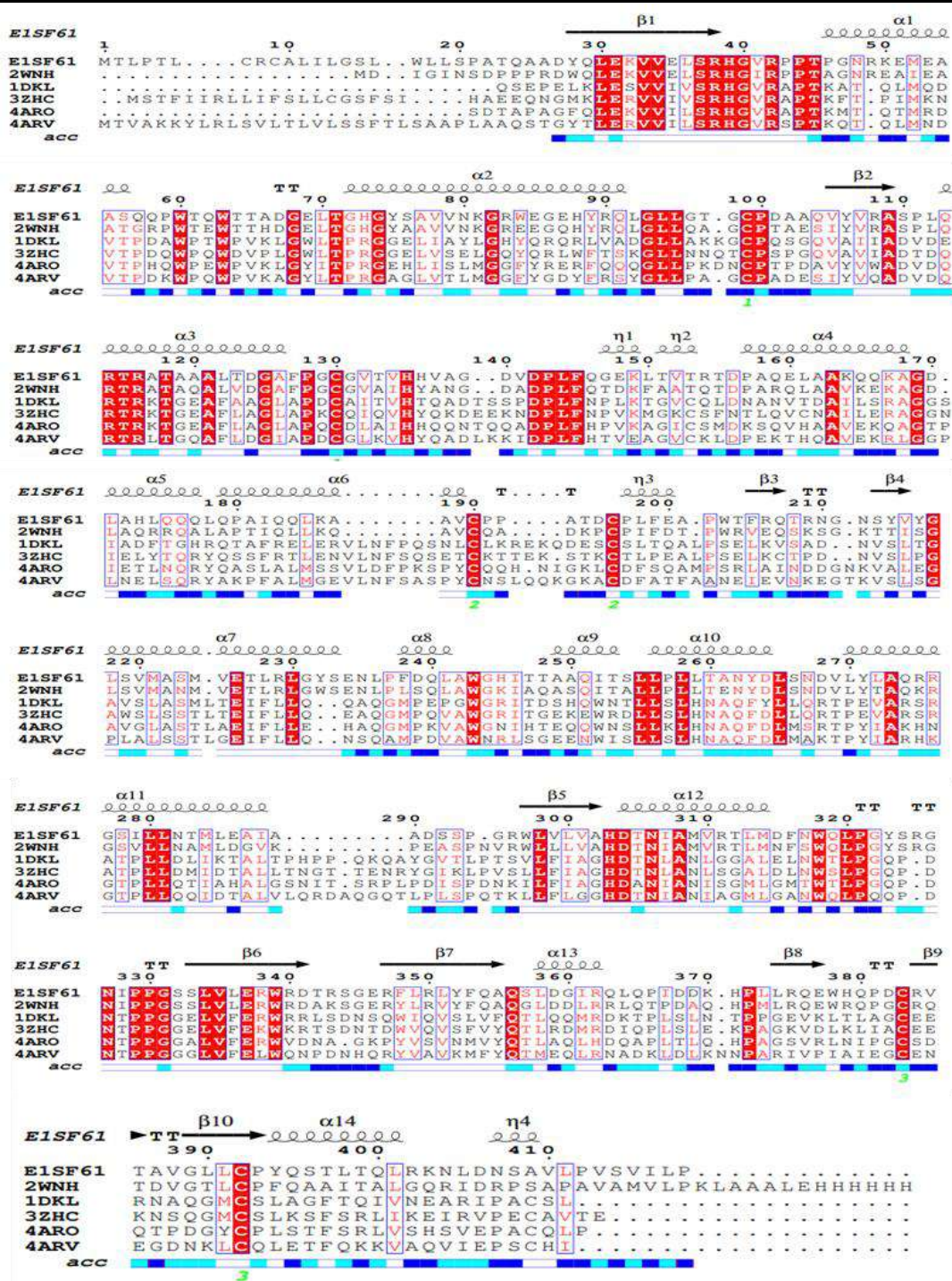


Figure 5. 4 Clustal O was used to generate multiple sequence alignment of PVP (UniProt accession no.: E1SF61) against the well-characterized phytases (retrieved from RCSB PDB) and edited by Easy Sequencing in PostScript (EScript). Note: well-characterized phytases from *E.coli* (PDB accession I.D: 1DKL), *Klebsiella* (PDB accession I.D: 2WNH), *Citrobacter* (PDB accession I.D: 3ZHC), *Hafnia* (PDB accession I.D: 4ARO), *Yersinia* (PDB accession I.D: 4ARV) was used for the alignment.

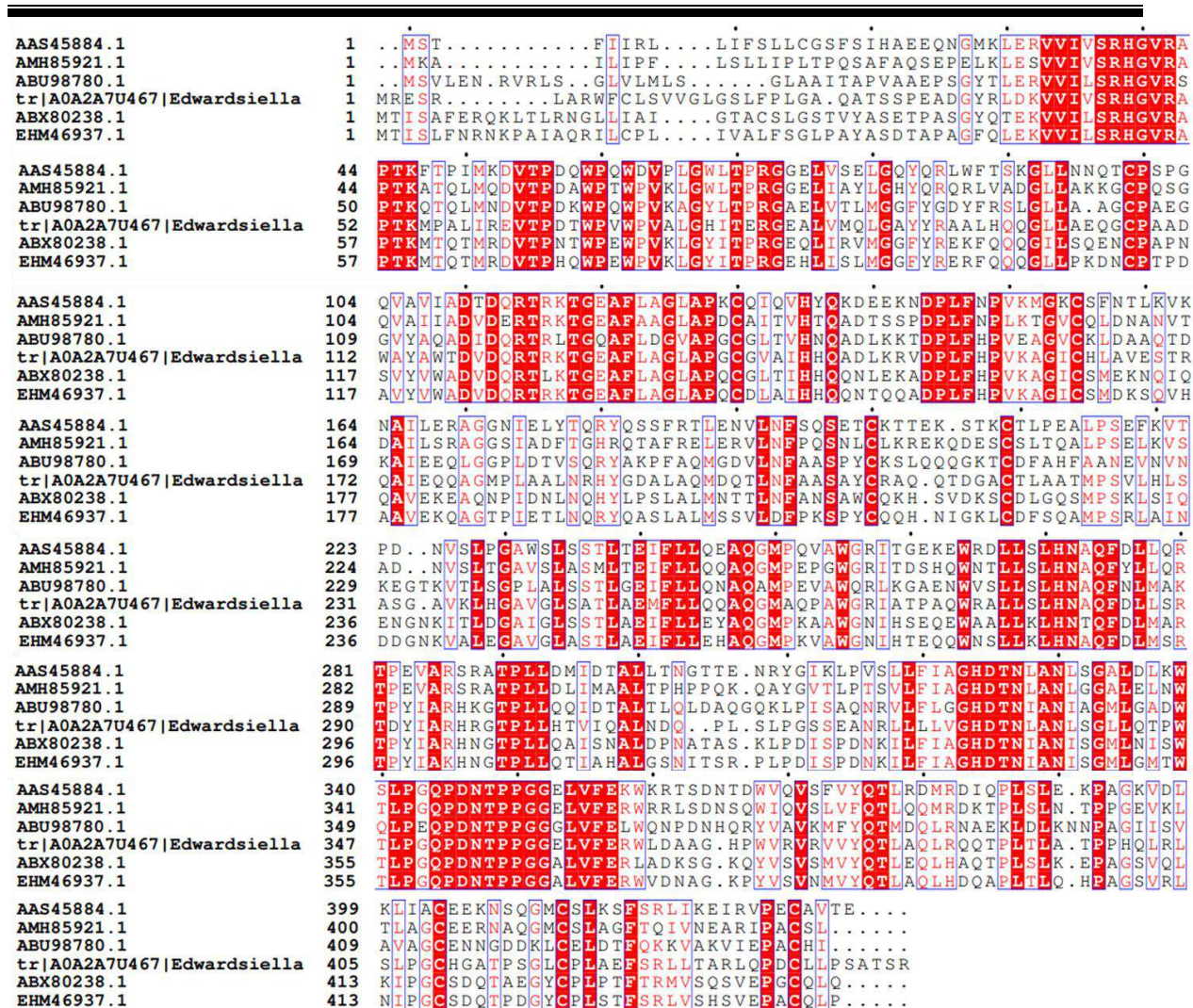


Figure 5. 5 Clustal O was used to generate multiple sequence alignment of ETP (Uniprot accession no.: A0A2A7U467) against the well-characterized phytases (retrieved from RCSB PDB) and edited by Easy Sequencing in PostScript (EScript). Note: well-characterized phytases from *Citrobacter braakii* phytase (NCBI accession I.D.: AAS45884.1), *Escherichia coli* phytase (NCBI accession I.D.: AMH85921.1), *Yersinia pestis* (NCBI accession I.D.: ABU98780.1), *Buttiauxella sp. GC21* phytase (NCBI accession I.D.: ABX80238.1), and *Hafnia alvei* ATCC 51873 (NCBI accession I.D.: EHM46937.1)

The neighbor-joining phylogenetic tree was constructed by using MEGA 7.0 software to reveal the evolutionary relationship of PVP (Figure 5. 6) and ETP (Figure 5. 7) with the well-characterized phytases. PVP was closely clustered with KPP, while ETP was closely clustered with well-characterized phytases.

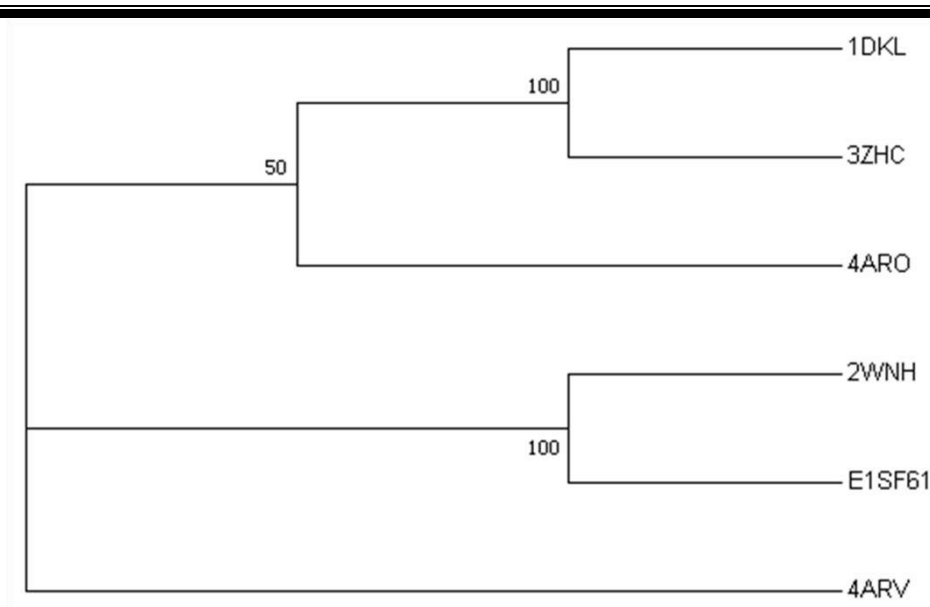


Figure 5. 6 The neighbor-joining phylogenetic analysis of PVP (Uniprot accession I.D.: E1SF61) along with reference sequences such as from *E.coli* (PDB accession I.D: 1DKL), *Klebsiella* (PDB accession I.D: 2WNH), *Citrobacter* (PDB accession I.D: 3ZHC), *Hafnia* (PDB accession I.D: 4ARO), and *Yersinia* (PDB accession I.D: 4ARV) was performed by using MEGA 7.0 software (Poisson correction method, 1000 bootstrap replicates).

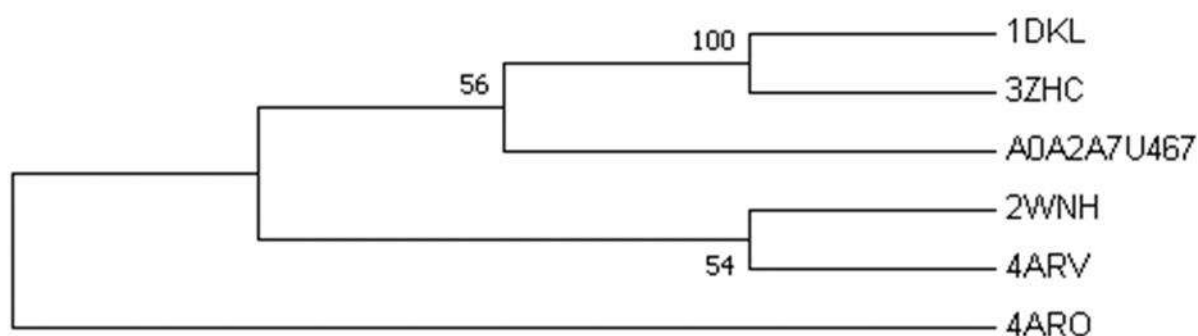


Figure 5. 7 The neighbor-joining phylogenetic analysis of ETP (Nucleotide accession I.D.: A0A2A7U467) along with reference sequences such as from *E.coli* (PDB accession I.D: 1DKL), *Klebsiella* (PDB accession I.D: 2WNH), *Citrobacter* (PDB accession I.D: 3ZHC), *Hafnia* (PDB accession I.D: 4ARO), *Yersinia* (PDB accession I.D: 4ARV) was performed by using MEGA 7.0 software (JTT matrix-based method, 1000 bootstrap replicates).

Primary sequence analysis was performed using the Protparam tool (Gasteiger et al. 2005) (<https://web.expasy.org/protparam/>) and Interproscan database (Mulder and Apweiler 2007) (<https://www.ebi.ac.uk/interpro/search/sequence/>). The Protparam tool was used to determine

Chapter 5

the amino acid composition, isoelectric point, molecular weight, aliphatic index, and the grand average of hydropathicities (GRAVY) for PVP, ETP, KPP, and HAP phytases. Note that the results of PVP were compared with KPP (used as reference phytase) whereas ETP characteristics were compared with HAP (used as reference phytase).

The ProtParam tool revealed that PVP, and ETP consisted of 418 and 444 amino acids which were comparable to their reference i.e., KPP (418 aa) and HAP (413 aa) respectively. The molecular weight of PVP, ETP, HAP, and KPP phytases were 46, 48, 45, and 46 kDa respectively. The pI of ETP, PVP, HAP, and KPP was 6.67, 6.79, 6.48, and 8.48 respectively. This indicated that PVP and ETP phytase were mildly acidic. The aliphatic index of ETP, PVP, HAP, and KPP was 80.05, 91.29, 82.71, and 84.59 respectively. This indicated that PVP was more thermostable as compared to ETP and references. The GRAVY index of PVP, ETP, HAP, and KPP was -0.279, -0.211, -0.283 and -0.378 respectively. This revealed that PVP and ETP both had better interactions with the surrounding aqueous environment. The details are mentioned in **Table 5. 1**. The amino acid composition results revealed that PVP, ETP, and HAP composed of the highest Leu amino acid whereas in the case of KPP, Ala amino acid was slightly higher than Leu (**Figure 5. 8**).

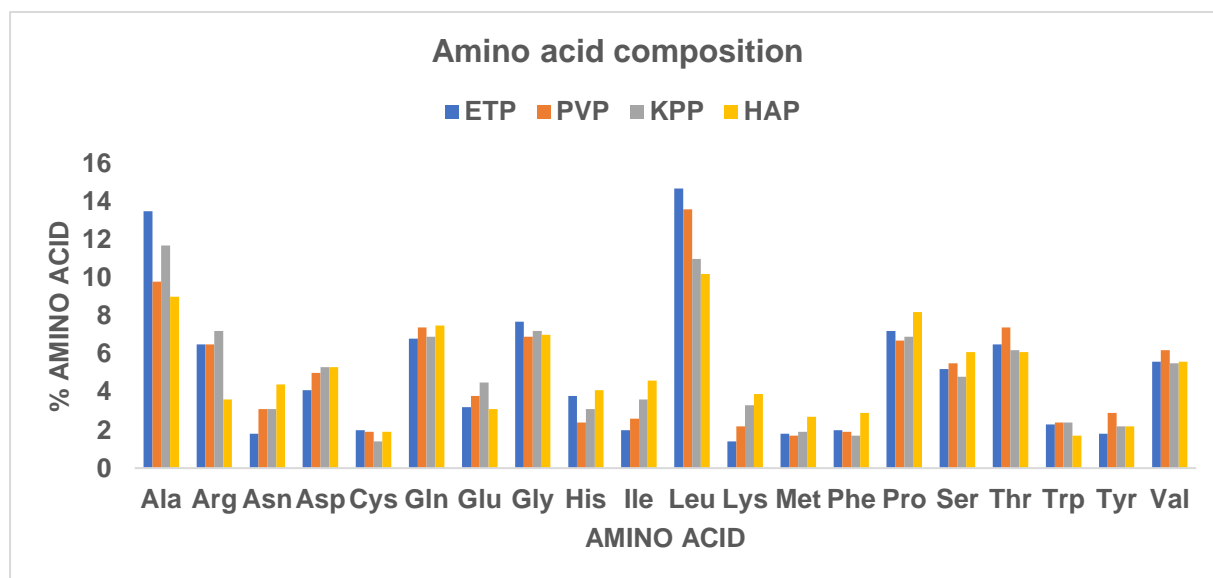


Figure 5. 8 The amino acid composition (%) of ETP, PVP, KPP, and HAP.

The overall ProtParam analysis revealed that the primary sequence properties of PVP and ETP were like those of the references. Hence, the sequence is reliable and can be used to study further.

Table 5. 1 The Protparam tool was used for the primary sequence analysis of PVP, KPP, ETP, and HAP.

Wildtype phytase	No. of amino acids	Molecular weight (kDa)	pI	Aliphatic index	Grand average of hydropathicity (GRAVY)
PVP	418	46	6.79	91.29	-0.211
KPP	418	46	8.48	84.59	-0.378
ETP	444	48	6.67	80.05	-0.279
HAP	413	45	6.48	82.71	-0.283

Additionally, the InterProScan tool was employed to identify the protein family, homologous superfamily, Pfam I.D., Prosite signature, and catalytic core of PVP, and ETP based on its FASTA sequence. According to the analysis, both PV and ET belong to the Histidine phosphatase superfamily and consist of Histidine acid phosphatase phosphohistidine signature. The details of PVP I.D. are here: Interprofamily I.D.: IPR029033, and clade2 I.D.: IPR000560, Histidine phosphatase domain (CDD: cd07061, Pfam:PF00328). The details of ETP I.Ds are here: Histidine Phosphatase superfamily (IPR029033), Pfam I.D. (PF00328), Prosite I.D. (PS00616), and CDD (cd07061).

In summary, *in-silico* characterization of PVP and ETP helped to gain information related to molecular properties and characteristics. The results are comparable to the known phytases. Hence it strengthens the conclusion drawn from the analysis.

5.2.3.2 Secondary Structure Analysis:

Psipred (**Figure 5. 9**) (**Figure 5. 10**), Polyview2D (**Figure 5. 11**) (**Figure 5. 12**), PDBsum (**Table 5. 2**) (**Figure 5. 13**) (**Figure 5. 14**), Proteus (**Figure 5. 15**) (**Figure 5. 16**), Porter (**Figure 5. 17**) (**Figure 5. 18**), Yet Another Scientific Artificial Reality Application (YASARA) software (**Figure 5. 19**) were used for the secondary structure analysis. Each tool/server/software employed specific algorithms and methodologies. These tools give information related to the percentage of helices, β -sheets, coils, and turns within PVP and ETP. Hence it is utilized to generate a consensus prediction of secondary structures.

Psipred and Proteus2, provided confidence scores for the prediction of each amino acid residue in PVP and ETP. The analysis from each tool provided reliable confidence scores for individual amino acids, enhancing the accuracy and reliability of the secondary structure prediction for ETP and PVP.

Table 5. 2 PDBsum analysis of PVP and ETP.

Parameters	PVP	ETP
Disulfide bonds	3	4
γ turns	4	2
β turns	32	31
helices	18	18
β strands	10	10
β bulges	3	4
psi loops	1	1
β hairpins	4	3
β sheets	2	2

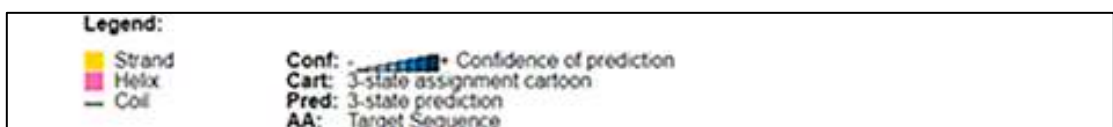
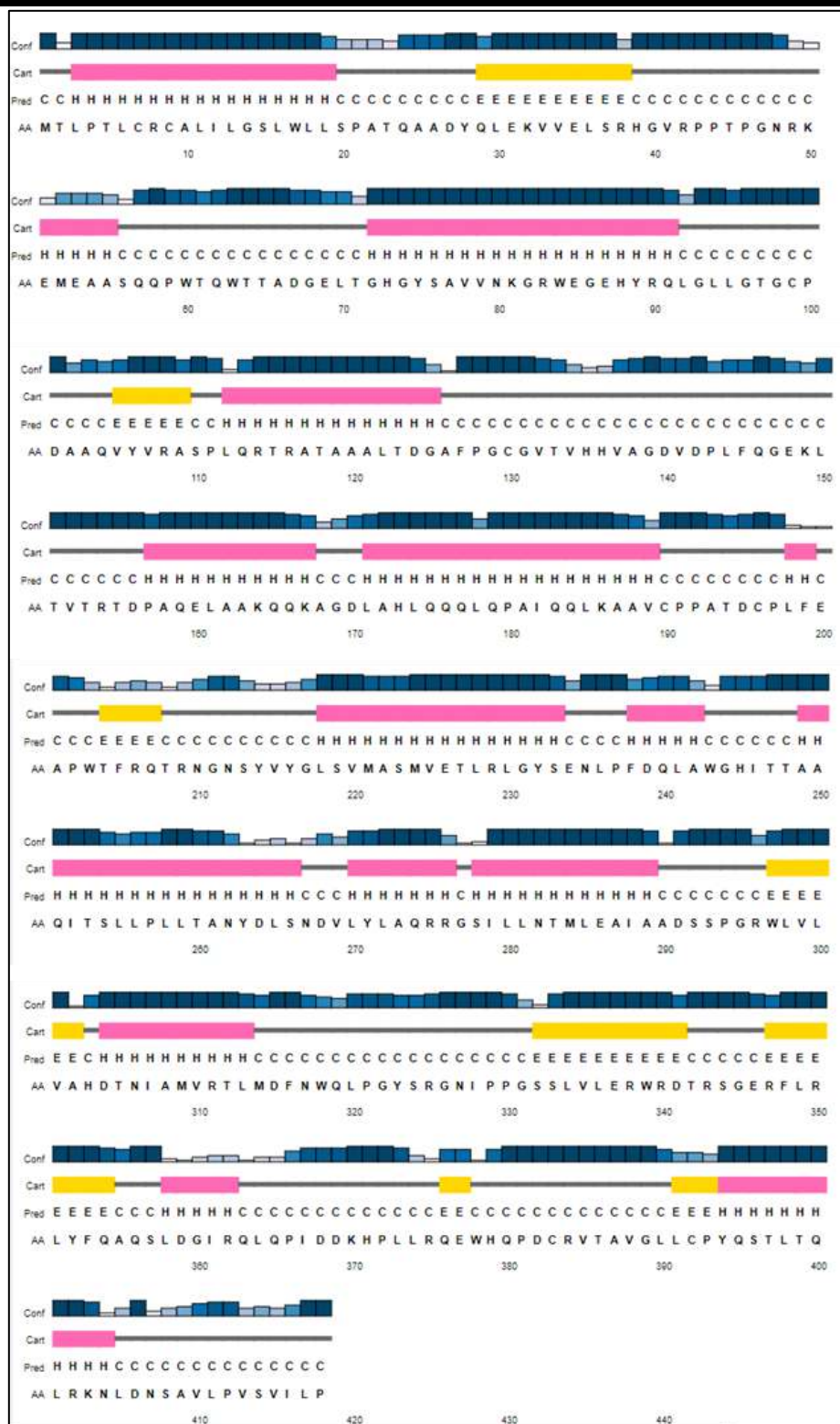
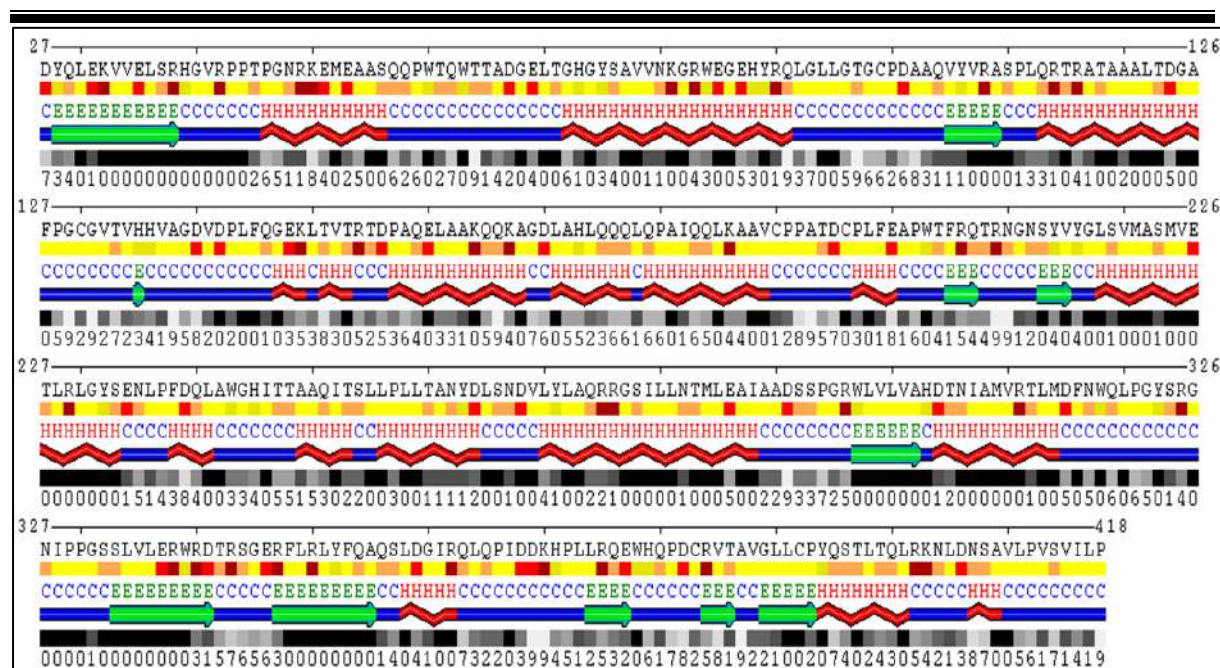
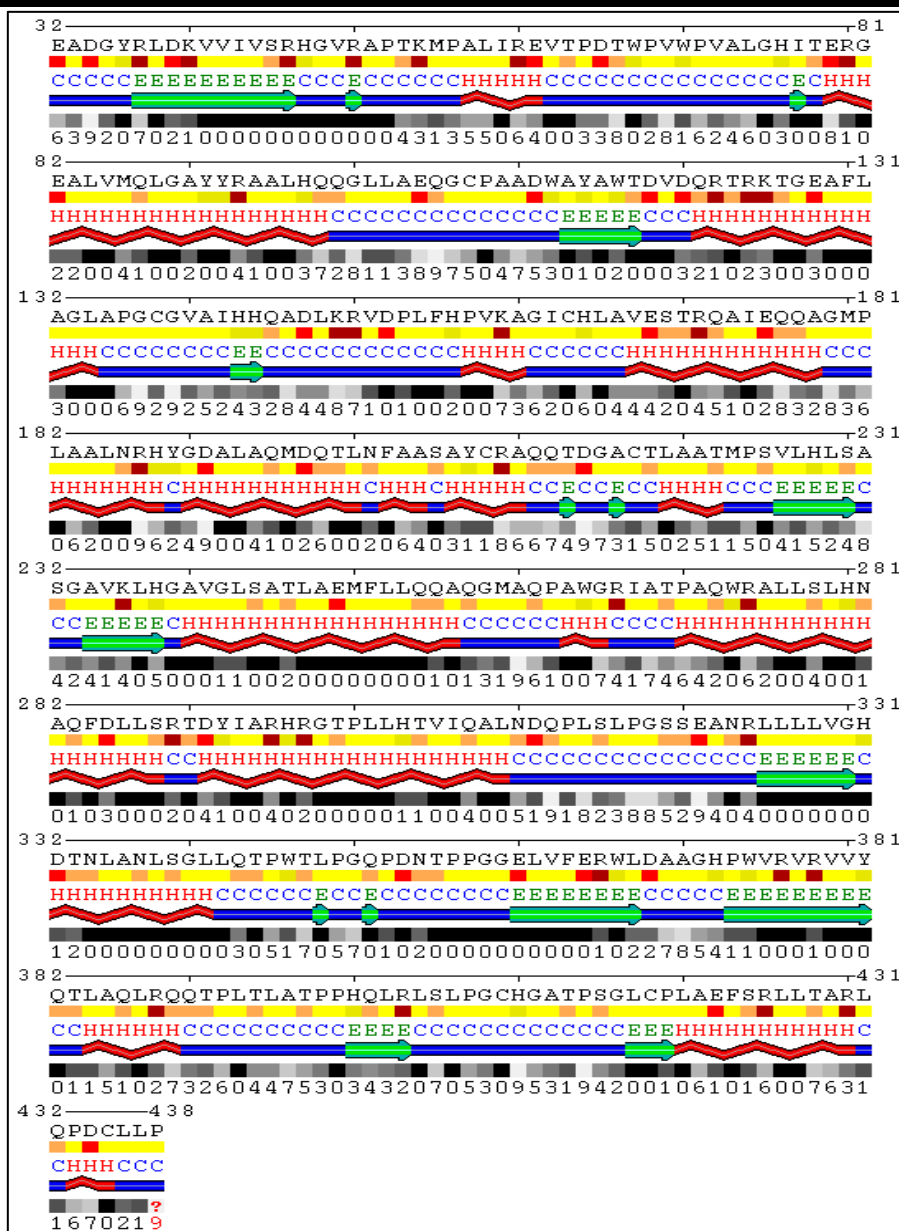


Figure 5. 9 Psipred was used to predict the secondary structure of PVP.



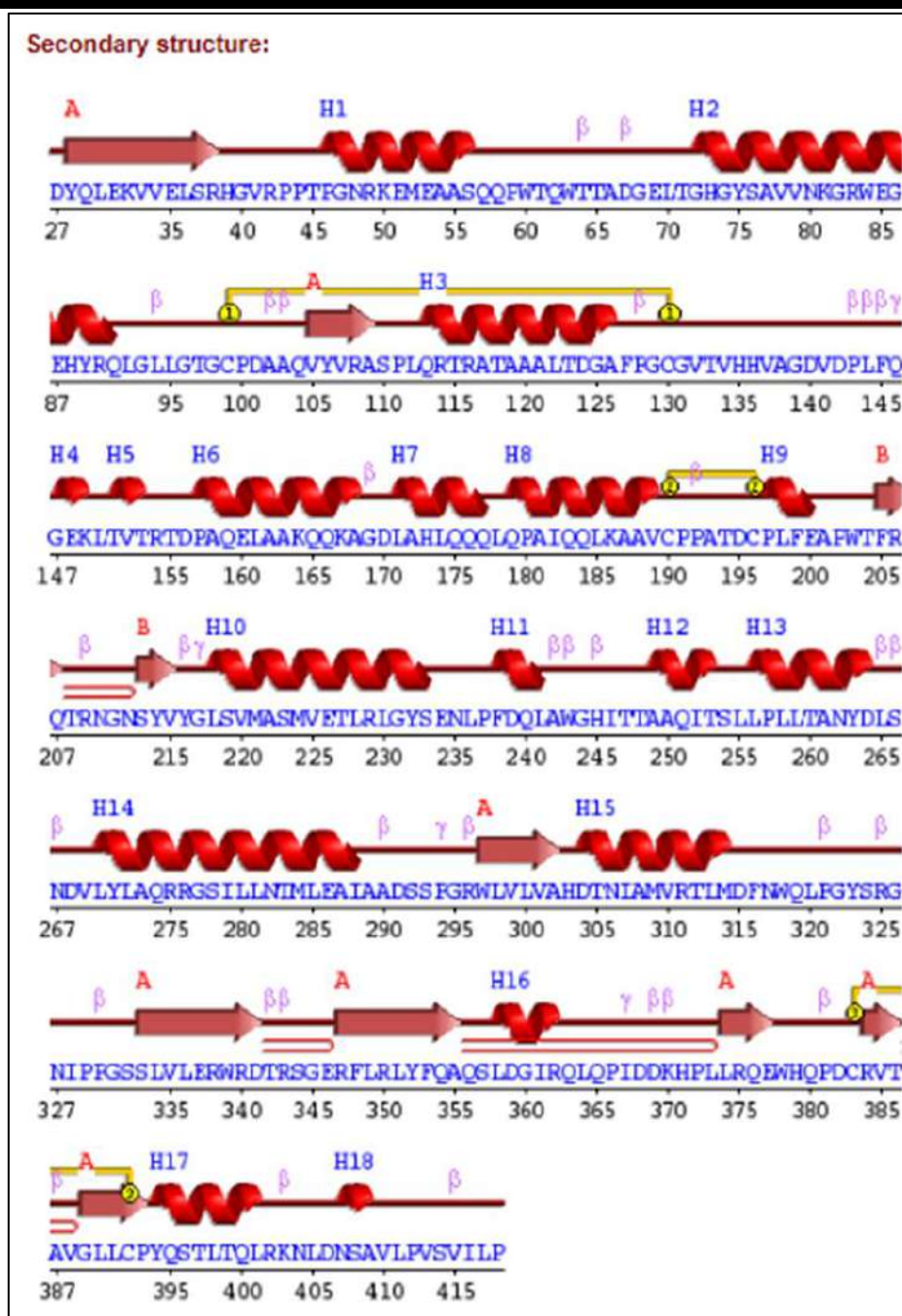
Legend	Description
1 —————>	Amino acid residue numeration
	Protein secondary structure
	H-alpha and other helices (model 1)
	H-alpha and other helices (model 2)
	E-beta-strand or bridge
	C-coil
	Relative solvent accessibility (RSA)
0 1 2 3 4 5 6 7 8 9	0-completely buried (0-9% RSA), 9-fully exposed (90-100% RSA)
	Physical-chemical properties
	H-hydrophobic: A,C,F,G,I,L,M,P,V
	A-amphipathic: H,W,Y
	P-polar: N,Q,S,T
	N/C-charged: D,E-neg; R,K-pos
	Confidence level of prediction
	0-the lowest level, 9-the highest level
GLCFEPFERL	Transmembrane domain

Figure 5. 11 Polyview2D was used to predict the secondary structure of PVP.





Legend	Description	Physical-chemical properties
1 —————>	Amino acid residue numeration	H-hydrophobic: A,C,F,G,I,L,M,P,V
	Protein secondary structure	A-amphipathic: H,W,Y
	H-alpha and other helices (model 1)	P-polar: N,Q,S,T
	H-alpha and other helices (model 2)	N/C-charged: D,E-neg; R,K-pos
	E-beta-strand or bridge	Confidence level of prediction
	C-coil	0-the lowest level,
	Relative solvent accessibility (RSA)	9-the highest level
0 1 2 3 4 5 6 7 8 9	0-completely buried (0-9% RSA),	
	9-fully exposed (90-100% RSA)	GLCFEPFERL Transmembrane domain

Figure 5. 12 Polyview2D was used to predict the secondary structure of ETP.



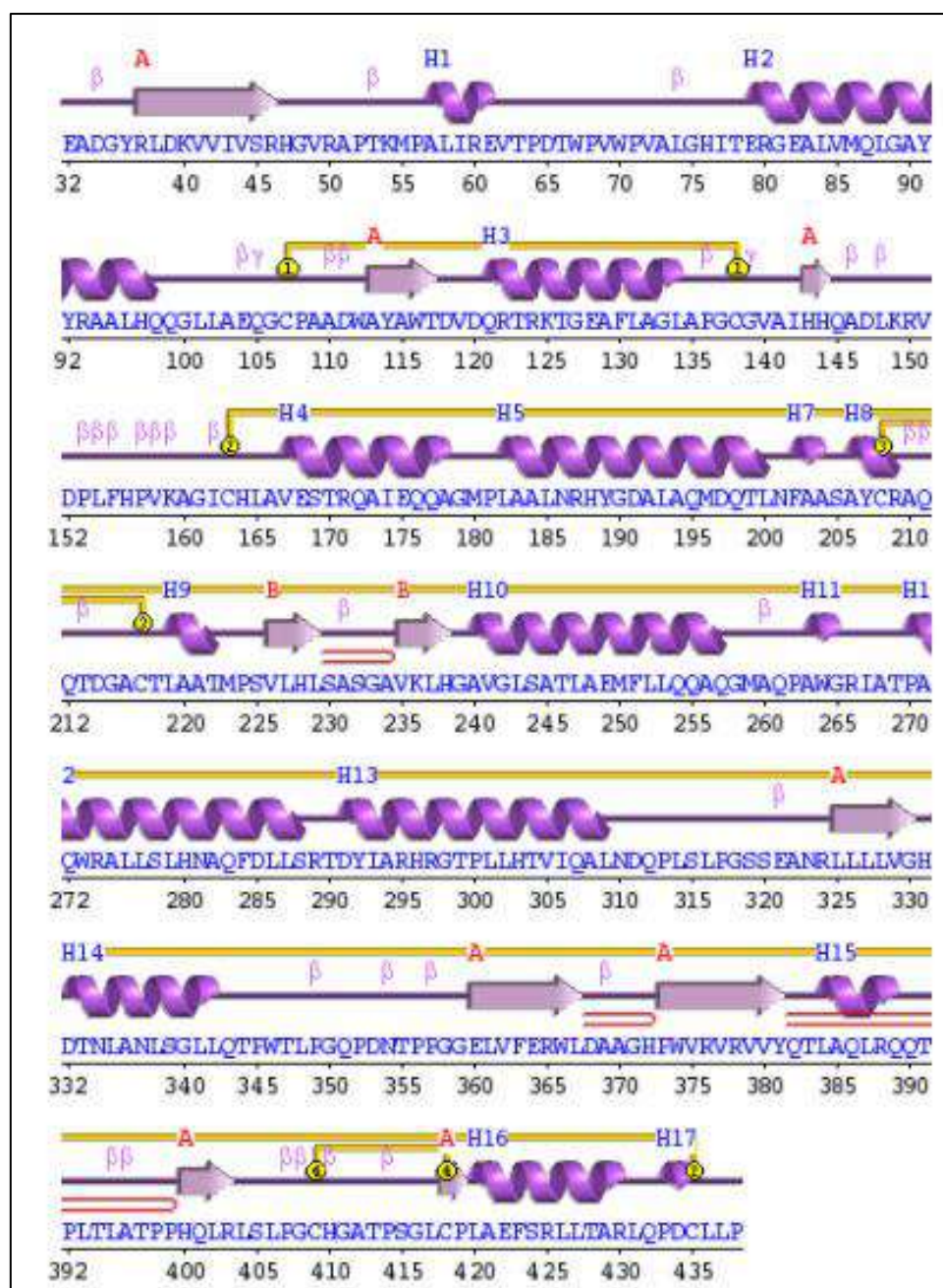
Key:


Sec. struc:  Helices labelled H1, H2, ... and strands by their sheets A, B, ...
 Helix Strand

Motifs: β beta turn γ gamma turn  beta hairpin


Disulphides:  disulphide bond

Figure 5. 13 PDBSum was used to predict the secondary structure of PVP.

**Key:**

Sec. struc:  Helices labelled H1, H2, ... and strands by their sheets A, B, ...

Helix Strand

Motifs: β beta turn γ gamma turn  beta hairpin

Disulphides:  disulphide bond

Figure 5. 14 PDBSum was used to predict the secondary structure of ETP.


```

MTLPTLCRCALILGSLWLLSPATQAADYQLEKWELSRHGVRPPTPGNRKEMEAASQQPW
CCHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHH
EEbbBBBBBBBBBBBBBBBBBebEEEEebeBebBBBBBBBBBBbBBbEEeeEeBEEbbEeEb

TQWTTADGELTGHGYSAVVNKGRWEGEHYRQLGLLGTGCPDAAQVYVRASPLQRTRATAA
CCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC
eEbEbEbBbBBEeBbeBBbbBBebBbebBeEeeBBEeEbbeEbBbBbBbbbbbBbbBBB

ALTDGAFPGCGVTVHHVAGDVDPLFQGEKLTVTRTDPAQELAAKQKAGDLAHLQQQLQP
HHHHCCCCCCCCCECCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCCHHHHHHHHH
BBBBBBBBBebEBebbbbeEEebbBBbBbEbEbEBbbeEeBbebbEEEEEbEbEEBbEeBeE

AIQQLKAAVCPATDCPLFEAPWTRFRQTRNGNSYVYGLSVMASMVETLRLGYSENLPFDQ
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
BBEBBEEbbbeEEeEbeBbEbEbeBebeEEEEebbbBBbBBbbBBBBBBBBBbBbEbEbEe

LAWGHITTAQITSLLPLLTANYDLSNDVLYLAQRSGSILLNTMLEAIAADSSPGRWLVL
CCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHH
bBbbbebEeeEbBeeBbeBbebbBbBbbeBeeBBebeBbeBBebBbEbBeEEEEebbBBBB

VAHDTNIAMVRTLMDFNWQLPGYSRGNIPGSSLVLERWRDTRSGERFLRLYFQAQSLDG
EECHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
BBBBBBBBBBBBBBbBBEBEbEbeEbbEEeBBBBBBBBBBBBbeeeEEeEbbBbBbBBBbBBeb

IRQLQPIDDKHPLLQEWHPDCRVAVGLLCPYQSTLTQLRKNLDNSAVLPVSVILP
HHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCECCHHHHHHHHHHHHHHHHHHHHHHHHHHH
BbEeEEbeeEEbbeEbEBEBEEbEEEEEEebBeBEeBbEbBeEbBeEbBBbBBeebeE

```

Figure 5. 17 Porter was used to predict the secondary structure of PVP.

Line 1: The 1-letter code of protein primary sequence.

Line 2: Secondary structure prediction by Porter 4.0:

H = helix : DSSP's H (alpha helix) + G (3-10 helix) + I (pi-helix) classes.

E = strand : DSSP's E (extended strand) + B (beta-bridge) classes.

C = the rest : DSSP's T (turn) + S (bend) + . (the rest).

Line 3: Relative Solvent Accessibility prediction by PaleAle 4.0:

B=completely buried (0-4% exposed)

b=partly buried (4-25% exposed)

e=partly exposed (25-50% exposed)

E=completely exposed (50+% exposed)


```

MRESRLARWFCLSVVGLGSLFPLGAQATSSPEADGYRLDKVVIVSRHGVRAPTCKMPALIR
CCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
EEEEbebbBbBBBBBBBBBBBBBBBBbbEbeEEEEbeBebBBBBBBBBBBbBbEeEEeBe

EVTPTWVWPVALGHI TERGEALVMQLGAYYRAALHQQLLAEQGC PAADWAYAWTDVD
CCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
EbeEeEbeEbEbebBbBBEeBbeBBbbBBebBbebBeEeEBBeEEebbeeEbBbBBBbbb

QRTRKTGEAFLAGLAPGCGVAIHQADLKRVDPLFHPVKAGICH LAVESTRQAIEQQAGM
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCC
bbBbbBBBBBBBBBBBebEbeBbbbEEbEEebbBBbBbebEbbEbbeEeBeEbBeEebEe

PLAALNRHYGDALAQMDQTLNFAASAYCRAQQT D GACTLAATMPSVLHLSASGAVKLHGA
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCC
ebEeBeEeBeEBbEbBeEBBEbeEbEbeEEEEeEEEEbeBbEeebeBebeEEEEebebEbB

VGLSATLAEMFLLQQAQGM AQP AWGRIATPAQWRALLSLHNAQFDLLSRTDYIARHRGTP
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCC
BeBBbBBBBBBBBBBbBbEbEbbbbbEbeeeEbBeeBbeBbbbbbBbBbbeBeeBBBeBbe

LLHTVIQALNDQPLSLPGSSEANRLLLLVGHDTNLANLSGLLQTPWTLPGQPDNTPPGGE
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCC
BBebBbEbBeEEeEBbEBEebBBBbBBBBBBBBBBBBBBbBbEbEbEbeEbbEeEBBBBBB

LVFERWLD AAGHPWVRVRYQT LAQLRQQTPLTLATPPHQLRLSLPGCHGATPSGLCPL
EEEEEECCCCCEEEEEEEECCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
BBBBbbeEEebBbBbBBBBBbebBbEeeEbeEEEbbeeEBEBEEbEEEEeEBbBebB

AEFSRLLTARLQPDCLLPSATSR
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCC
EeBbEbBeEbbeEEbbbBbEeE
    
```

Figure 5. 18 Porter was used to predict the secondary structure of ETP.

Line 1: The 1-letter code of protein primary sequence.

Line 2: Secondary structure prediction by Porter 4.0:

H = helix : DSSP's H (alpha helix) + G (3-10 helix) + I (pi-helix) classes.

E = strand : DSSP's E (extended strand) + B (beta-bridge) classes.

C = the rest : DSSP's T (turn) + S (bend) + . (the rest).

Line 3: Relative Solvent Accessibility prediction by PaleAle 4.0:

B=completely buried (0-4% exposed)

b=partly buried (4-25% exposed)

e=partly exposed (25-50% exposed)

E=completely exposed (50+% exposed)

According to the analysis by using YASARA software it was found that PVP (**Figure 5. 19**) and ETP (**Figure 5. 19**) consisted of a higher percentage of helices compared to its other secondary structural elements. Similar observations were found in the case of their respective reference sequences i.e., KPP and HAP.

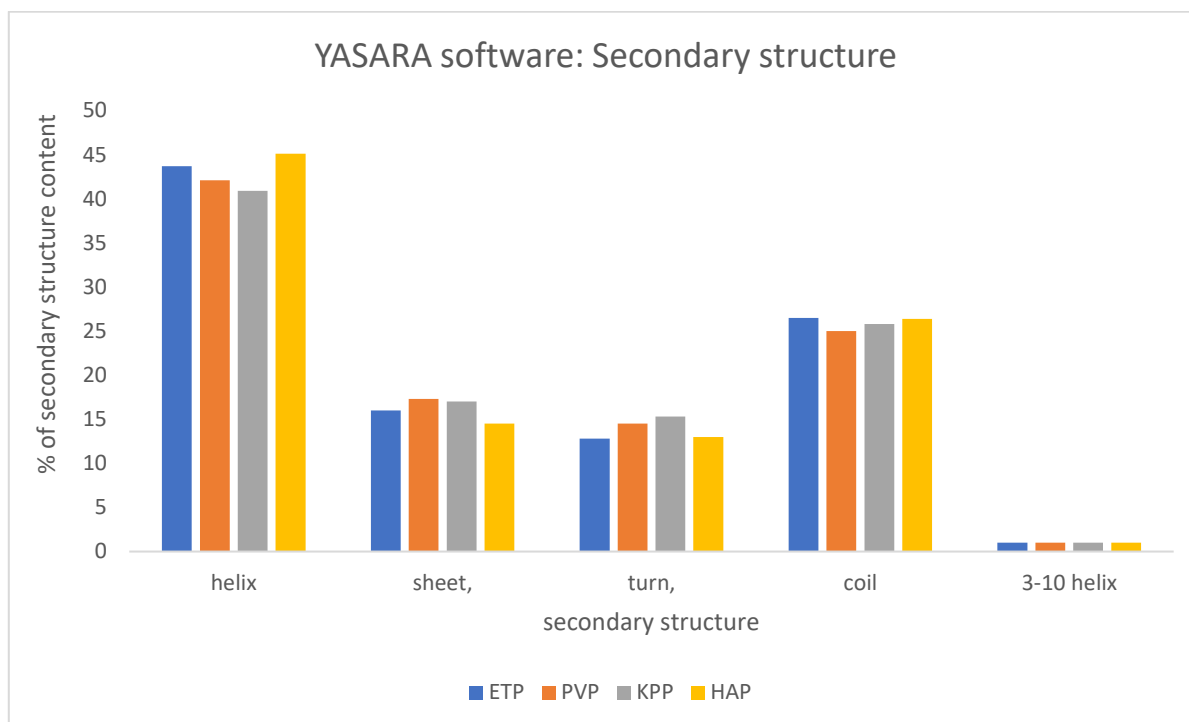


Figure 5. 19 Secondary structure analysis of ETP, PVP, KPP, and HAP by using YASARA software.

In summary, secondary structure analysis encompassing the evaluation of the topology of PVP and ETP revealed a noteworthy similarity and comparable % of secondary structure content between these phytases and the reference sequences.

5.2.3.3 Tertiary structure analysis and validation:

Tertiary structure modeling of PVP and ETP was performed via. Funfold, Hhpred, Phyre, Swiss Model, Fold and Function Assignment System (FFAS), Iterative Threading ASSEMBly Refinement (I-TASSER), Topmodel, and Intfold. The tertiary structure of phytases was examined using these servers, which followed a specific workflow consisting of several fundamental steps. These steps included input sequence, typically in FASTA format, selecting a suitable template, constructing a model, minimizing energy, assessing quality and validity, and visualizing the results.

Chapter 5

The modeling of PVP and ETP involved the use of multiple software/servers/tools, and each model structure was primarily validated using the Structure Assessment tool provided by the EXPASY web portal (<https://swissmodel.expasy.org/assess>). This tool facilitated the evaluation of the Ramachandran plot, which is instrumental in assessing the conformation of backbone dihedral angles (phi and psi) in proteins. The allowed regions on the plot indicate combinations of phi-psi angles that are energetically favorable for proper protein folding. The PVP and ETP structures modeled using various tools exhibited a range of 87.28 to 97.22% and 80 to 97.53% of amino acid residues within the allowed regions respectively (**Table 5. 3**).

Table 5. 3 Preliminary screening of modeled structures by using the Structure Assessment tool provided by the EXPASY web portal.

Sr.No.	Method	% Ramchandran allowed/favored region of	
		PVP	ETP
1	Fold and Function Assignment System (FFAS)	97.22	97.53
2	FunFold	91.9	92.97
3	Hhpred	97.44	96.54
4	IntFold	97.36	96.83
5	Phyre	95.65	95.31
6	SWISS MODEL	96.41	95.8
7	Tr-rosetta	96.88	97.96
8	I-TASSER	87.28	80

Chapter 5

According to the Ramachandran plot, the percentage of residues in the allowed regions in the case of modeled structures PVP and ETP via SWISS-MODEL were 96.41 and 95.8 respectively and that of KPP and HAP were 98.74 and 99.49 respectively. This indicated that the structure was reliable as the % of residues in the allowed regions of PVP and ETP were like their references i.e., KPP and HAP respectively.

The modeled structures of PVP and ETP, along with reference sequences (KPP and HAP) were further validated by using ProSa, ERRAT (%), and QmeanDisco (**Table 5. 4**). The ProSaZ scores of PVP and ETP were -8.92 and -8.77 respectively, while those of references KPP and HAP were -9.18, and -9.2 respectively. The accepted range of %ERRAT was >50. The %ERRAT value of PVP and ETP was 94.24 and 90.95 respectively, whereas that of KPP, and HAP was 96.124 and 95.58 respectively. QMEANDisco score of PVP, ETP, KPP, and HAP was 0.88 ± 0.05 , 0.84 ± 0.05 , 0.93 ± 0.05 , and 0.94 ± 0.05 respectively. The ProSa Z-score values, ERRAT (%), and QMEANDisco score were close to their references which indicated that the structures were reliable.

The structures of PVP (**Figure 5. 20**) and ETP (**Figure 5. 21**) modeled by using SWISS-MODEL were for molecular dynamic simulation studies because SWISS-MODEL is a well-known and widely used freely available server for modeling structure via homology modeling. Reliable models were constructed in less time.

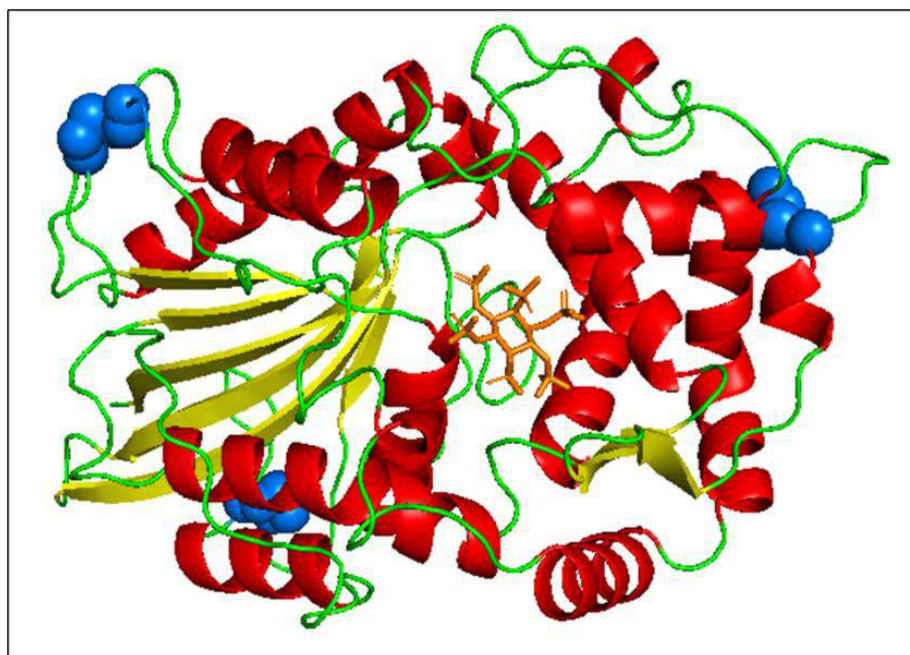


Figure 5. 20 The solid ribbon representation of PVP docked with phytate (brown): red α -helix, yellow β -sheet, blue disulfide bonds, and green loops.

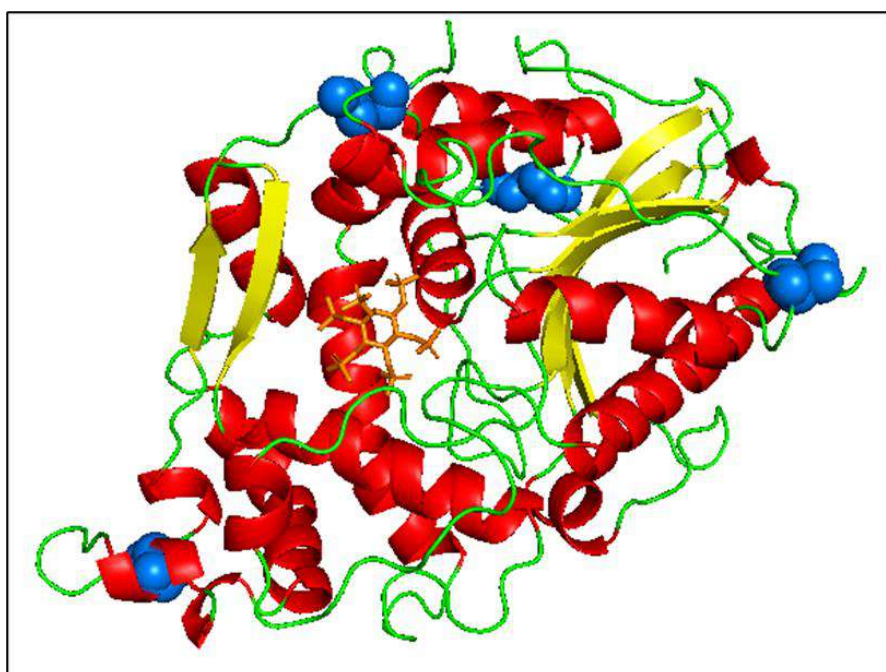


Figure 5. 21 The solid ribbon representation of ETP docked with phytate (brown): red α -helix, yellow β -sheet, blue disulfide bonds, and green loops.

Table 5. 4 Tertiary structure validation of shortlisted Swiss model and template structures by using ProSa, ERRAT, and QmeanDisco.

SWISS-MODEL structure	Model name	Structure Assessment (%)	ProSa Z-score	ERRATA (%)	QmeanDisco
Model	PVP	96.41	-8.92	94.241	0.88 ± 0.05
	ETP	95.8	-8.77	90.956	0.84 ± 0.05
Template	KPP (2WNH)	98.74	-9.18	96.124	0.93 ± 0.05
	HAP (4ARO)	99.49	-9.2	95.58	0.94 ± 0.05

The superposition of PVP (**Figure 5. 22**) and ETP (**Figure 5. 23**) on their respective references i.e., KPP and HAP revealed structural similarities between them though the sequence identity was less than 70% in both cases. PVP and ETP both possessed α and $\alpha\beta$ domains like their reference structures.

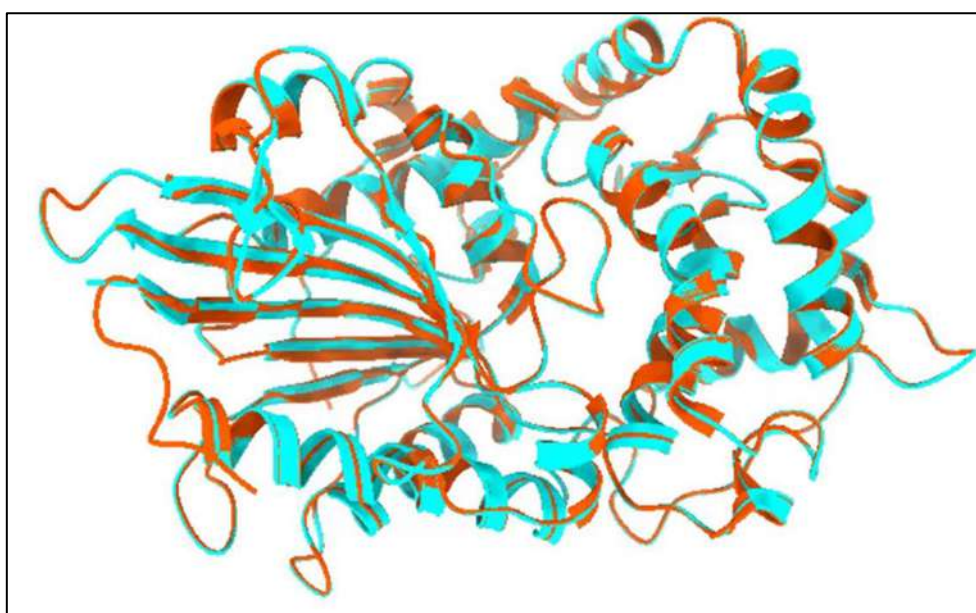


Figure 5. 22 Superimposition of the modeled tertiary structure of PVP (cyan) and KPP (brown).

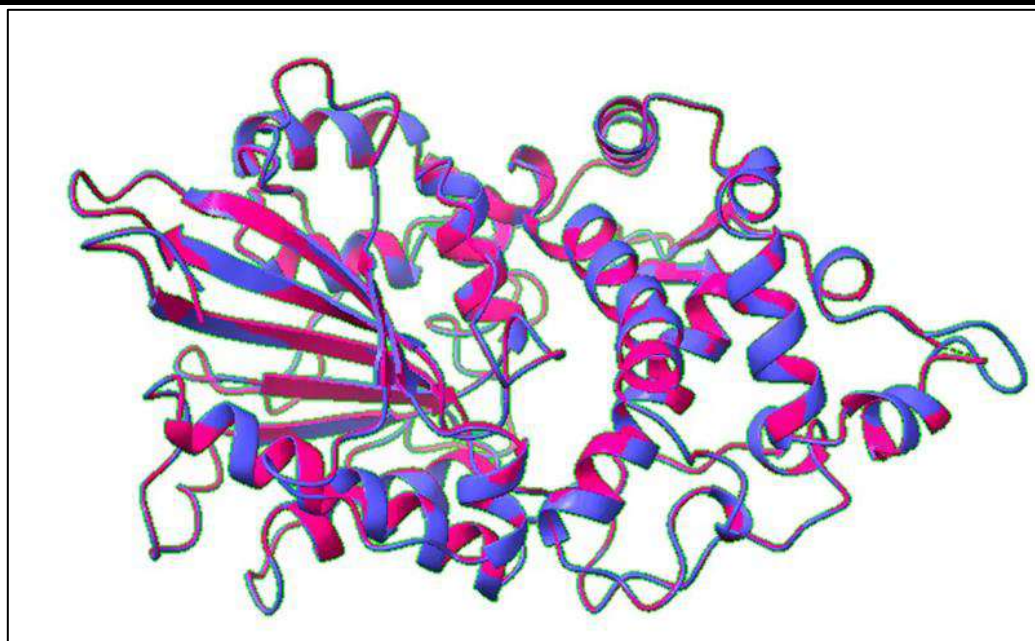


Figure 5. 23 Superimposition of the modeled tertiary structure of ETP (Blue) and HAP (Pink).

In summary, the tertiary structure analysis revealed that the SWISS-MODEL structures of PVP and ETP were reliable as both were comparable to their reference structures. Hence, PVP and ETP structures can be used further for molecular dynamics studies.

5.2.3.4 Functional analysis:

The conserved catalytic core present in PVP was R38, H39, R42, R114, H303, and D304. The catalytic core present in ETP was R46, H47, G48, P52, H331 and D332. The catalytic site was the same as the known HAP phytases such as *E.coli* (Lim et al., 2000) and *Klebsiella* phytases (Böhm et al., 2010a). This indicated that PVP and ETP have the potential to hydrolyze phytate via two-step mechanisms. The first step involved the formation of a phosphohistidine intermediate via nucleophilic attack on the phosphorous of phytate by the N-terminal conserved histidine. This intermediate was hydrolyzed by the action of C-terminal conserved aspartic acid (proton donor to the oxygen atom of the scissile phosphomonoester bond). This results in the release of lower myoinositols, inorganic phosphate, and bound minerals (Lim et al., 2000) (Böhm et al., 2010a).

According to Castp analysis, the total number of pockets identified in-case of PVP and ETP was 53 and 68 respectively. The first pocket of PVP occupied was 744.160 Å² area and 700.187 Å³ Volume (**Table 5. 5**) (**Figure 5. 24**). The first pocket of ETP occupied was 681.749 Å² Area and Volume 673.731 Å³ (**Figure 5. 25**) (**Table 5. 6**). The motif ‘LdkVviVsRHGvRaP’ and

LDKVVIVSRHGVRAP were present in PVP and ETP according to ScanProsite, and MotifFinder respectively. According to the analysis via VMD, the total number of salt bridges present in PVP and ETP was 17, and 11 respectively.

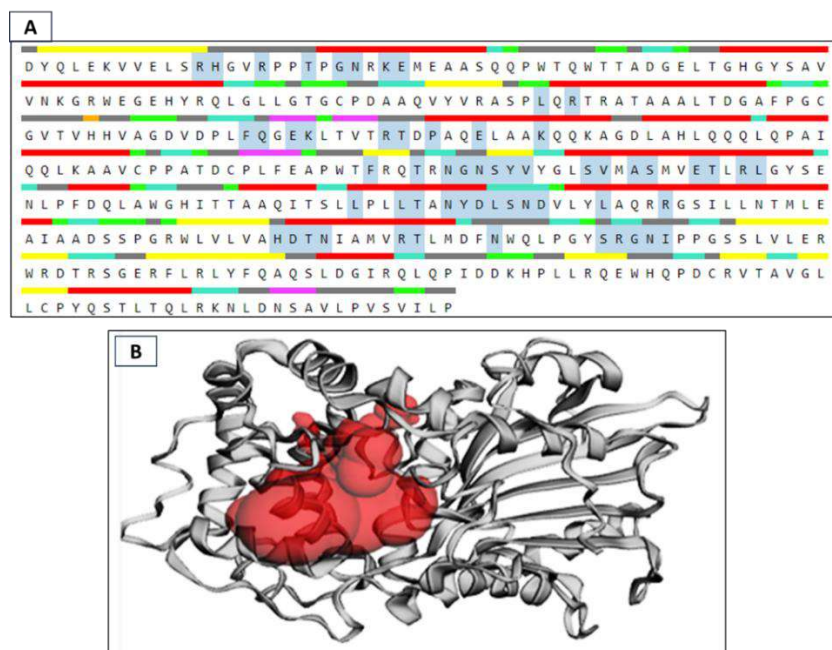


Figure 5. 24 The Castp analysis of PVP structure: A. Grey highlighted residues involved in catalysis of phytate, B. tertiary structure of PVP (grey), and highlighted catalytic pocket (red).

Table 5. 5 Residues present in the 1st pocket and predicted to be involved in catalysis according to the Castp analysis of PVP structure.

SeqID	AA	SeqID	AA	SeqID	AA	SeqID	AA
38	ARG	155	THR	223	SER	272	LEU
39	HIS	157	PRO	226	GLU	276	ARG
42	ARG	160	GLU	227	THR	303	HIS
45	THR	164	LYS	229	ARG	304	ASP
47	GLY	205	PHE	230	LEU	305	THR
48	ASN	208	THR	256	LEU	306	ASN
50	LYS	210	ASN	259	LEU	311	ARG
51	GLU	211	GLY	260	THR	312	THR
112	LEU	212	ASN	262	ASN	317	ASN
114	ARG	213	SER	263	TYR	324	SER
145	PHE	214	TYR	264	ASP	325	ARG
146	GLN	215	VAL	265	LEU	326	GLY
148	GLU	219	SER	266	SER	327	ASN
149	LYS	220	VAL	267	ASN	328	ILE
154	ARG	222	ALA	268	ASP		

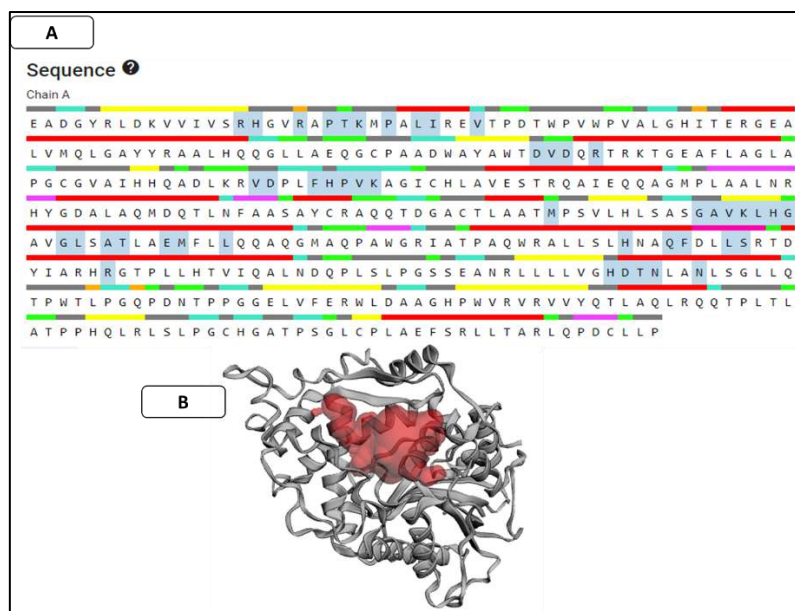


Figure 5. 25 A. Displayed the sequence of ETP: grey highlighted residues that were involved in catalysis according to Castp analysis. B. The tertiary structure of ETP: Highlighted in Red is the predicted pocket.

Table 5. 6 Residues present in the 1st pocket and predicted to be involved in catalysis according to the Castp analysis of ETP structure.

SeqID	AA	SeqID	AA	SeqID	AA
46	ARG	152	ASP	245	ALA
47	HIS	155	PHE	246	THR
50	ARG	156	HIS	249	GLU
52	PRO	157	PRO	250	MET
53	THR	158	VAL	253	LEU
54	LYS	159	LYS	280	HIS
56	PRO	223	MET	283	GLN
58	LEU	234	ALA	284	PHE
59	ILE	235	VAL	287	LEU
62	VAL	236	LYS	288	SER
118	ASP	237	LEU	297	ARG
119	VAL	238	HIS	331	HIS
120	ASP	239	GLY	332	ASP
122	ARG	242	GLY	333	THR
151	VAL	243	LEU	334	ASN
				337	ASN

Chapter 5

The SignalP analysis revealed that PVP (**Figure 5. 26**) and ETP (**Figure 5. 27**) also possessed signal peptides. The cleavage site presented in PVP was between A25 and A26. The cleavage site presented in ETP was between A27 and T28. This suggested that ETP and PVP had the potential for further investigation regarding their structural and functional properties, as they both possessed high solubility and could be produced in large quantities.

The total number of disulfide bonds present in PVP and ETP were 3 and 4 respectively. According to ClustalO and superimposition of PVP with KPP and ETP with HAP, the conserved cysteines in PVP were: C99, C130, C190, C196, C383, C392, and in ETP were C107, C143, C163, C217, C208, C417, and C443. According to Pymol and PIC analysis, the disulfide bonds present in PVP were 190C-196C, 383C-392C, 99C-130C, and that in ETP were C409-C418, C163-C435, C107-C138, and C208-C217.

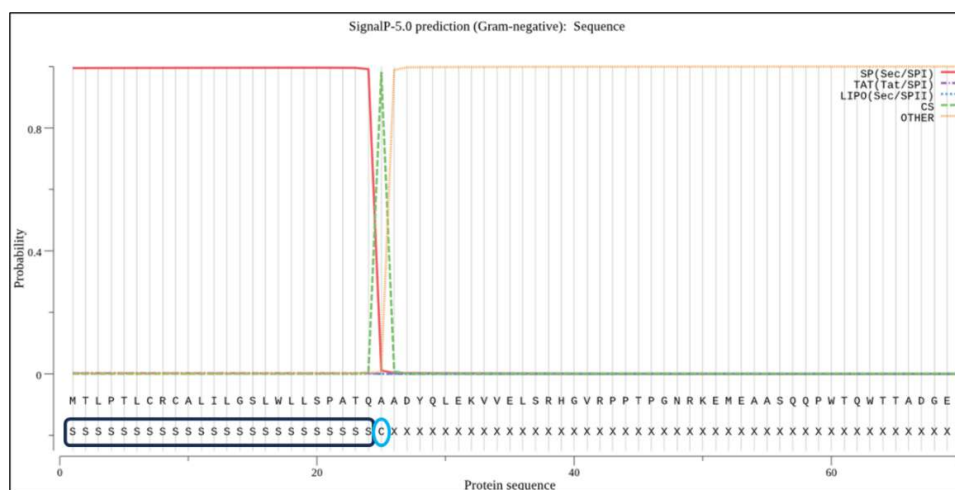


Figure 5. 26 SignalP analysis of PVP.

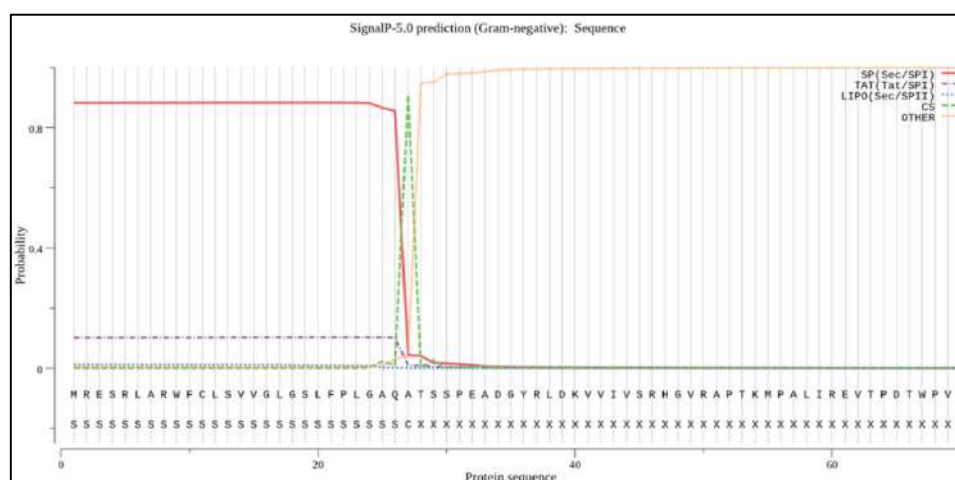


Figure 5. 27 SignalP analysis of ETP.

Chapter 5

According to the SCooP analysis of PVP (Figure 5. 28) and ETP (Figure 5. 29), the melting temperature (T_m) was predicted to be 65°C and 63.4°C respectively.

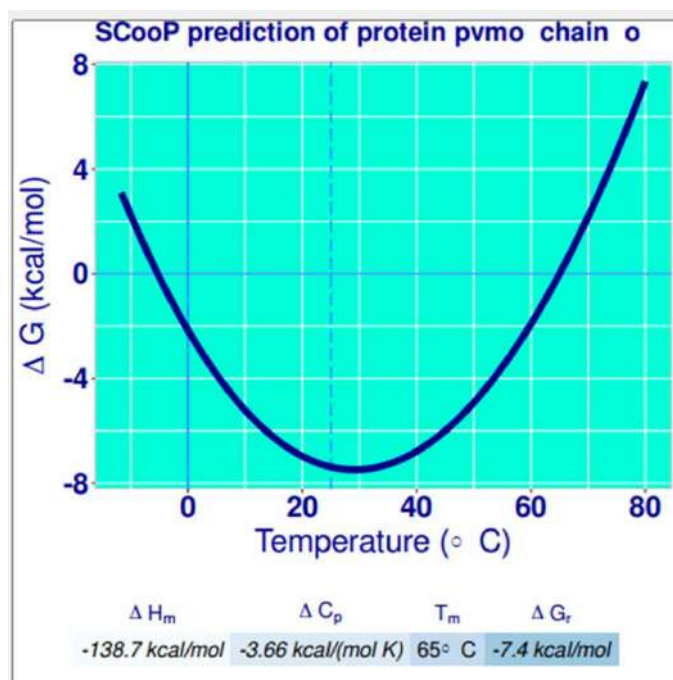


Figure 5. 28 SCooP analysis of PVP

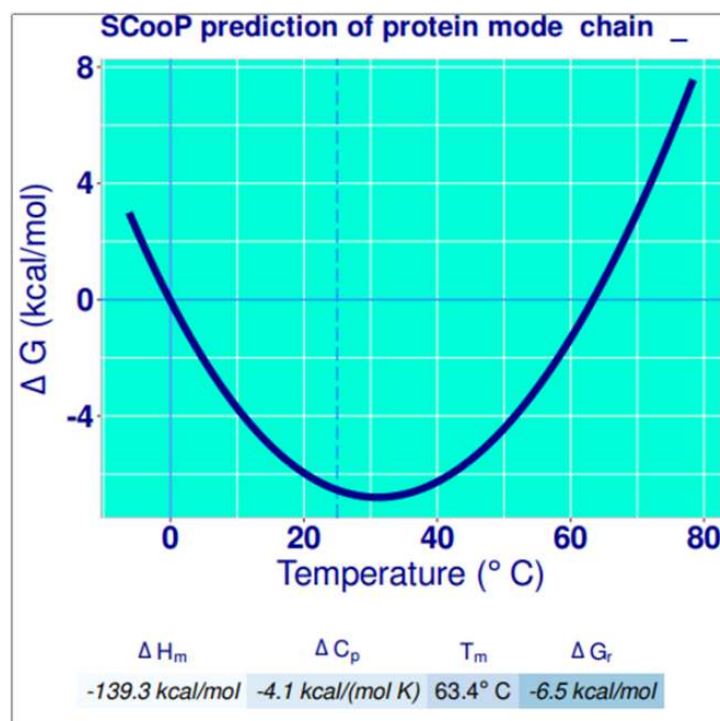


Figure 5. 29 SCooP analysis of ETP.

To summarize, the functional analysis of PVP and ETP provided valuable information regarding several aspects. It identified conserved residues involved in catalysis and assessed solubility. This also provided information related to the potential application of PVP and ETP.

5.2.3.5 Molecular docking and Molecular dynamics simulation:

Molecular Docking: Autodock 4 was used to investigate the binding of phytate. The calculated binding affinity of PVP (**Figure 5. 30**) and ETP (**Figure 5. 31**) was -9.1 and -7.7 kcal/mol respectively. ETP and PVP both exhibited the presence of conserved N terminal (RHGXRX) and C terminal (HD). These conserved residues were predicted to interact with phytate via salt bridges, attractive charges, hydrogen bonds, carbon-hydrogen bonds, and Pi-Anion interactions. There were other residues that helped to enhance the binding of phytate to the catalytic pocket of ETP and PVP. In-case of PVP, the predicted residues involved in catalysis were: R38, H39, R42, N48, R114, S266, N267, R276, and H303. In-case of ETP, the predicted residues involved in catalysis were: R46, H47, R50, R122, H156, H331, R297, T333, S288, Q283, K236, K54, and F284.

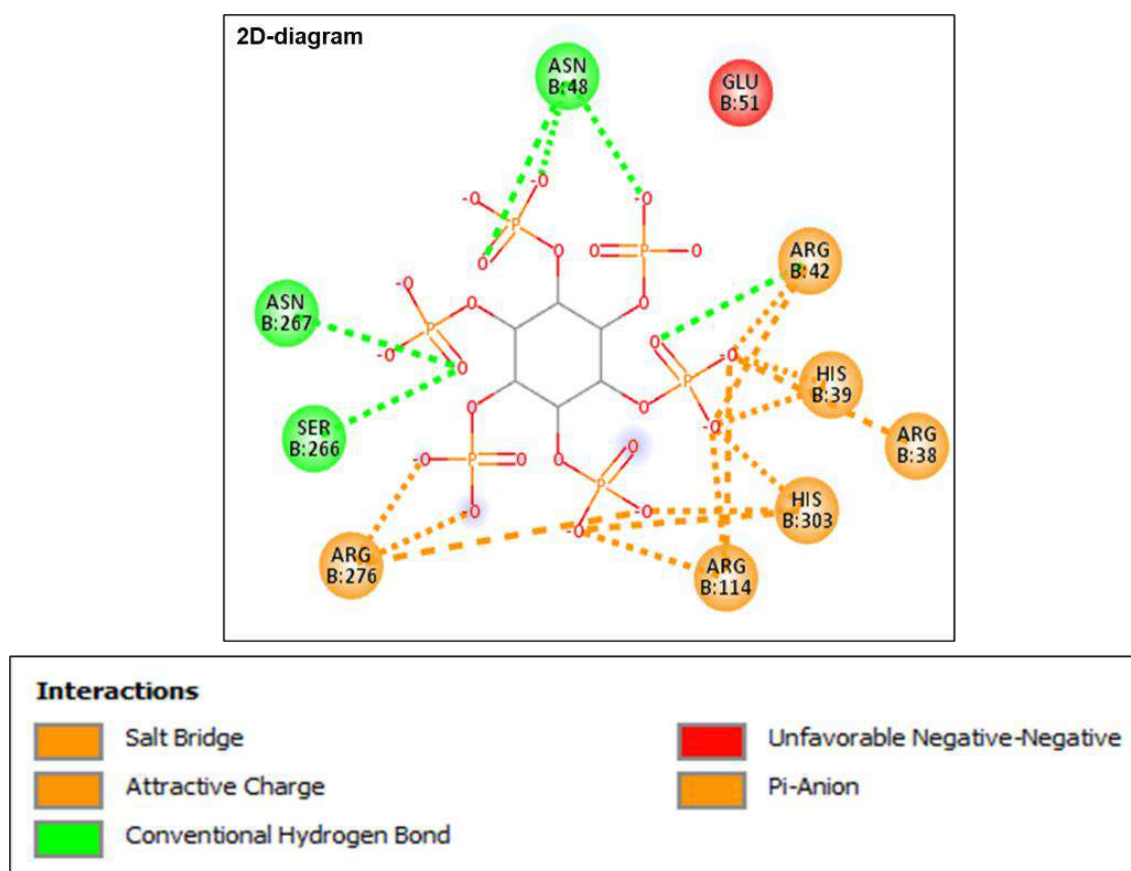


Figure 5. 30 BioVia Discovery studio used to generate the 2D structure of the PVP phytate complex to identify interactions.

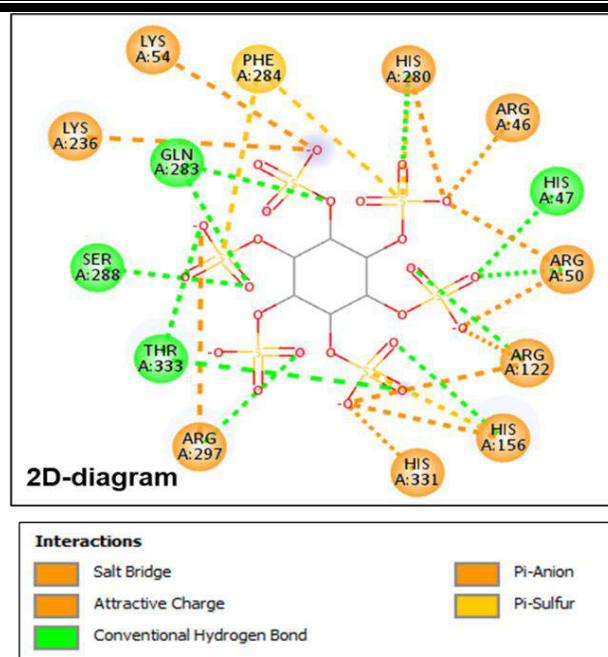


Figure 5. 31 BioVia Discovery studio used to generate the 2D structure of ETP phytate complex to identify interactions.

In summary, the conserved C-terminal, and N-terminal residues, along with other residues interacted with phytate to enable stable binding to the catalytic pocket. The docked structures were subjected to molecular dynamics simulation studies at different temperatures to understand the stability of PVP and ETP.

Molecular dynamics simulation:

The docked PVP and ETP structures were further subjected to molecular dynamic simulation at temperatures 300, 350, 400, and 450K for 30ns using GROMACS software (<https://www.gromacs.org/>). The analysis performed consisted of Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Solvent Accessible Surface Area (SASA), and Radius of Gyration (Rg)

RMSD: RMSD analysis reveals the overall conformational dynamics of the protein. RMSD of the backbone was calculated. The average RMSD of both PVP (**Figure 5. 32**) and ETP (**Figure 5. 32**) was found to increase as the temperature increased from 300K to 450K. This was due to increased thermal energy which consequently increased molecular motions of PVP and ETP. The average RMSD of PVP at 300, 350, 400, and 450K was 0.120, 0.150, 0.219, and 0.415nm respectively (**Figure 5. 32**). The average RMSD of ETP at 300, 350, 400, and 450K was 0.16, 0.22, 0.31, and 0.45nm respectively (Figure 5.32). However, there was an increase in the

average backbone RMSD of ETP as compared to PVP. This indicated that the molecular motions caused due to thermal energy were less in PVP as compared to ETP at temperatures 350, 400, and 450K. In summary, RMSD analysis revealed that the overall conformational changes in PVP were less than that of ETP. This indicated that PVP was more stable than ETP.

RMSF: The analysis of root-mean-square fluctuation (RMSF) provided valuable insights into the fluctuations and flexibility of PVP and ETP at the residue level, particularly under different thermal conditions. RMSF was measured w.r.t backbone. RMSF of both PVP and ETP increased due to the increase in the temperature of the system (**Figure 5. 33**). Residues exhibiting RMSF values above 0.2 nm were identified as highly flexible regions, indicating increased fluctuation. On the other hand, residues with RMSF values below 0.2 nm were considered stable regions, suggesting reduced flexibility. The average RMSF backbone of PVP was 0.074, 0.097, 0.129, and 0.240nm at 300,350,400, and 450K respectively (**Figure 5. 33**). The average RMSF backbone of ETP was 0.082, 0.116, 0.160, and 0.287nm respectively at 300, 350, 400, and 450K (**Figure 5. 33**). The average backbone RMSF of PVP was less compared to ETP which suggested that PVP was more stable than ETP at higher temperatures especially at 400 and 450K. In summary, this analysis provides a detailed understanding of the dynamic behavior of ETP and PVP, highlighting regions of high and low flexibility in the protein structure.

Rg: The radius of gyration (Rg) is a parameter that reflects the compactness of a protein structure and provides insights into its conformational flexibility and folding/unfolding behavior. Rg was measured for the backbone. Rg of PVP and ETP structures increased as the temperature of the system increased which indicated conformational changes in the PVP and ETP structure (**Figure 5. 34**). The average backbone Rg of PVP at 300, 350, 400, and 450K were 2.12, 2.13, 2.13, and 2.15nm respectively (**Figure 5. 34**). The average backbone Rg of ETP at 300, 350, 400, and 450K were 2.187, 2.200, 2.214, and 2.216 respectively (**Figure 5. 34**). However, the average backbone Rg of PV was less than that of ET which indicated that the structure of PVP is stable at all temperatures as compared to ETP (**Figure 5. 34**).

SASA: The solvent-accessible surface area (SASA) represents the total surface area of a protein that is accessible to the surrounding solvent molecules. Variations in protein surface exposure, influenced by factors such as temperature, provide insights into conformational changes, unfolding/folding, and overall protein dynamics. SASA of PVP and ETP increased as the temperature increased which indicated that the structure was unfolding at higher

Chapter 5

temperatures (**Figure 5. 35**). The average SASA of PVP at 300, 350, 400, and 450K were 171, 173, 174, and 183 nm² respectively. The average SASA of ETP at 300, 350, 400, and 450K were 181.1, 181, 183, and 192.5nm² respectively (**Figure 5. 35**). However, the average SASA of PVP structure was less than that of ETP which indicated less conformational changes in PVP compared to ETP (**Figure 5. 35**). This suggested that PVP was more stable than ETP. In summary, molecular dynamics simulation analyses such as RMSD, RMSF, Rg, and SASA suggested that PVP structure was stable compared to ETP at different temperatures.

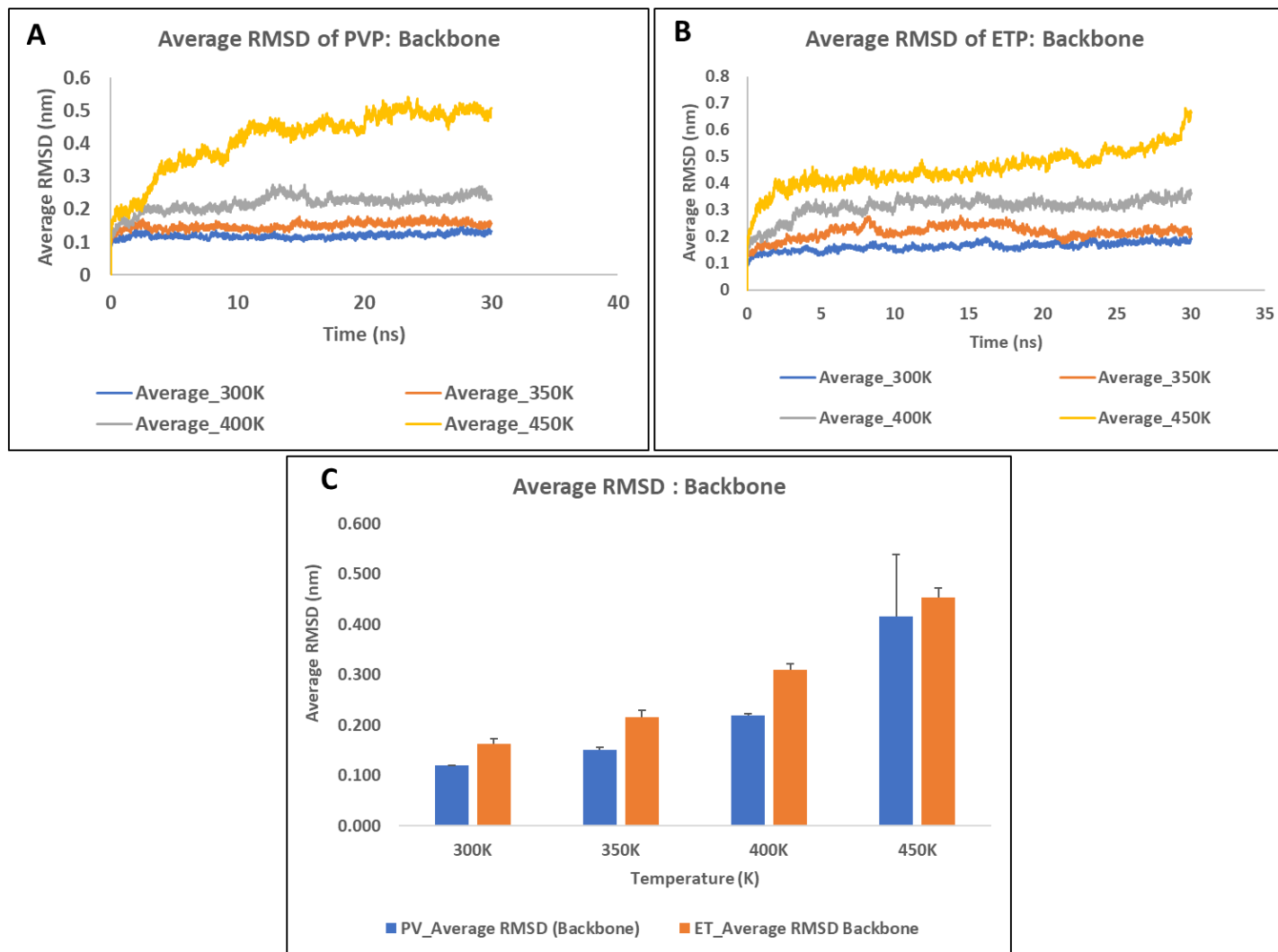


Figure 5. 32 Average RMSD analysis of A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP RMSD values.

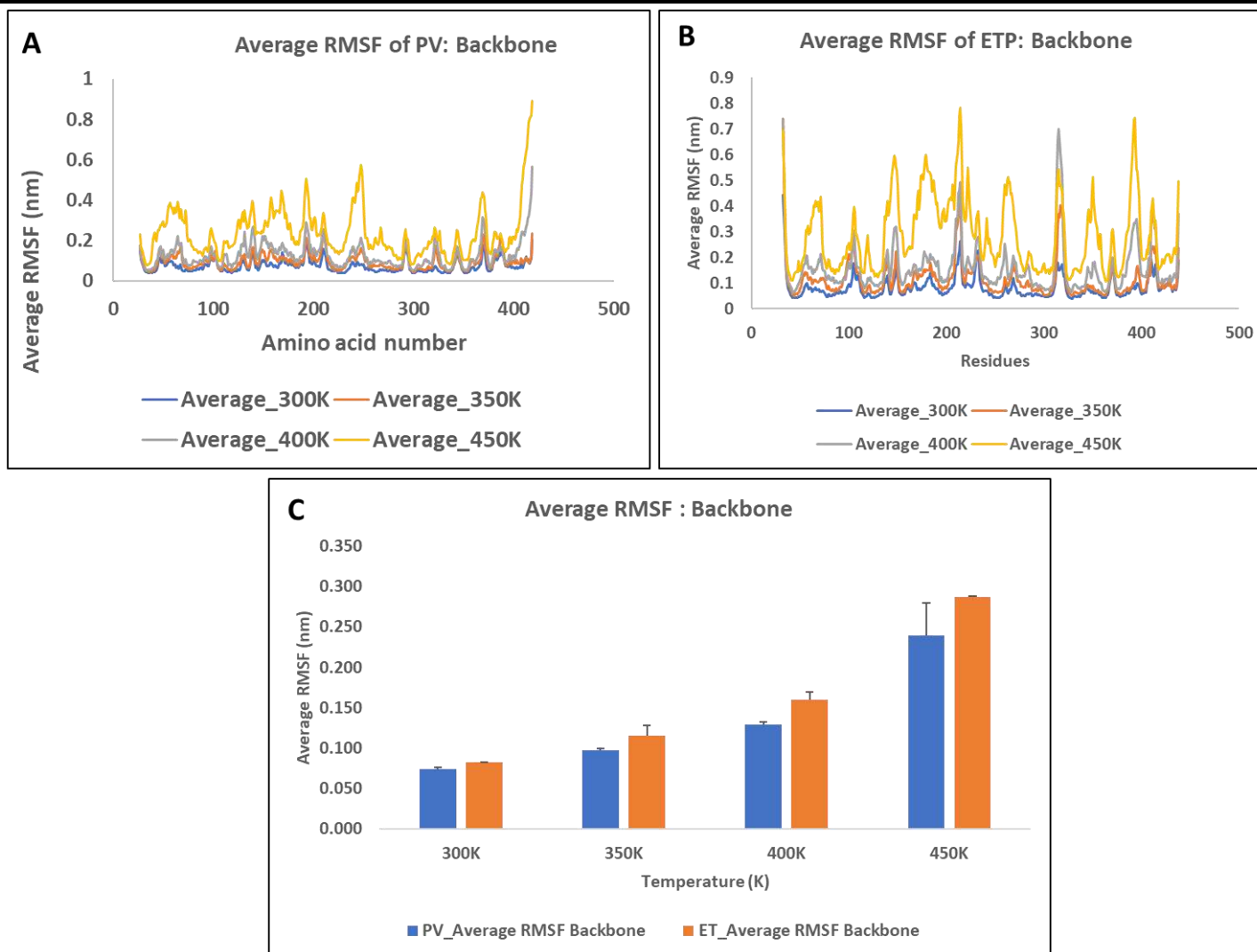


Figure 5.33 Average RMSF of A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP

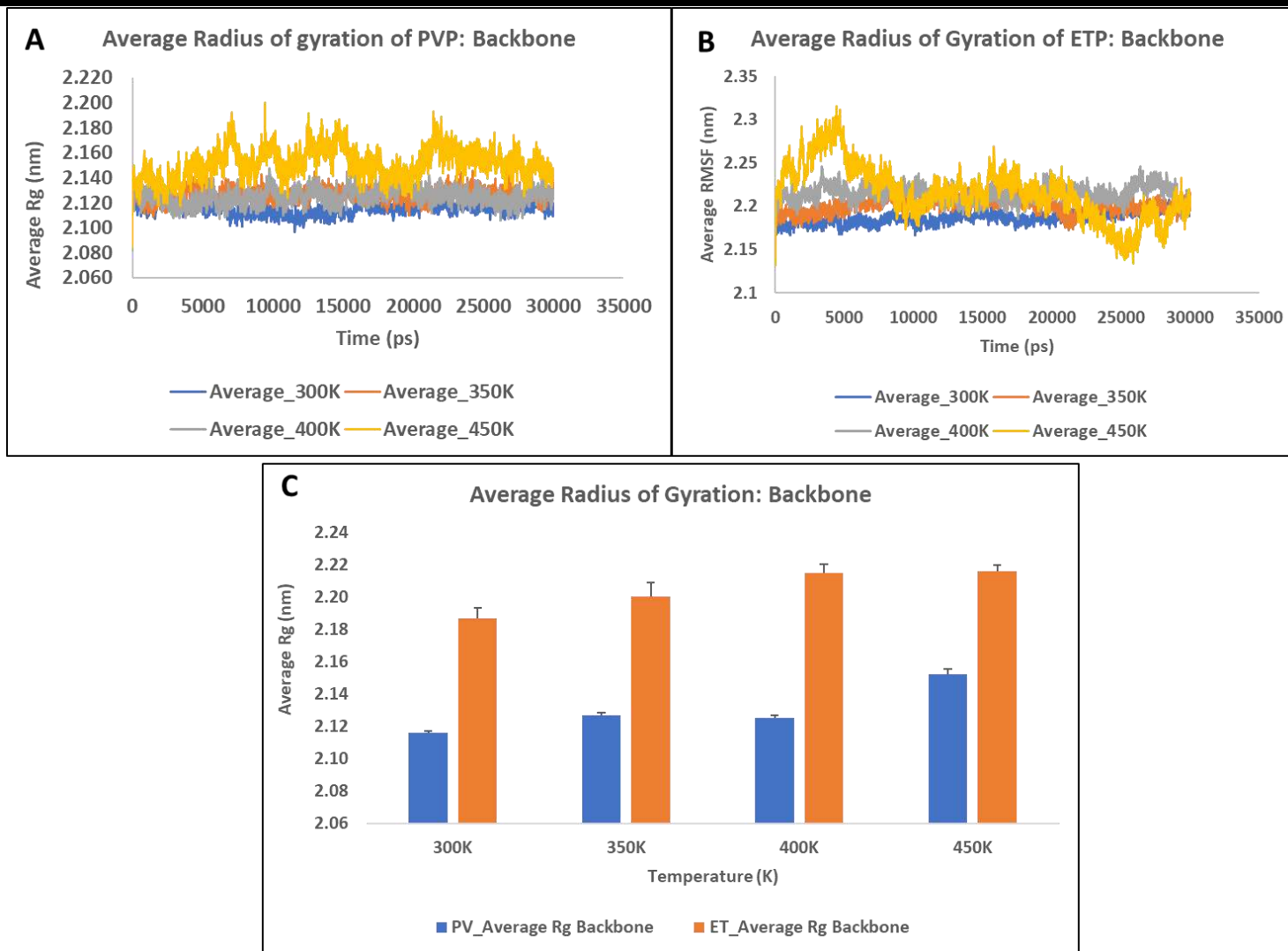


Figure 5.34 Average Rg of A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP

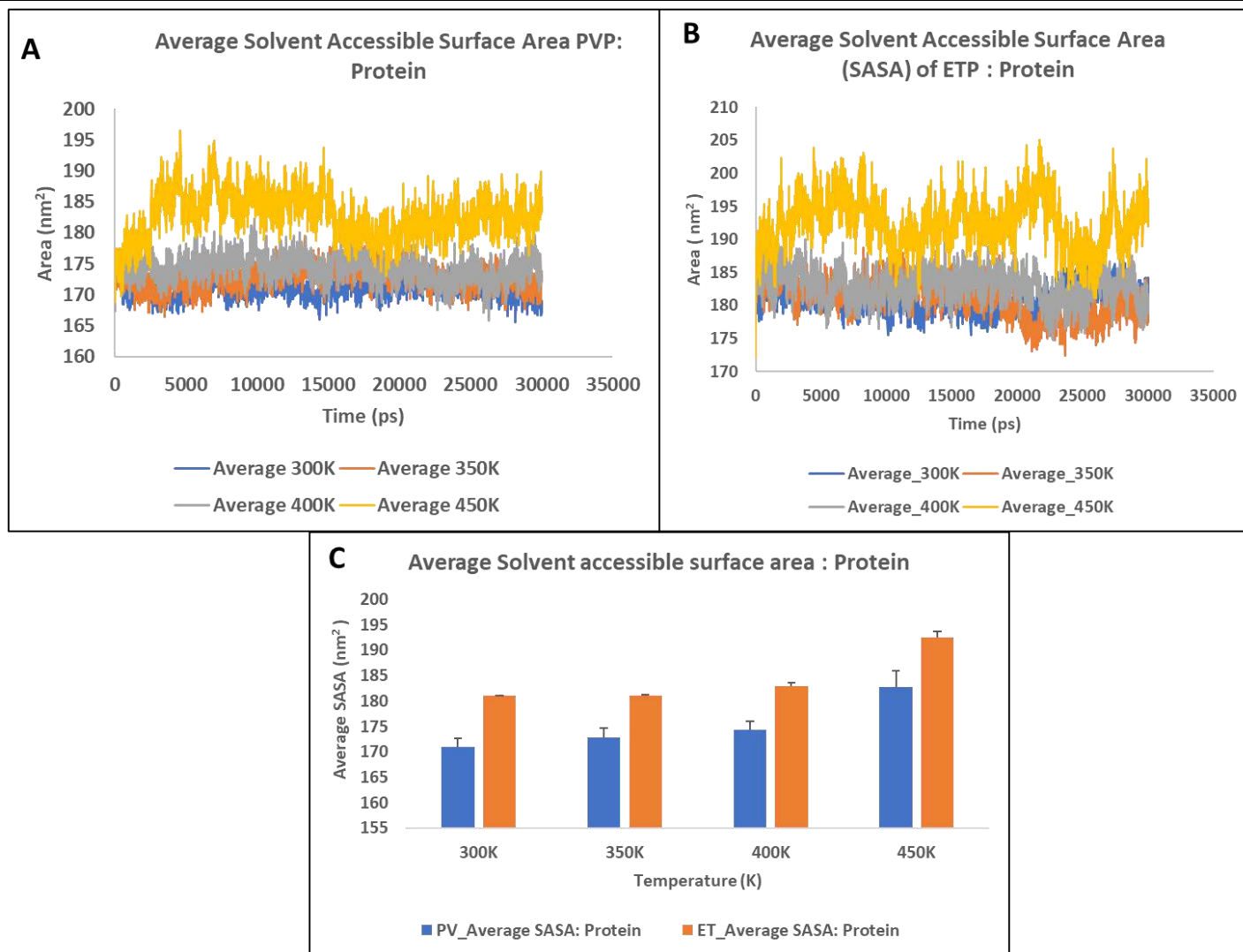


Figure 5.35 Average SASA of: A. PVP backbone, B. ETP backbone, C. bar graph: comparison of PVP vs ETP

5.3 Experimental validation:

5.3.1 Introduction:

Experimental validation experiments are performed along with *in-silico* characterization. Experimental validation includes confirmation about phytase gene in *P.vagans* and *E.tarda*, cloning of phytase gene from *P.vagans* and *E.tarda*, followed by over-expression in *E.coli* BL21(DE3) system, and biochemical characterization (Figure 5. 36).

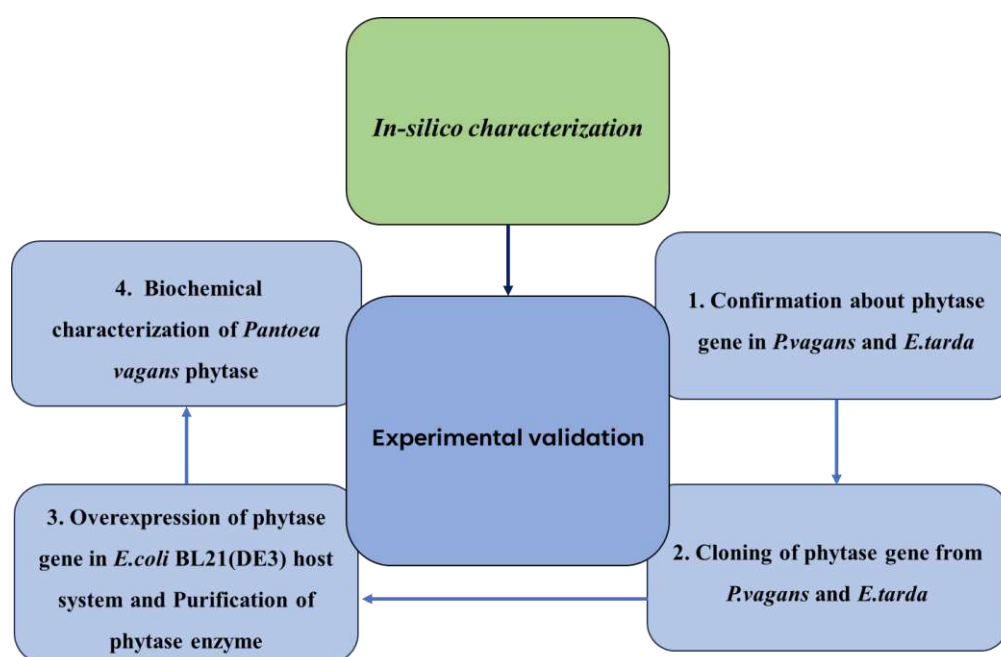


Figure 5. 36 Overview of the experimental validation.

5.3.2 Materials and Methods:

5.3.2.1 Materials (Chemicals, kits, reagents, solutions, plasticware):

PureLink® Genomic DNA kit (Cat. No. K182001), Primers (Eurofins), 2X PCR Emerald mix, NucleoSpin® Gel and PCR Clean-up kit, GeneJET Plasmid Miniprep kit, Thermo Fisher (Catalog: K0503), GeneAll® Expin™ PCR SV kit, T4 DNA ligase (Thermo Scientific), 10X T4 DNA ligase buffer (Thermo Scientific, Catalog number: B69), Luria Bertin media (HiMedia), Isopropyl β-D-Thiogalactopyranoside (IPTG) (MP Biomedicals), Pierce™ Disposable Columns (Catalog number: 29922). All chemicals were procured from either Sigma-Aldrich, MP Biomedicals, Loba Chemie, or MolyChem. All plasticware utilized in the experiments, including 15ml and 50ml falcon tubes, were obtained from SPL Life Sciences

Pvt. Ltd., India. NcoI (10U/μl) Thermo Scientific (Catalog number: ER0571), NotI (10U/μl) Thermo Scientific (Catalog number: ER0591).

5.3.2.2 Confirmation of phytase gene in *P.vagans* and *E.tarda*:

The gDNA from *P.vagans* and *E.tarda* was isolated by using the PureLink® Genomic DNA kit (Cat. No. K182001). The integrity of gDNA was assessed by electrophoresis on 1% agarose gel. These extracted gDNA were used as a template to amplify its phytase gene. *P.vagans* consisted of a *phyk* gene that encodes for phytase, whereas *E.tarda* consisted of an *appA* gene that encodes for phytase.

To design primers, the nucleotide sequence of *P.vagans* (EMBL: ADO08555.1) and *E.tarda* (Accession I.D.: PEH73109) was retrieved from EMBL embedded in the UniprotKB site. The extracted gDNA from *P.vagans* and *E.tarda* was used to amplify the full length phytase gene by using primers as mentioned in **Table 5. 7**.

Table 5. 7 Primers used to amplify the full-length phytase gene of *P.vagans* and *E.tarda*.

Full length	Forward primer for <i>P.vagans phyk</i> gene amplification: FP_Pvagansc91_fulllength	5'- ATG ACG CTG CCA ACC CTG TG -3'
	Reverse primer for <i>P.vagans phyk</i> gene amplification : RP_Pvagansc91_fulllength	5'- GGG CAA TAT AAC CGA AAC CGG CA -3'
	Forward primer for <i>E.tarda appA</i> gene amplification: FP_E.tarda_fulllength	5'- ATG CGT GAG AGT AGG TTA GCT CGG T -3'
	Reverse primer for <i>E.tarda appA</i> gene amplification: RP_E.tarda_fulllength	5'-GCG CGA TGT CGC GCT GGG TAA CA-3'

Chapter 5

The PCR conditions and reaction mixture are mentioned in **Table 5. 8** and **Table 5. 9** below.

Table 5. 8 PCR reaction mixture to amplify phytase gene.

Components	Volume/ reaction
Template*	1 μ l
Forward primer (as mentioned in Table 5. 7)	1 μ l
Reverse primer (as mentioned in Table 5. 7)	1 μ l
2X PCR Emerald mix	10 μ l
Sterile distilled water	7 μ l
Total reaction volume	20 μ l

Template*: gDNA of *P.vagans* and *E.tarda*.

Note: Non-template control (NTC) was also included in the process. In the case of an NTC reaction mixture included all the above components except the template. Distilled water (1 μ l) was added instead of the template in the NTC reaction mixture.

Table 5. 9 Phytase gene PCR amplification conditions.

Steps	Temperature ($^{\circ}$ C)	Time
Initial denaturation	95	3 min
Denaturation	95	45 sec
Annealing	For <i>P.vagans phyk</i> gene amplification: 54, 56, 58, 60, 62 64.	30 sec
	For <i>E.tarda appA</i> gene amplification: 57, 58, 59, 62, 64, 66	
Initial extension	72	1 min 30 sec
Final extension	72	10min
Store	4	infinite

} 30 cycles

Chapter 5

The PCR products were electrophoresed in a 0.8% agarose gel with 1X TAE as the buffer medium, at 100 volts against 1 kbp DNA ladder. The PCR product was purified using the gel extraction method via NucleoSpin® Gel and PCR Clean-up kit. The resultant amplicons of the *P.vagans phyk* and *E.tarda appA* gene were also subjected to Sanger's sequencing with forward primer. The Sanger's sequenced *P.vagans* full-length phytase gene was aligned with the nucleotide sequence of *P.vagans* (ENA database) to find the presence of conserved N-terminal 'RHGVRPP' sequence. The presence of conserved residues confirmed the presence of the phytase gene in *P.vagans* and *E.tarda*.

5.3.2.3 Cloning of phytase gene from *P.vagans* and *E.tarda*:

Amplification of insert with flanked restriction sites sequence at both ends:

The above amplicons were used as a template to re-amplify phytase gene from *P.vagans* and *E.tarda* with restriction sites (3'NcoI and 5'NotI) flanked at both sides by using primers mentioned in **Table 5. 10** below:

Table 5. 10 Primers used to amplify the phytase gene of *P.vagans* and *E.tarda* with flanked restriction sites at both ends.

<i>P.vagans phyk</i> gene amplification with flanked restriction sites	Forward primer: FP_PvagansNco1	5'- ATT CCA TGG CGA ATG ACG CTG CCA ACC-3'
	Reverse primer: RP_PvagansNot1	5'- ATT GCG GCC GCA ATG GGC AAT ATA ACC GA -3'
<i>E.tarda appA</i> gene amplification with flanked restriction sites	Forward primer: FP Nco1_ <i>E.tarda</i>	5'- ATT CCA TGG CGA ATG CGT GAG AGT AGG -3'
	Reverse primer: RP with Not1_ <i>E.tarda</i>	5'- ATTGCGGCCGCAATGCGC GATGTCGCGCT-3'

Chapter 5

The PCR conditions and reaction mixture are mentioned in **Table 5. 11** and **Table 5. 12** below.

Table 5. 11 PCR reaction mixture to amplify phytase gene with flanked restriction sites at both ends.

Components	Volume/ reaction
Template*	1 μ l
Forward primer (as mentioned in the Table 5. 10)	1 μ l
Reverse primer (as mentioned in the Table 5. 10)	1 μ l
2X PCR Emerald mix	10 μ l
Sterile distilled water	7 μ l
Total reaction volume	20 μ l

Template:* The full-length PCR amplicon of *P.vagans* and *E.tarda* phytase gene.

Note: Non-template control (NTC) was also included in the process. In the case of NTC reaction mixture included all the above components except the template. Distilled water (1 μ l) was added instead of the template in the NTC reaction mixture.

Table 5. 12 Phytase gene PCR amplification conditions:

Steps	Temperature (°C)	Time
Initial denaturation	95	3 min
Denaturation	95	45 sec
Annealing	For <i>P.vagans phyk</i> gene amplification: 58, 60, 62, 64, 66.	30 sec
	For <i>E.tarda appA</i> gene amplification: 56, 58, 60, 62, 64, 66	
Initial extension	72	1 min 30 sec
Final extension	72	10min
Store	4	infinite

The PCR products (insert) were electrophoresed in a 0.8% agarose gel with 1X TAE as the buffer medium, at 100 volts against a 1 kbp DNA ladder. The PCR product was purified using the gel extraction method via NucleoSpin® Gel and PCR Clean-up kit.

Isolation of plasmid for cloning and expression studies:

E.coli DH5 α containing pET29-b was inoculated in LB-Kanamycin broth. This culture was used to isolate pET29-b by using the GeneJET Plasmid Miniprep kit, Thermo Fisher (Catalog: K0503). The volume of reagents was optimized for a higher yield of the pET29b vector.

In brief, 5ml of *E.coli* DH5 α containing pET29-b culture was processed. The pellet was suspended in 400 μ l resuspension solution (RNase A was added to the solution), followed by the addition of lysis solution. The mixture was gently mixed by inverting 4 times. The tubes were incubated for 5min at RT (do not exceed more than 5 min). To neutralize the mixture, 560 μ l of Neutralization buffer was added and the mixture was mixed gently. It was centrifuged

Chapter 5

at 13,000 rpm for 5 minutes at 4°C. The supernatant was removed gently and loaded on the spin column without the cloudy precipitate. The column was centrifuged at 11,000 rpm for 1 minute at room temperature. The flowthrough in the collection tube was discarded, and 700 µL pf washing buffer was added. The column was centrifuged at 11,000 rpm for 30 seconds at room temperature. This step was repeated twice. The flowthrough in the collection tube was discarded, and the spin column was centrifuged at 14,000 rpm for 1 minute at room temperature. The spin column was transferred in a fresh Eppendorf tube and 60 µL of elution buffer was added. The tube was incubated for 1 minute at room temperature. The tube was centrifuged at 14,000 rpm for 2 minutes at room temperature. The isolated pET29-b were electrophoresed in a 0.8% agarose gel with 1X TAE as the buffer medium, at 100 volts against 1 kbp DNA ladder.

Restriction Digestion of pET29b and insert:

Restriction digestion was performed as mentioned in **Table 5. 13**:

Table 5. 13 Restriction digestion of insert and vector (pET29b).

Reagent	Insert*	pET29b
Sample	40µl	40µl
Buffer O	5µl	5µl
NcoI (10U/µl)	4µl	4µl
NotI (10U/µl)	1µl	1µl
Total Volume	50µl	50µl

Note: Insert: phyk gene flanked with NcoI and NotI, and appA gene flanked with NcoI and NotI.

The above mixture was incubated for 4 hours at 37°C in a water bath. It was stored at -20°C.

The restricted products were purified by using GeneAll® Expin™ PCR SV kit. The purified restricted digested products were used for ligation.

Chapter 5

Ligation of purified restricted digested insert and pET29b:

The ligation reaction of purified restricted digested and insert (Test) was set up along with the following controls:

PC: Plasmid control; IC: Insert control; LC: Ligase control.

The ligation reaction was setup as mentioned in **Table 5. 14**.

Table 5. 14 Ligation reaction setup:

Reagent	Test	PC	IC	LC
Insert* (μl)	2	2	-	2
Purified restricted digested pET29b (μl)	3	-	3	3
T4 DNA ligase (μl)	0.5	0.5	0.5	-
10x T4 DNA ligase buffer (μl)	1	1	1	1
Nuclease-free water (μl)	3.5	6.5	5.5	4.5
Total volume (μl)	10			

Note: Insert: purified restricted digested phyk and appA*

The above reaction was incubated at 16°C for 16 hrs and stored at -20°C.

Competent cells preparation:

A single colony of *E.coli* DH5 α was inoculated in LB-kanamycin broth and incubated at 37°C, 180rpm, 16-18hrs (primary culture). The primary culture (1:100) was inoculated in the fresh 100ml LB-Kanamycin broth (Secondary culture). The O.D._{600nm} of the culture should be 0.04-0.05. The secondary culture was incubated at 37°C, 180rpm till the O.D._{600nm} reached up to 0.3-0.4. It was then incubated in ice for 10min. It was transferred to a prechilled 50ml sterile falcon and centrifuged at 6000rpm, 5min, 4°C. The supernatant was discarded, and the pellet was resuspended in total 4ml of chilled 0.1M CaCl₂ (In case of 100ml secondary culture, i.e., in each 50ml sterile falcon containing pellet: 2ml of the chilled 0.1M CaCl₂ was added). The resuspended pellet was incubated in ice for 30 minutes (gentle tap mix in between). It was then centrifuged at 6000rpm, 5min, 4°C. The supernatant was discarded, and the pellet was resuspended in a total 2ml chilled freezing mixture (0.1M CaCl₂ +15% Glycerol). The mixture

Chapter 5

was aliquoted (100µl) in each prechilled Eppendorf. The competent cells were spread-plated on LB-kanamycin and LB plates. It was stored at -80°C for long-term storage.

Note: all the reagents, Eppendorf, and falcons used in the competent cell preparation should be chilled.

Transformation:

Competent cells (100µl) were thawed in ice for 5-10min. The ligated products (Test) and controls (PC, IC, and LC) (5µl) each were added in the different competent cell's aliquots. The mixture was gently tapped and mixed. It was incubated in ice for 30 mins. *Note: Do not vortex competent cells in the process.* The heat-shock procedure was conducted for efficient transformation. After 30 minutes of incubation, the mixture (Competent cells + ligated products) was subjected to 42°C for 90 seconds. It was then immediately transferred onto ice and incubated for 5-6min. For the recovery of the cells, 900 µL of LB broth (making up the volume to 1000 µL) was added in each tube and incubated at 37°C for 1 hour at 180 rpm in an incubator shaker. After the recovery step, 100 µL from each tube was spread plated onto the LB-Kanamycin agar plates. The remaining 900 µL mixture was centrifuged at 10000 rpm for 2 minutes. 800 µL of the supernatant was removed from each tube. The pellet was resuspended in the remaining 100 µL of the LB broth and 100 µL of this resuspended mixture from each tube was spread onto the LB-Kanamycin agar plates. The growth of the colonies was observed.

Single Colony PCR

Single colonies from the test plate were randomly selected and patch-plated on an LB-Kanamycin agar plate. The plate was incubated at 37°C, overnight. These colonies were used for single colony PCR (SCP).

Each of these patch-plated colonies was suspended in 50µl of sterile distilled water with the help of sterile 10µl tips. Only sterile distilled water was kept in control. Individual transformants were lysed in water with a short heating step at 95°C for 10 minutes. Centrifuge at 10,000 rpm for 5 minutes. This brief spin will settle down the debris. 5µl of supernatant was used as the template. The following was the PCR reaction mixture (**Table 5. 15**) and conditions (**Table 5. 16**).

Chapter 5

Table 5. 15 Single colony PCR reaction mixture.

Reagents	Volume per reaction
Emerald 2X	10 μ l
T7 promoter*	1 μ l
T7 terminator*	1 μ l
Distilled water	3 μ l
Template	5 μ l
Total Volume	20 μ l

Note:

*T7 promoter: TAATACGACTCACTATAGGG ;

*T7 terminator: GCTAGTTATTGCTCAGCGG.

Table 5. 16 Single colony PCR reaction conditions.

Steps	Temperature ($^{\circ}$ C)	Time	Cycles
Initial Denaturation	95 $^{\circ}$ C	3 min	1
Denaturation	95 $^{\circ}$ C	45 sec	30
Annealing	55 $^{\circ}$ C	30 sec	
Initial Extension	72 $^{\circ}$ C	90 sec	
Final Extension	72 $^{\circ}$ C	10 min	1
Storage	4 $^{\circ}$ C	infinite	

PCR products were resolved in 0.8% agarose gel. The positive clones were identified if the amplicon size was 1.5kbp, whereas the negative clones were identified if the amplicon size was 250bp.

5.3.2.4 Over-expression of phytase in *E.coli* BL21(DE3) host system and purification of phytase enzyme:

Over-expression of phytase in *E.coli* BL21(DE3) host system:

The recombinant constructs (phyk-pET29b and appA-pET29b) were isolated by using GeneJET Plasmid Miniprep kit, Thermo Fisher (Catalog: K0503). It was transformed into the host system i.e., *E.coli* BL21(DE3) competent cells. The transformed colonies were patch-plated on an LB-Kanamycin agar plate. Individual colonies were inoculated in 5ml LB-Kanamycin broth and incubated at 37°C, 180rpm, overnight. They were isolated on an LB-Kanamycin agar plate. A single colony from each plate was inoculated in fresh 5ml LB-Kanamycin broth and incubated at 37°C, 180rpm, 16-18hrs (primary culture). 1ml of the primary culture was inoculated in fresh 99ml of LB-Kanamycin broth (i.e., 1:100 dilution) (secondary culture). It was then incubated at 37°C, 180rpm. The O.D._{600nm} of the secondary culture should be 0.6-0.8, followed by induction. The culture containing phyk-pET29b was induced with 1mM IPTG and incubated at the optimized conditions i.e., 33°C, 120rpm, 20hrs (for phyk-pET29b recombinant constructs). In the case of appA-pET29b recombinant construct, we performed induction at different IPTG concentrations (0.25, 0.5, 1, and 1.5mM) and temperatures (15, 20, 25, and 30°C).

Cell lysis:

In both cases, the induced culture was pelleted by centrifugation at 10,000 rpm, 10 minutes, at RT. Each pellet was resuspended in 3ml of lysis buffer (25mM Tris-base + 300mM NaCl). In total 30ml cell suspension was collected in the glass beaker. 200 µL of 1X EDTA-free protease inhibitor cocktail was added. It was then subjected to sonication by using a Q125 Sonicator-#4420 probe. The sonication conditions were 125 watts, 20 kHz, and 30% intensity, 10sec ON and 10sec OFF for a total of 1hr. After sonication, the cell lysate was centrifuged at 11,000 rpm for 30 minutes at 4°C. The supernatant was collected and stored at -20°C for further steps of purification.

Purification of phytase by using Immobilized metal affinity chromatography:

His60NiSuper Flow resin was used for purification. 1ml of resin was poured into the 5ml Pierce™ Disposable Columns (Catalog number: 29922). The storage buffer was allowed to pass through the column. The beads trapped in the column were washed with 10ml Milli-Q water. 5ml of equilibration buffer (50mM Tris, 300mM NaCl, 10mM Imidazole, pH7) was then

added to the column and the flow-through was discarded. This helps in equilibrating the beads. The beads were then removed from the column by adding approx. 2-3 ml of cell lysate and gently mixed. This was collected in the fresh 50ml falcon. This process was done multiple times till all the beads were removed from the column. This mixture was then incubated at 4°C overnight for binding on the rocker at a gentle speed. The mixture was added into the column again, in batches. The flow through of this step was collected and labeled as (FT). 10ml of wash buffer (50mM Tris, 300mM NaCl, 40mM Imidazole, 5% glycerol, pH7) was added to the column and the flow-through was collected (Labeled as Wash). The nozzle was kept closed as 5ml of elution buffer (50mM Tris, 300mM NaCl, 450mM Imidazole, 5% glycerol, pH7) was added to the column. The nozzle was opened to collect the flow-through after five minutes and this elution was collected in a volume of 0.5ml in each microcentrifuge tube (E1-9: 9 elution fractions were collected). The protein concentration was measured using nanodrop and the absorbance of the samples was taken at $A_{280\text{nm}}$. The E1-E9, FT, and wash fractions were resolved on 10% SDS-PAGE gel.

The dialysis was performed by using a dialysis buffer (20mM Tris, 150mM NaCl, 3% glycerol, pH7). For dialysis: Elution fractions 1, 2, 3, and 4 were combined as they exhibited higher protein concentration as compared to other elution fractions. This combined fraction was dialyzed (Labeled: fraction 1). Elution fractions 4-8 were pooled together in another dialysis bag (labelled: fraction 2). The dialyzed fractions were resolved on 10% SDS-PAGE gel. The protein concentration was checked after dialysis using nanodrop at $A_{280\text{nm}}$.

5.3.2.5 Biochemical characterization of phytase:

Standard KH_2PO_4 reaction setup:

The standard to estimate the released inorganic phosphate was constructed by using potassium dihydrogen phosphate (KH_2PO_4) at concentrations ranging from 0.1 $\mu\text{moles/ml}$ to 2 $\mu\text{moles/ml}$ (**Table 5. 17**). The termination buffer was freshly prepared by adding 5% ammonium molybdate: 5N H_2SO_4 : Acetone 100% (1:1:2)

Chapter 5

Table 5. 17 Reaction setup for KH_2PO_4 standardization. Stock solution of KH_2PO_4 : 50mM.

Stock solution of phytate: 44mM. The absorbance was measured at 400nm.

KH_2PO_4 concentration ($\mu\text{moles/ml}$)	KH_2PO_4 (ml)	Distilled water (ml)	Phytate (ml)	Distilled water (ml)	Incubation (Temperature and Time)	Termination buffer (ml)
0.1	0.002	0.048	0.05	0.4	37°C for 30 min	0.5
0.25	0.005	0.045				
0.5	0.010	0.040				
1	0.020	0.030				
1.5	0.030	0.020				
2	0.040	0.010				
50	0.050	-				
Blank	-	0.050				

Standard reaction setup to estimate phytase activity of purified phytase:

The modified Heinonen method (Suleimanova et al., 2015) (Nagar et al., 2021) was used to estimate the release of inorganic phosphate. The standard reaction setup for estimating total inorganic phosphate in the test sample (purified PVP was used) was prepared as per the following **Table 5. 18**. One unit (U) of phytase activity was defined as the amount of phytase required to liberate $1\mu\text{mol}$ of inorganic phosphorous per minute at 37°C.

Table 5. 18 Standard reaction setup for estimating phytase activity of purified phytase.

Reagents	Test	Enzyme blank
Substrate (μl)	50	50
Enzyme/ dialysis buffer (μl)	2 of purified phytase	2 of dialysis buffer
Reaction buffer (μl)	448	
Incubate at 37°C for 30min		
CRS (μl)	500	
Absorbance measurement	400nm	

Chapter 5

Effect of pH on phytase activity:

To estimate optimum pH: 0.1M Glycine-HCl (pH 3), and 0.1M Sodium acetate (pH 4, 4.5, 5.5, and 6) pH buffer systems were used in the above standard reaction setup to estimate phytase activity of purified PVP. The graph of % relative specific activity vs pH was plotted.

Effect of temperature on phytase activity:

To estimate optimum temperature: the standard assay was carried out at 25, 30, 40, 50, 55, 60, 65°C. The graph of % relative specific activity vs temperature (°C) was plotted.

5.3.3 Results

5.3.3.1 Confirmation of phytase gene in *P.vagans* and *E.tarda*:

According to UniProtKB, phytase genes presented in *P.vagans* and *E.tarda* were *phyk* and *appA* respectively. The study started with amplification of phytase gene from *P.vagans* and *E.tarda* by using gDNA (Figure 5. 37) as a template.

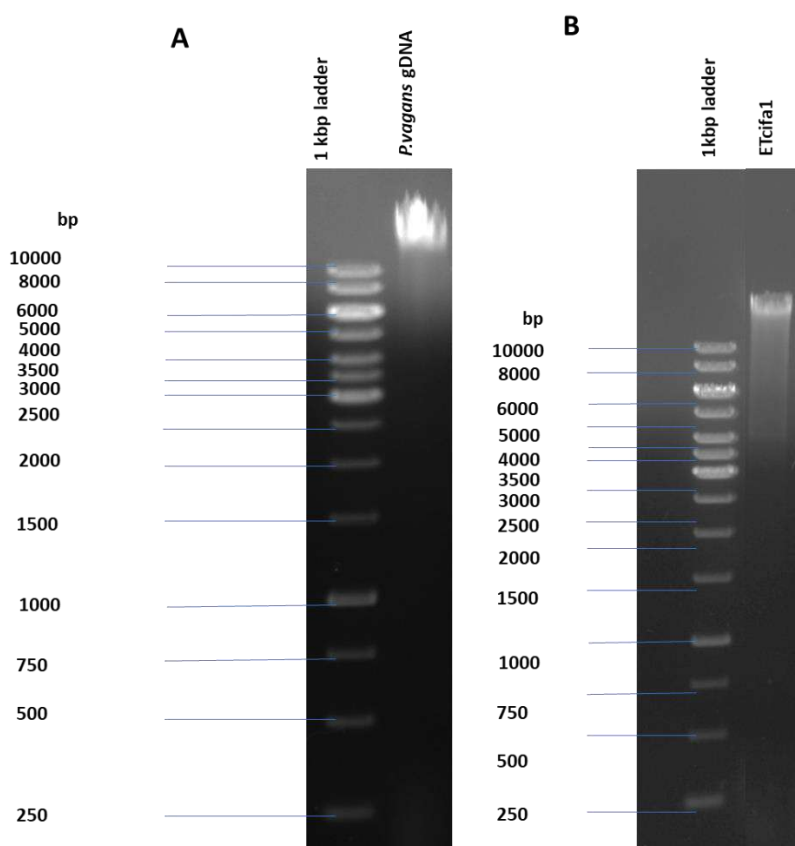


Figure 5. 37 Isolation of gDNA from A. *P.vagans*, B. from *E.tarda* by using the PureLink® Genomic DNA kit.

Chapter 5

The primers were designed to amplify full-length phytase genes. The expected size of amplicon (approximately 1.3 kbp -1.4 kbp phytase gene of *P.vagans* and *E.tarda*) was observed on 0.8% agarose gel (**Figure 5. 38**).

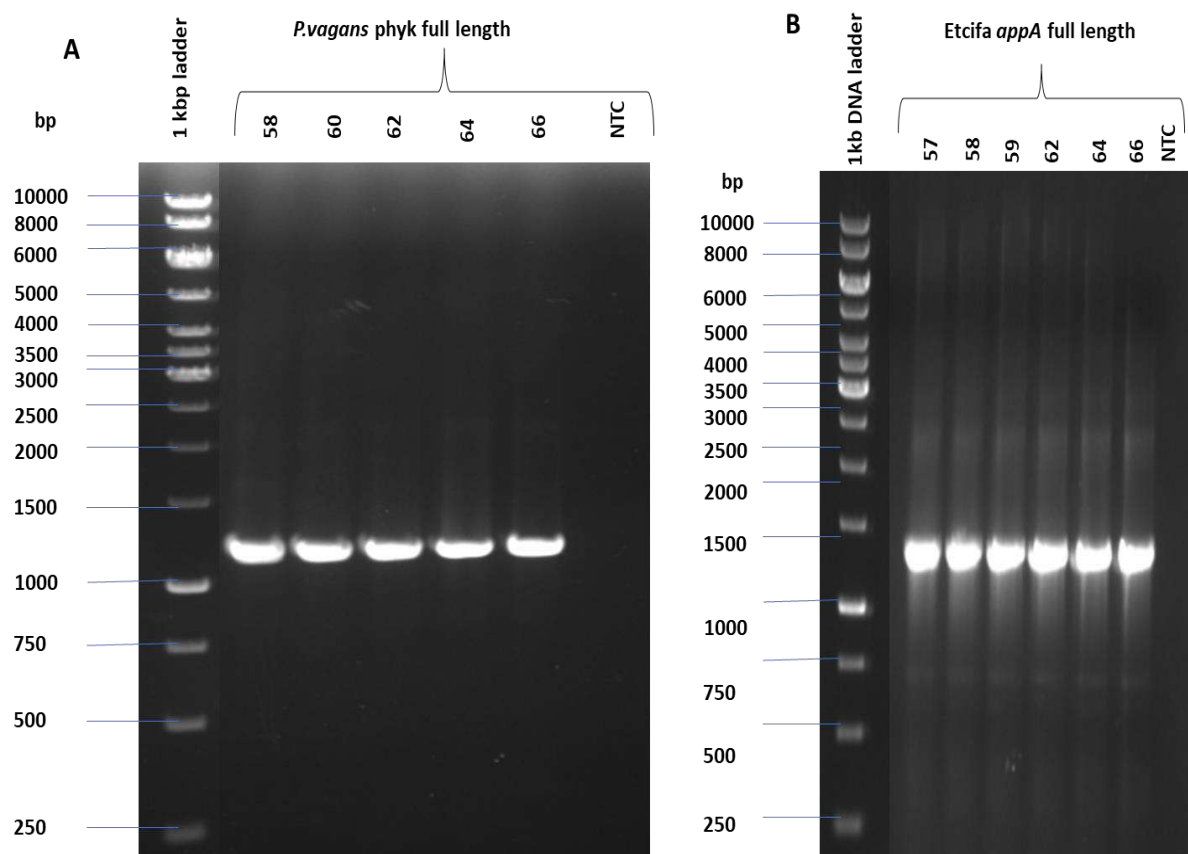


Figure 5. 38 A. Full-length *P.vagans* phyk gene PCR amplicons at different annealing temperatures. B. Full-length *E.tarda* appA gene PCR amplicons at different annealing temperatures.

These amplicons were gel purified (**Figure 5. 39**) and subjected to Sanger's sequencing (**Figure 5. 40**). The sequencing results revealed the presence of conserved residues which were important for catalysis. Blastn analysis of the sequenced full-length *P.vagans* amplicon and *E.tarda* amplicon was performed. In -the case of *P.vagans* analysis, the hit found was *P.vagans* C9-1 (Sequence I.D.: CP002206.1), and in-case of *E.tarda* analysis, the hit found was *E.tarda* (Sequence I.D.: CP084506.1). In summary, the PCR amplification and sequencing results confirmed the presence of the phytase gene in *P.vagans* and *E.tarda*. The full-length phytase gene from both *P.vagans* and *E.tarda* was then cloned and over-expressed as described below.

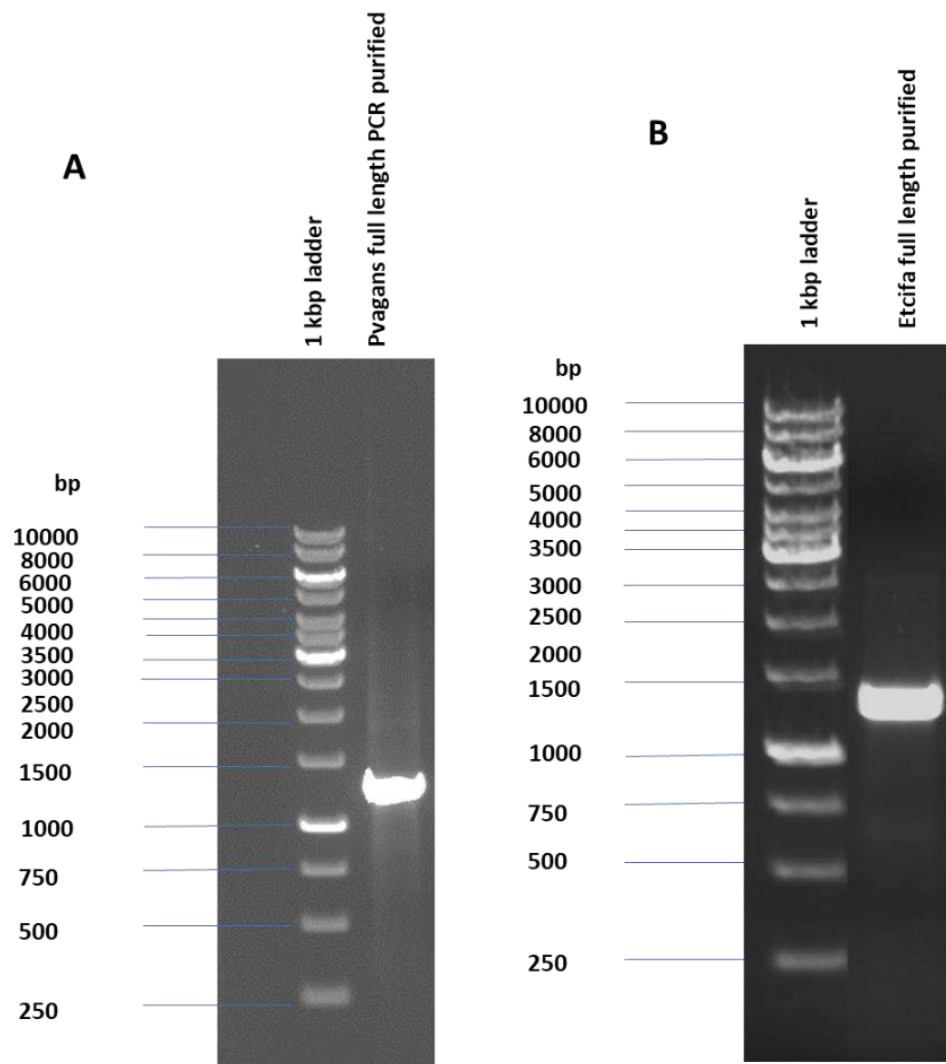


Figure 5. 39 Gel purified A. *P.vagans* phytase gene PCR amplicon , B. *E.tarda* phytase gene used for Sanger's sequencing.

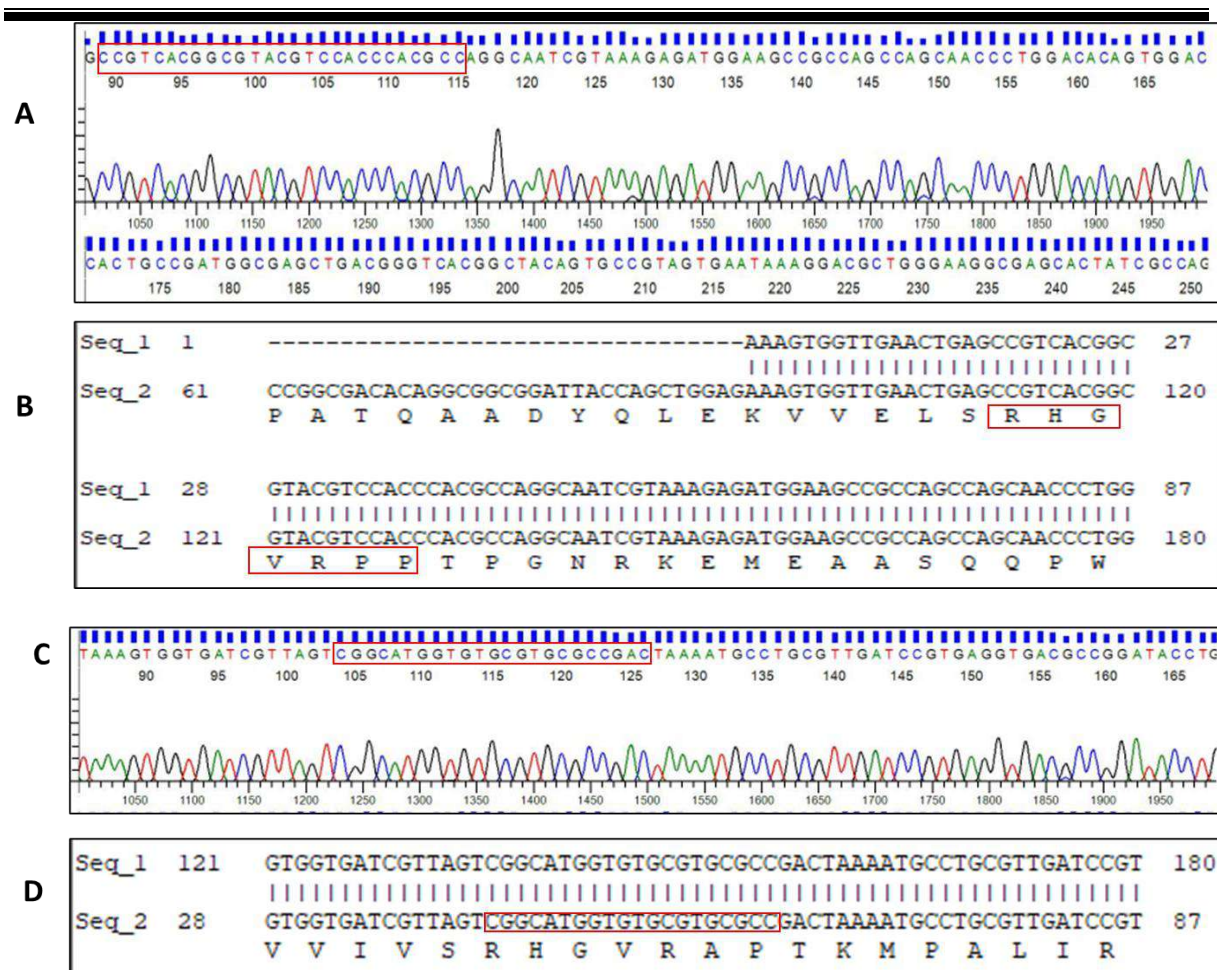


Figure 5. 40 A. Snippet of the *P.vagans* phytase gene electrogram: The red highlighted nucleotide sequence codes for the conserved N-terminal ‘RHGVRPP’ protein sequence. The presence of this sequence confirms the presence of the phytase gene in *P.vagans*. B. Snippet of the alignment constructed in Serial cloner: Seq_1: Sanger’s sequenced *P.vagans* full-length phytase gene aligned with the Seq_2: Nucleotide sequence of *P.vagans* (ENA database) which was used for primer designing. The *P.vagans* full-length phytase gene exhibited the presence of a conserved N-terminal ‘RHGVRPP’ protein sequence (Highlighted in red).C. Snippet of the *E.tarda* phytase gene electrogram: The red highlighted nucleotide sequence codes for the conserved N-terminal ‘RHGVRAP’ protein sequence. The presence of this sequence confirms the presence of phytase gene in *E.tarda*. D. Snippet of the alignment constructed in Serial cloner: Seq_1: Nucleotide sequence of *E.tarda* from ENA database which was used for primer designing aligned with the Seq_2: Sanger’s sequenced *E.tarda* full-length phytase gene. The *E.tarda* full-length phytase gene exhibited the presence of a conserved N-terminal ‘RHGVRAP’ protein sequence (Highlighted in red).

5.3.3.2 Cloning of phytase gene from *P.vagans* and *E.tarda*:

This study involved re-amplification of the phytase gene from both *P.vagans* and *E.tarda* (obtained from the previous step) by using primers possessing restriction sites on both ends (3'NcoI and 5'NotI) (Figure 5. 41).

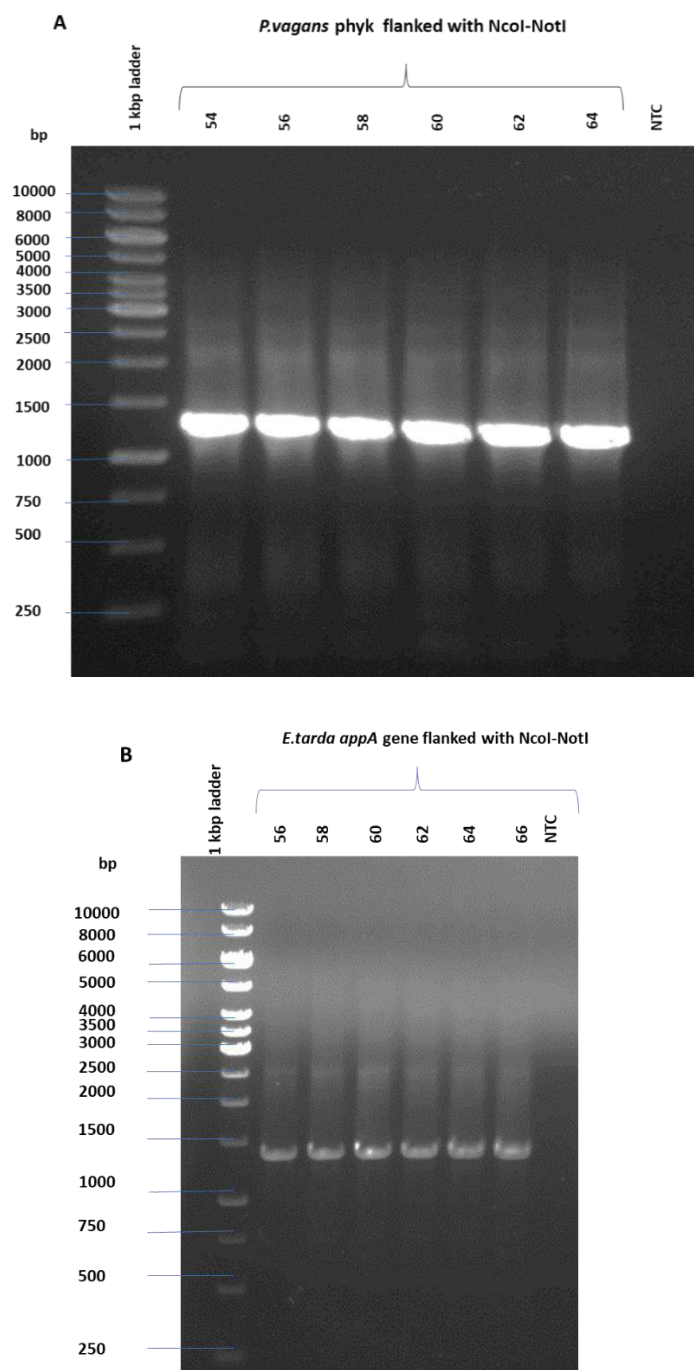


Figure 5. 41 Full-length phytase gene from: A. *P.vagans* and, B: *E.tarda* flanked with NcoI and NotI restriction sites at the different annealing temperatures.

Chapter 5

pET29b plasmid was used for cloning and over-expression of phytase from both *P.vagans* and *E.tarda*. It was isolated by using GeneJET Plasmid Miniprep kit, Thermo Fisher (**Figure 5. 42**) . The supercoiled pET29b form was more than 90%.

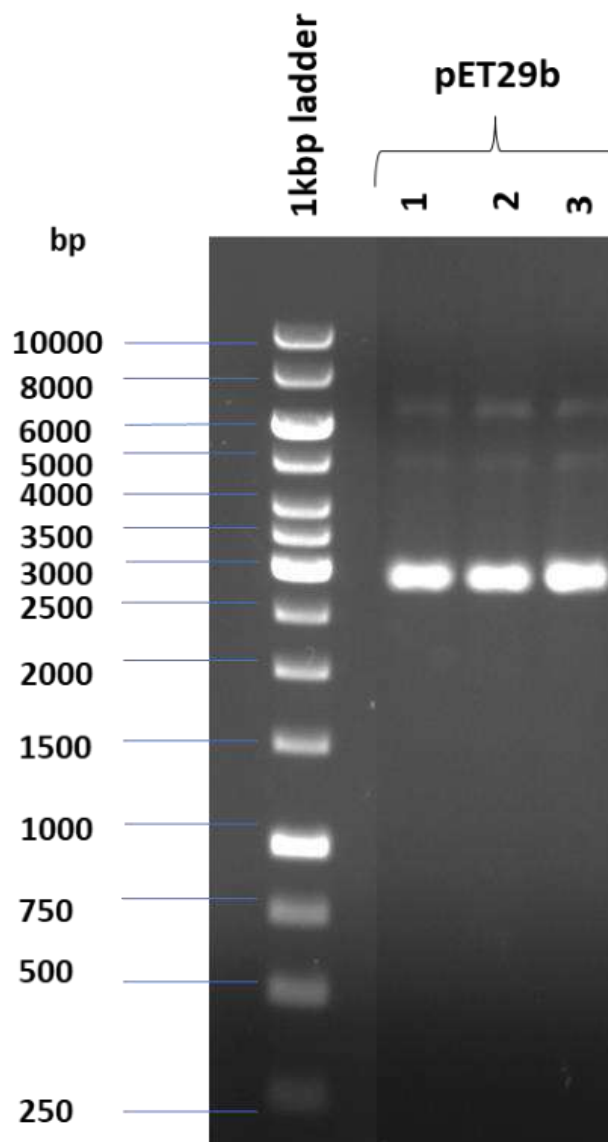


Figure 5. 42 pET29-b plasmid isolated by using GeneJET Plasmid Miniprep kit, Thermo Fisher.

The resultant amplicons and isolated vectors i.e., pET29-b were subjected to restriction digestion by using restriction enzymes NcoI and NotI. The restriction products were purified and separated into 0.8% agarose gel (**Figure 5. 43**).

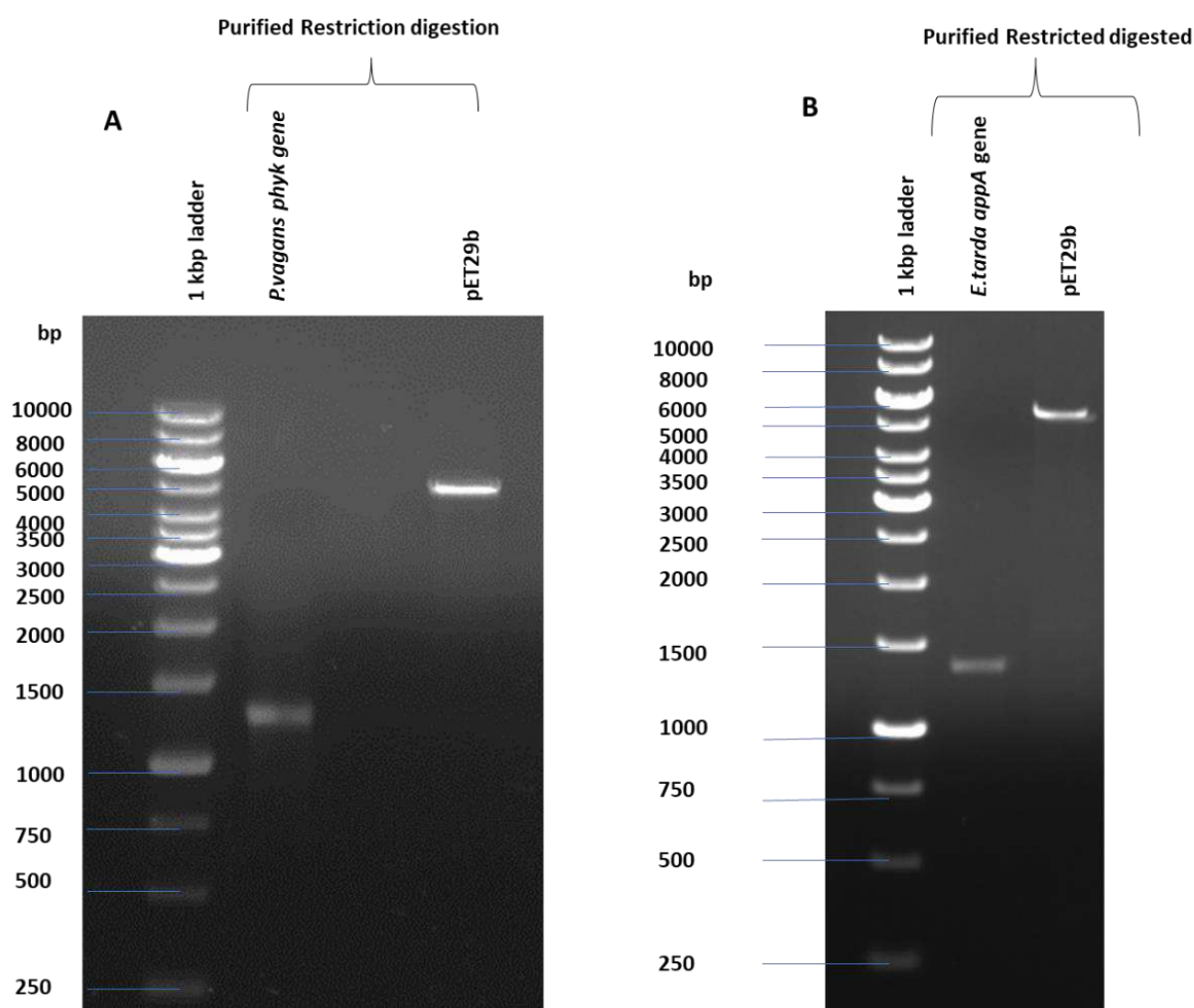


Figure 5. 43 Restriction digestion with NcoI and NotI restriction enzymes of the following samples: A. purified *P.vagans* phyk gene, and isolated pET29b, B. Purified *E.tarda* appA gene and isolated pET29b.

The restricted amplicons and vector were ligated and resulted in two systems i.e., phyk-pET29b recombinant construct (from *P.vagans*) and appA-pET29-b construct (from *E.tarda*). These recombinant constructs were separately transformed in *E.coli* DH5 α host system. Positive clones were identified by single colony PCR (**Figure 5. 44**). According to the SCP, the positive clones (recombinant construct: phyk-pET29b and appA-pET29-b) exhibited an amplicon size of around 1.4kbp, whereas negative clones (only pET29b) exhibited an amplicon size of around 250bp.

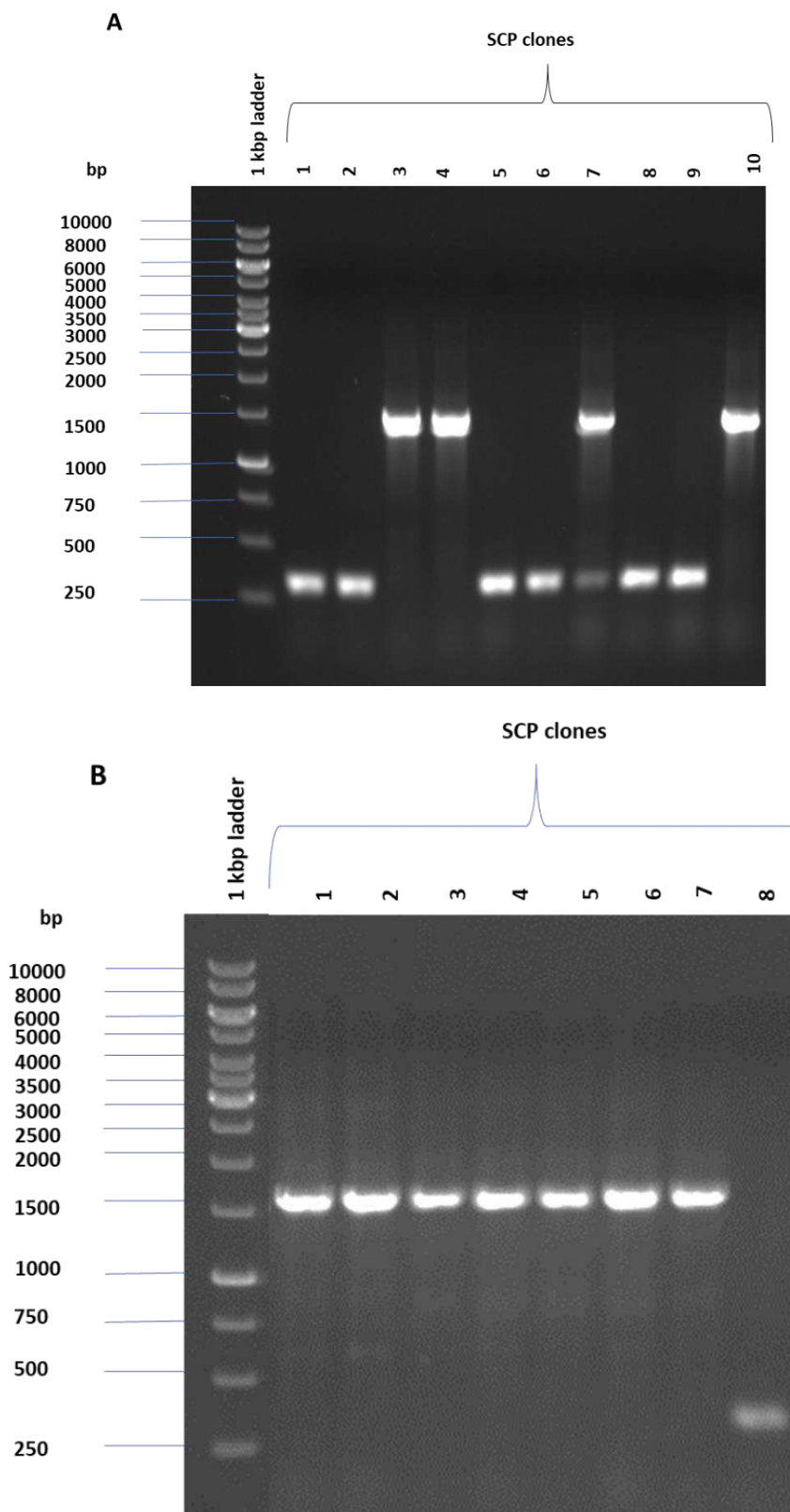


Figure 5. 44 Single Colony PCR of: A. phyk-pET29b recombinant construct (from *P.vagans*) and, B. appA-pET29-b construct (from *E.tarda*).

The positive phyk-pET29-b and appA-pET29-b constructs were used for over-expression in the *E.coli* BL21(DE3) system. The sequencing of the recombinant construct enabled us to confirm that there was a no-frame shift and expected to yield an active phytase enzyme upon over-expression (**Figure 5. 45**).

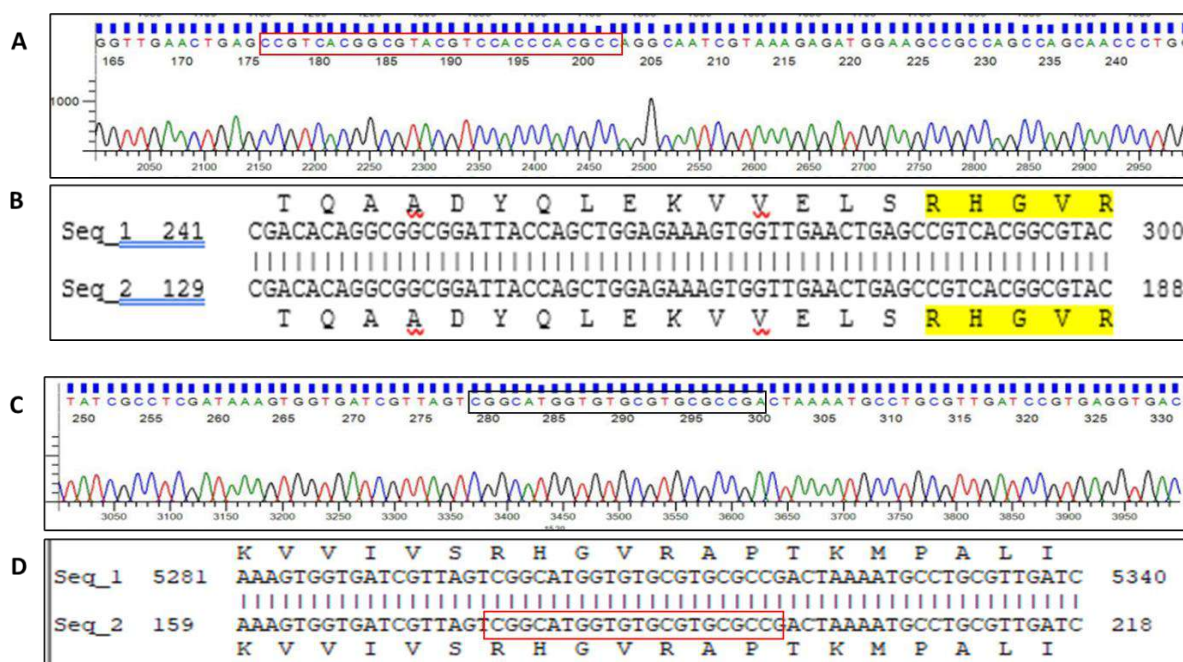


Figure 5. 45 A. Snippet of the phyk-pET29-b (from *P.vagans*) gene electrogram: The red highlighted nucleotide sequence codes for the conserved N-terminal ‘RHGVRPP’ protein sequence. B. Snippet of the alignment constructed in Serial cloner: Seq_1: Sanger’s sequenced phyk-pET29-b (from *P.vagans*) gene aligned with the Seq_2: In-silico ligated product (phyk-pET29b) built-in Serial cloner. The phyk-pET29b construct exhibited the presence of a conserved N-terminal ‘RHGVRPP’ protein sequence (Highlighted in yellow). C. Snippet of the *E.tarda* phytase gene electrogram: The black highlighted nucleotide sequence codes for the conserved N-terminal ‘RHGVRAP’ protein sequence. D. Snippet of the alignment constructed in Serial cloner: Seq_1: Nucleotide sequence of *E.tarda* from ENA database which was used for primer designing aligned with the Seq_2: Sanger’s sequenced appA-pET29b construct. The appA-pET29b construct exhibited the presence of a conserved N-terminal ‘RHGVRAP’ protein sequence (Highlighted in red).

5.3.3.3 Over-expression of phytase gene in *E.coli* BL21(DE3) host system and purification of phytase enzyme:

This study involved the transformation of positive recombinant constructs phyk-pET29-b and appA-pET29-b in separate *E.coli* BL21(DE3) host systems. The over-expression of phytase was performed by using 1mM IPTG as inducer at 33°C, overnight. The over-expressed PVP was purified using the Immobilized Metal affinity chromatography technique (IMAC) via His60NiSuperflow resins. PVP (approximately 43kDa) was found to be overexpressed in the *E.coli* BL21(DE3) host system (**Figure 5. 46**). It was dialyzed for further analysis (**Figure 5. 47**). The quantity of purified PVP after dialysis was measured at absorbance (A_{280} nm) by using an Epoch microplate reader. We obtained 0.6mg/ml (Fraction 1) and 0.1 mg/ml (fraction 2) purified PVP. The purified PVP was then subjected to biochemical characterization i.e., pH and temperature optima.

ETP failed to over-express to the optimum level even after multiple attempts of varying temperatures, rpm conditions, and IPTG concentrations (**Figure 5. 48**). Hence, was did not select it for rational engineering.

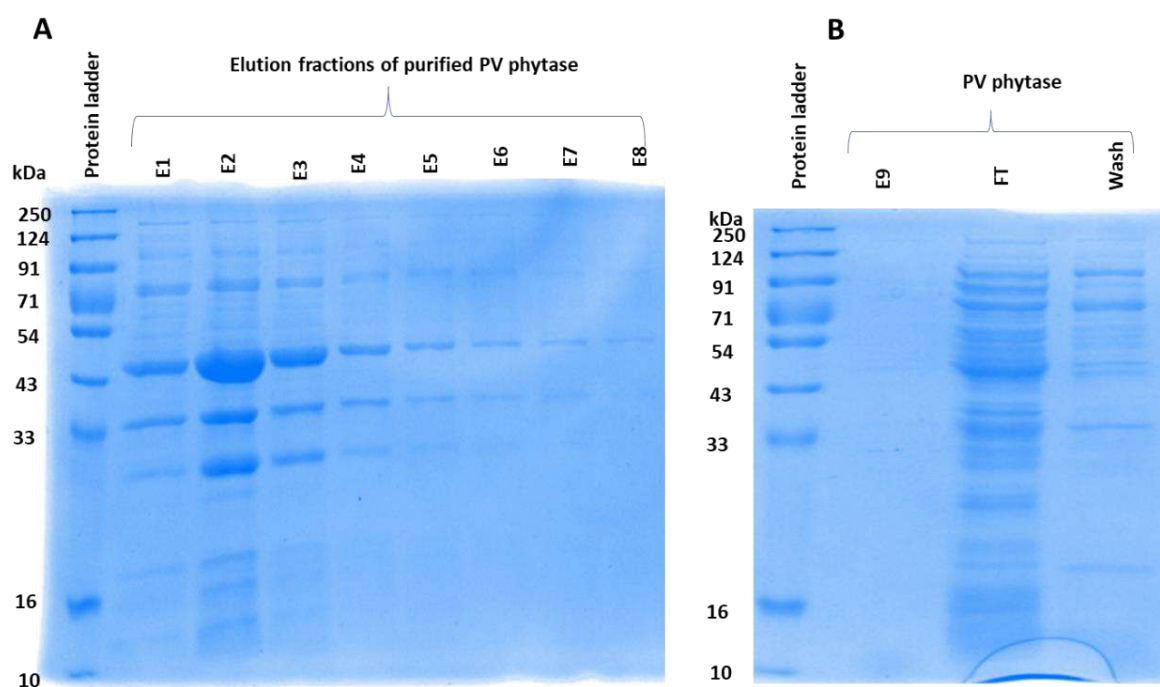


Figure 5. 46 PVP purified by using Immobilized Metal affinity chromatography technique (IMAC) via His60SF resins and resolved on 10% SDS-PAGE gel: A. Elution fractions (E1-E8), B. Elution fraction 9, along with Flowthrough (FT), and Wash fraction (Wash).

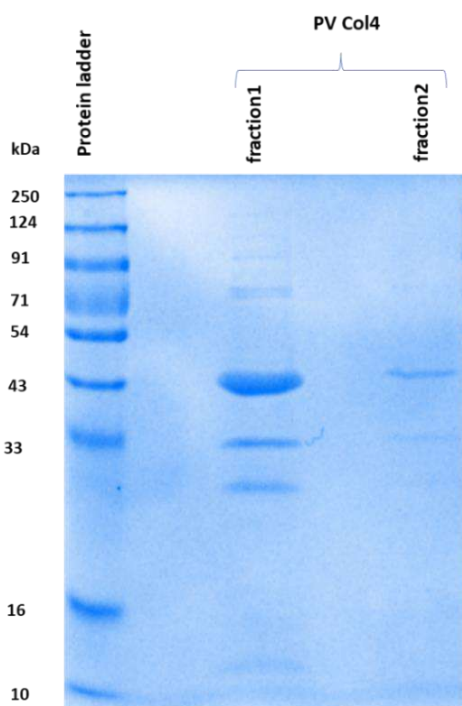


Figure 5. 47 Purified PVP after dialysis (Fraction 1 and 2) resolved on 10% SDS-PAGE gel.

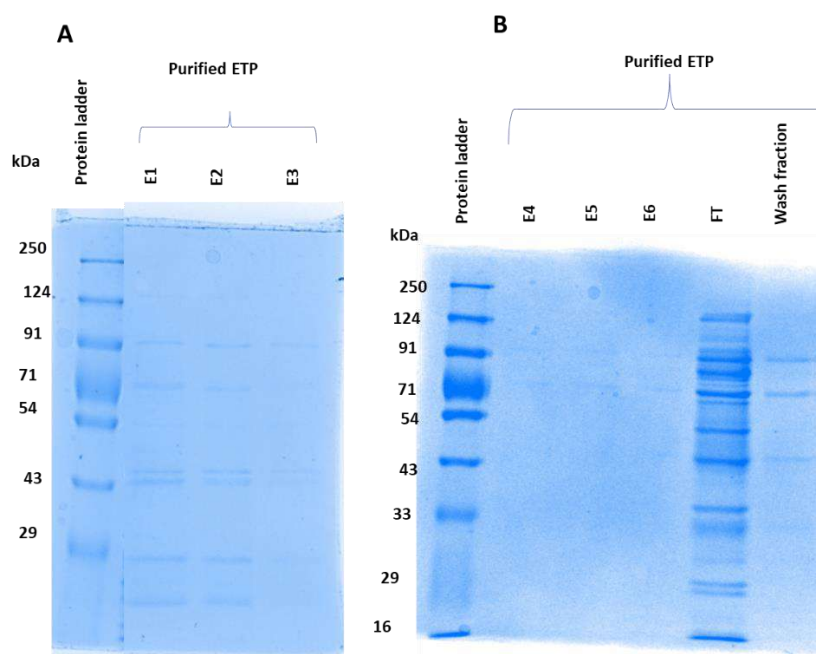


Figure 5. 48 ETP purified by using Immobilized Metal affinity chromatography technique (IMAC) via His60NiSF resins and resolved on 10% SDS-PAGE gel: A. Elution fractions (E1-E8), B. Elution fraction 9, along with Flowthrough (FT), and Wash fraction (Wash).

5.3.3.4 Biochemical and Biophysical characterization of PV phytase:

The optimum pH and temperature of PV phytase were pH4, (**Figure 5. 49**) and 40°C (**Figure 5. 50**) respectively. KH_2PO_4 was kept as a standard to analyze the release of inorganic phosphate (**Figure 5. 51**).

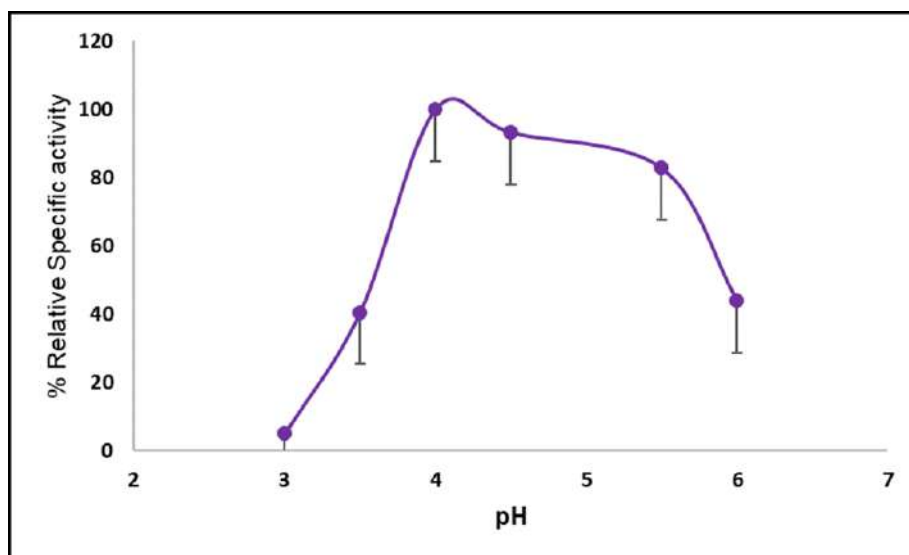


Figure 5. 49 Effect of different pH buffer (3, 3.5, 4, 4.5, 5, 5.5, and 6) conditions on PVP activity.

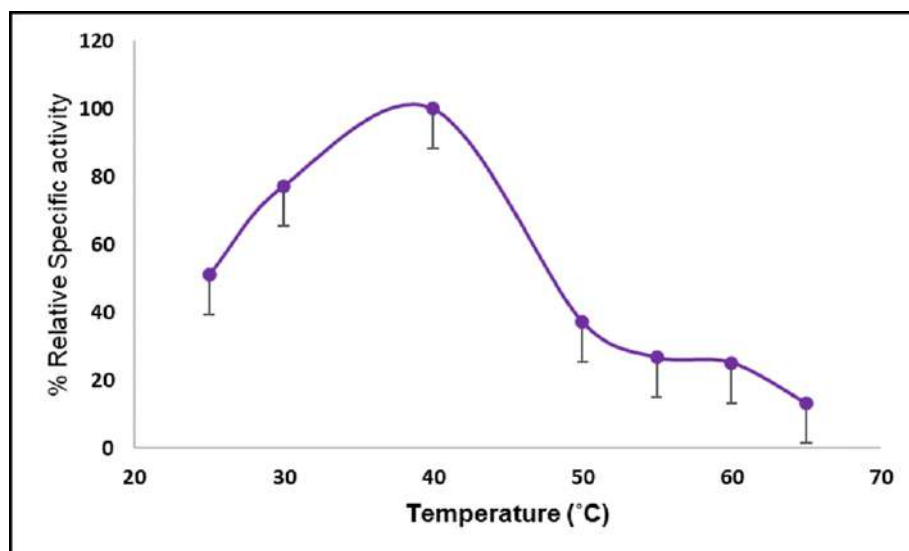


Figure 5. 50 Effect of different temperature conditions on PVP.

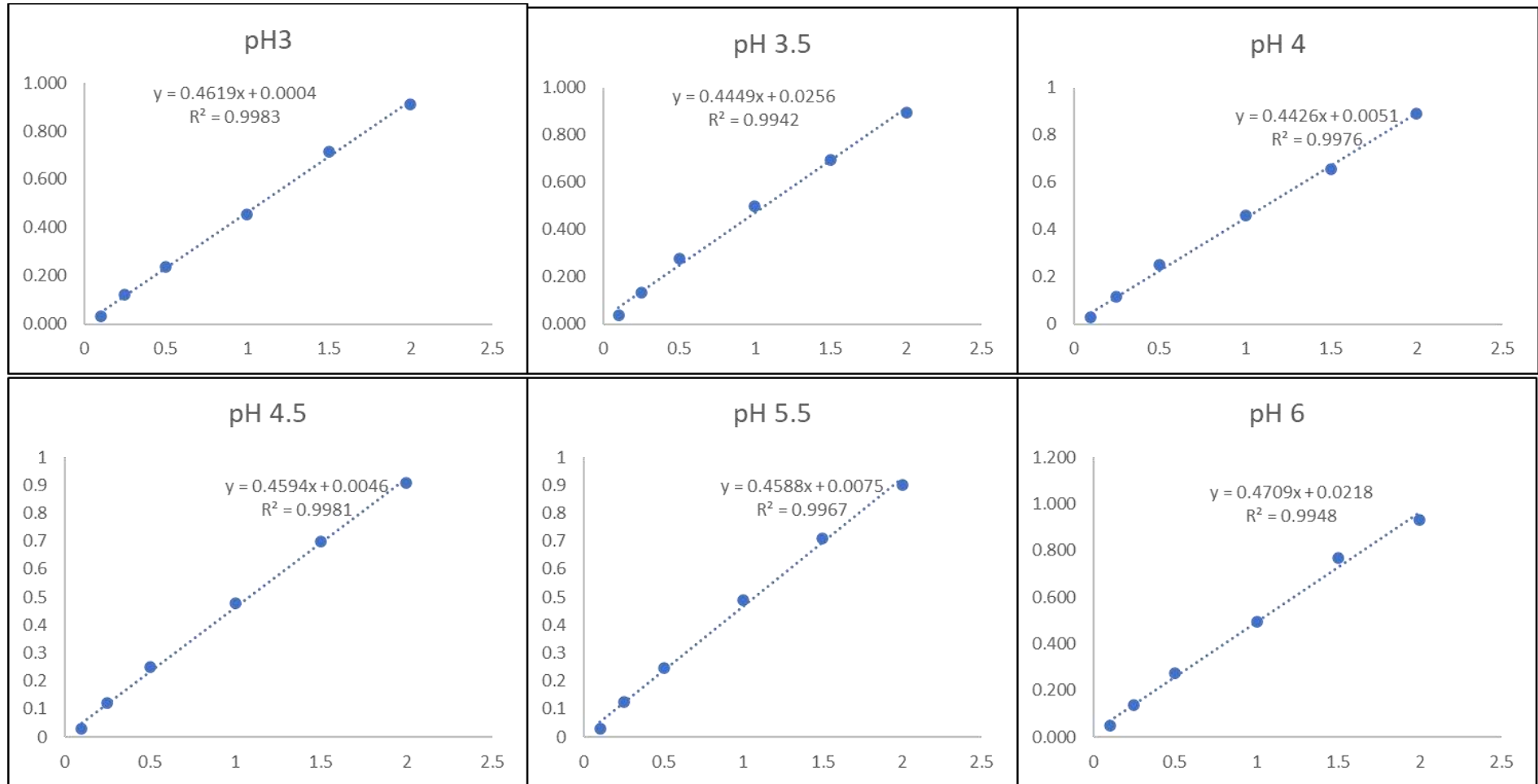


Figure 5. 51 KH_2PO_4 was used as a standard to analyze the release of inorganic phosphate at different pH conditions (pH3, 3.5, 4, 4.5, 5, 5.5, and 6).

5.4 Discussion:

In-silico characterization of novel enzymes helped to gain information about their primary sequence, secondary structure, and tertiary structure along with the functional, evolutionary relationship with the experimentally proven homologs, and dynamics studies via molecular dynamic simulations. E.g., Pramanik et.al. performed *In-silico* characterization of *Pseudomonas aeruginosa* phosphatase by using various computational tools to exploit its physicochemical characteristics, secondary and tertiary structure, phylogenetic assessment, and functional analysis (Pramanik et al., 2017). There are various other enzymes such as thermostable chitinase II (Iqbal et al., 2015), Cellobiohydrolase (Dadwal et al., 2022) including phytases (Gontia-Mishra et al., 2014) which were characterized by using computational tools/servers/software.

In the present study, *In-silico* characterization of PVP and ETP was conducted to gain insights into the overall structural, functional, and dynamics behavior. This study involved primary sequence, secondary structure analysis, tertiary structure modeling plus validation, and phylogenetic and functional assessment by using various computational tools/servers/software. It also involved molecular docking and molecular dynamic simulation studies to examine its behavior at different temperatures.

The primary sequence analysis suggested that PVP and ETP both belonged to the histidine acid phosphatase superfamily. The closest homolog of PVP and ETP were KPP and HAP respectively. The known experimentally validated phytases such as *Klebsiella* (Böhm et al., 2010b), *Hafnia* (Ariza et al., 2013), *E.coli* (Lim et al., 2000), and many others also belong to the same superfamily. The overall primary sequence analysis also suggested that PVP was thermostable and exhibited better interaction with the surrounding solvent as compared to ETP. The secondary structure analysis revealed that the topology of ETP and PVP were like each other.

The analysis of tertiary structures demonstrated that PVP and ETP possess similar characteristics, consisting of α and α/β domains, with the catalytic site situated at the interface between these domains. The crystal structures of *Klebsiella* (Böhm et al., 2010b) and *Hafnia* (Ariza et al., 2013) also exhibited a comparable organization to PVP and ETP. Examination of the catalytic sites in PVP and ETP revealed the presence of conserved regions, namely RHGXRXP at the N-terminal and HD at the C-terminal, which resemble the corresponding

regions in *Klebsiella* (Böhm et al., 2010b) and *Hafnia* (Ariza et al., 2013). These findings strongly suggest that PVP and ETP can degrade phytate through a two-step mechanism, similar to the experimentally characterized phytases of *Klebsiella* (Böhm et al., 2010b) and *Hafnia* (Ariza et al., 2013). Despite having a sequence identity below 70% with the template, PVP and ETP exhibited structural similarity to their template structures. The validation of the tertiary structure suggests the accuracy and reliability of the modeled tertiary structure. The validation results revealed that the PVP structure was more reliable as compared to that of ETP.

T_m values of protein determine the thermal stability of the protein. According to SCOP analysis, PVP was thermostable compared to that of ET. The results of the primary sequence analysis corroborate these observations.

Molecular dynamic simulation of protein helps to understand the conformational dynamics of protein. It also provides insights into structural changes and flexibility at different temperatures. E.g., Wang et.al. used molecular dynamic simulation (MDS) to understand the effect of temperature on the structure and function of Chitosanase (Wang et al., 2023). Similarly, Hua et.al. used MDS to study the stability of Glutamate Decarboxylase (Hua et al., 2020). Many researchers have utilized molecular dynamics simulation to gain insights into the conformational dynamics of various enzymes including phytases such as *Aspergillus niger* phytase (K. Kumar et al., 2015) and *E.coli* (Shivange et al., 2010) phytase. In the present study, molecular dynamic simulation was conducted to understand the conformational dynamics of PVP and ETP at different temperatures (300 to 450K for 30ns). The MDS analysis, especially the average R_g, SASA analysis revealed that PVP was stable as compared to ETP.

Overall, the combined analysis of primary sequence, tertiary structure, validation, and molecular dynamics simulation provides valuable insights into the structural characteristics and dynamics of PVP and ETP, highlighting PV's thermostability and stability compared to ETP.

Experimental validation along with *in-silico* characterization enables researchers to acquire a more comprehensive understanding of enzyme characteristics. In this study, experimental validation of both PVP and ETP was conducted to shortlist one candidate (either PVP or ETP) based on the results. The steps involved in experimental validation were full-length phytase gene confirmation via PCR amplification by using suitable primers, cloning, and over-expression, followed by biochemical characterization (effect of pH and temperature). The full-length phytase genes of PVP and ETP were approximately 1.3 and 1.4kbp in size. This size of the gene was comparable to that of experimentally validated phytase genes e.g., *Hafnia* phytase

(Ariza et al., 2013), *E.coli* phytase (Abeldenov et al., 2017) and *Klebsiella* phytase (Sajidan et al., 2004). The full-length PVP and ETP genes were cloned in the pET29-b vector and transformed in the *E.coli* DH5 α host system and over-expressed in *E.coli* BL21(DE3). PVP was found to be produced in large quantities as compared to that of ETP. Hence, only PVP was purified using IMAC. The size of PVP was found to be approximately 43kDa which was comparable to its closet homolog *Klebsiella* phytase (Sajidan et al., 2004) (Böhm et al., 2010b). The purified phytase was biochemically characterized. The optimum pH of PVP was found to be pH4 and the optimum temperature was 40°C which was similar to *Klebsiella* phytase (optimum pH 5 and optimum temperature 45°C) (Böhm et al., 2010b) .

In summary, the selection of PVP for rational engineering was based on a combination of *in silico* characterization and experimental validation analyses.

5.5 Summary and Conclusion:

In our present study, we focused on *in-silico* characterization of PVP and ETP using various computational tools to understand their structural, functional, and dynamic behavior. PVP and ETP belong to the histidine acid phosphatase superfamily, sharing similarities with experimentally validated phytases. Analysis revealed that PVP exhibited thermostability and stronger solvent interactions compared to ETP. Tertiary structure examinations indicated that both PVP and ETP possess similar structures and conserved catalytic sites, suggesting a shared phytate degradation mechanism akin to known phytases. Despite sequence differences, both enzymes showed structural resemblance to their templates, validated through modeling assessments.

Additionally, thermal stability analysis and molecular dynamics simulations supported PV's stability over ETP. Experimental validation involving gene confirmation, cloning, over-expression, and biochemical characterization favored PVP, which produced larger quantities and exhibited properties akin to *Klebsiella* phytase, a well-established enzyme. Notably, PV's optimum pH and temperature closely resembled that of *Klebsiella* phytase.

The integrated approach of *in-silico* characterization and experimental validation led to the selection of PVP for potential rational engineering.

Chapter 6
***In-silico* hotspots
identification
and
experimental
validation of
shortlisted phytase**

In-silico hotspot identification and experimental validation of shortlisted phytase

6.1 Introduction:

This study aimed to investigate and identify regions or specific amino acid residues in PVP that contribute to its thermos-sensitivity. Various *in-silico* tools, software, and servers were utilized to analyze and identify these regions. Subsequently, computational methods were employed to suggest amino acid substitutions that could enhance stability.

The computational methods provided a list of mutant variants based on the calculated ΔG and $\Delta\Delta G$ (kcal/mol) values using the FoldX software. The mutants and wild type were synthesized and subjected to experimental validation to gain a deeper understanding of the stability of the novel PVP variants.

6.1.1 *In-silico* hotspot identification:

Protein engineering methodologies, such as rational or semi-rational strategies, and directed evolution, provide avenues to improve enzyme attributes like thermostability, resistance to degradation, and achieving stability across different pH ranges. E.g., Huang et.al. enhanced the thermostability of phospholipase D (PLD) by using directed evolution (Huang et al., 2020). Shivange et.al., used the directed evolution method to improve the thermostability of *Yersinia mollaretii* phytase (Shivange et al., 2012). Li et.al. improved the activity and stability of alkaline protease from *Bacillus clausii* by using directed evolution (Li et al., 2021). Directed evolution necessitates screening many colonies to identify enzymes exhibiting traits, resulting in increased expenses and time consumption (L. Li et al., 2022). Hence, rational engineering is used to overcome the limitations of directed evolution. E.g., Tan et.al. used rational engineering to enhance the thermostability of a novel acidobacteria-derived phytase (Tan et al., 2016). Zu et al. enhanced the thermostability of xylanase XynA through a rational engineering approach, which involved decreasing surface entropy (Zhu et al., 2023).

6.1.2 Experimental validation:

Experimental validation plays a vital role in enzyme engineering by confirming predictions, evaluating performance, assessing stability and functionality, refining design approaches, and ensuring the stability of mutants. This validation process significantly improved the credibility

and practical applicability of enzyme engineering strategies. E.g., Li et.al., identify hotspots in *Candida rugosa* lipases to enhance thermostability, and further experimentally validate the predictions via site-directed mutagenesis to generate predicted mutants, protein over-expression, purification, estimate pH, temperature, catalytic efficiency, and activity (G. Li et al., 2018a). Yang et.al., modified α -L-arabinofuranosidase by using structure-based rational engineering. In this study, the predicted mutants were constructed via site-directed mutagenesis, cloned, expressed, purified, and characterized (Yang et al., 2016). There are many other enzymes such as xylanase (Yang et al., 2007), β -Glucosidase (Zong et al., 2015), lipase (Zhang et al., 2019), feruloyl esterase (Yang et al., 2022), and phytase (Gordeeva et al., 2019) that were also experimentally validated.

6.2 *In-silico* hotspot identification:

6.2.1 Introduction:

The *in-silico* hotspot identification involves the use of diverse tools, servers, or software like Disulfide by Design 2.0, Consensus Finder, Consurf, manual consensus, Fireprot, Conformational Alphabet-Based Structure Flexible Fitting (CABSFLEX), and FOLDX position scan (**Figure 6. 1**). These methods collectively aid in pinpointing hotspots within PVP. Further details regarding each method are outlined below.

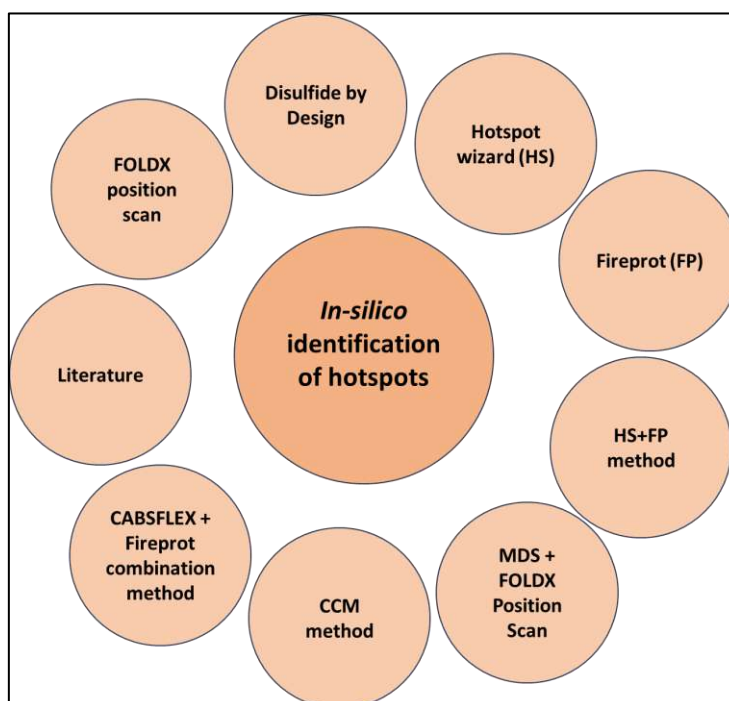


Figure 6. 1 Overview of *in-silico* hotspot identification by hotspots by various methods.

Chapter 6

Note: CCM: Consensus Finder + Consurf + manual consensus, MDS: Molecular Dynamics Simulations, CABSFLEX (Conformational Alphabet-Based Structure Flexible Fitting), Common of hotspot wizard and firepot (HS-FP) method.

6.2.2 Materials and Methods:

6.2.2.1 General Procedure:

- a. Identified *in-silico* hotspots within the PVP structure using the methods mentioned below.
- b. Performed desired substitutions of amino acid(s) within the wild-type protein sequence.
- c. Uploaded the modified/mutant sequence onto the SWISS-MODEL online platform to generate homology models of the structure.
- d. Employed the FOLDX repair module to rectify the mutant and wild-type PVP structures.
- e. Compare the ΔG value of the repaired wildtype (WT) PVP, which was -87.4kcal/mol, with the ΔG value of the mutants. Mutants with a ΔG value <-87.4 kcal/mol were shortlisted for further consideration.

We have listed below different methods to find potential hotspots, thereby generating mutants to improve the stability of PVP:

6.2.2.2 Disulfide by Design (D2D)

The structure of PVP was uploaded on the D2D online platform (<http://cptweb.cpt.wayne.edu/DbD2/>) (Craig and Dombkowski, 2013) (Dombkowski, 2003).

The criteria to generate the first list were as follows:

- a. Optimum Chi3 angle: +97/-87 with Tolerance: 5
- b. Optimum Ca-Cb-S angle: 114.60 with Tolerance: 6

This list underwent filtration based on the subsequent criteria:

- a. Exclusion of hotspots involved in native disulfide bond formation.
- a. Elimination of hotspots lacking a distance of at least 20 amino acids between them.
- b. Removal of amino acids positioned within proximity of 5Å from the catalytic site.

The mutants shortlisted through this method underwent modeling and subsequent processing following the steps outlined in the 'Common steps' methodology mentioned earlier.

6.2.2.3 Consensus Finder, Consurf, and manual consensus (CCM) method:

- a. Consensus Finder (<http://kazlab.umn.edu/>) (Xiong et al., 2022): The PVP sequence file was uploaded to the online server, generating an initial list of potential mutants. This list underwent refinement utilizing Consurf, as detailed below.
- b. Consurf (https://consurf.tau.ac.il/consurf_index.php) (Ashkenazy et al., 2016): Hotspot residues (from the initial list) with a conservation score exceeding 6 were excluded. This process resulted in the creation of a second list of shortlisted mutants, which then underwent manual consensus.
- c. Manual consensus:

A multiple sequence alignment was established by aligning PVP with reference sequences like 1DKM (*Escherichia coli* phytase), 3ZHC (*Citrobacter braakii* phytase), 4ARV (*Yersinia kristensenii* phytase), and 4ARS (*Hafnia alvei* phytase). The consensus residues of PVP hotspots (shortlisted in the second list) were noted within each alignment. Each of these identified hotspots underwent substitution by the recommended amino acids (as suggested by the manual consensus and Consensus Finder method) to enhance the stability of the PVP structure.

The mutants shortlisted through this method were then subjected to modeling and processed following the 'Common steps' methodology mentioned previously.

6.2.2.4 Literature:

This method involved modifying PVP by replicating specific mutations previously documented in *E. coli* phytase, which are known to enhance thermostability (Li et al., 2019a); (Kim and Lei, 2008a); (Chen and Tw, 2018); (Wang et al., 2015); (Fakhravar and Hesampour, 2018); (Zhang et al., 2016). These mutations were introduced into the corresponding residues of PVP.

The mutants shortlisted through this approach underwent modeling and further processing following the 'Common steps' methodology mentioned earlier.

6.2.2.5 Hotspot Wizard (HS):

The Hotspot Wizard (Bendl et al., 2016) incorporated several features like Functional hotspots, Stability hotspots by structural flexibility, and stability hotspots by sequence consensus to detect hotspots within PVP.

- a. For Functional hotspots and Stability hotspots by structural flexibility, residues involved in correlated positions, catalytic pockets, and tunnels were excluded.

- b. Regarding stability hotspots by sequence consensus features, correlated positions, catalytic pockets, and tunnels were also excluded.

Following the exclusion criteria, the mutants shortlisted by this method underwent modeling and further processing as detailed in the previously mentioned 'Common steps' methodology.

6.2.2.6 Fireprot (FP):

The PVP structure was used to identify mutations by considering energy, evolution, and a combination of both features (Musil et al., 2017). Automated calculations were utilized, applying specific parameters. Key residues involved in catalysis (R38, H39, G40, V41, R42, P43, P44, H303, and D304) were specifically selected.

The initial list of mutants was generated using all three approaches. This list underwent refinement based on the following criteria:

- a. For the combined mutant list: Only mutants exhibiting the highest stability concerning FOLDX/ROSETTA values were retained.
- a. For the Evolution mutant list: Only mutants with the highest FOLDX (kcal/mol) value combined with the highest BTC by majority and BTC by ratio were selected.

Subsequently, the mutants shortlisted using this method underwent modeling and further processing following the 'Common steps' methodology outlined earlier.

6.2.2.7 Common mutation between HS and FP:

Several mutants were shared between the HS and FP methods. These shortlisted mutants, identified through this commonality, underwent the modeling and processing steps outlined in the previously described 'Common steps' methodology.

6.2.2.8 CABSFLEX (Conformational Alphabet-Based Structure Flexible Fitting) + Fireprot combination method:

The PVP structure was utilized as an input, using default parameters, which generated a list of amino acid numbers paired with their respective RMSF values. A selection was made of amino acids showing RMSF values exceeding 1. This list was compared to the mutant list generated by the Fireprot method. Mutants that corresponded to RMSF values >1 were identified and shortlisted.

Subsequently, the mutants shortlisted through this process were subjected to modeling and processed following the 'Common steps' methodology as previously outlined.

6.2.2.9 FOLDX position scan:

All suggested mutants from various methods, excluding those forming disulfide bonds, were compiled. Hotspots located more than 5Å away from the catalytic site were chosen, resulting in the initial list of mutants. Mutants showing positive $\Delta\Delta G$ values were removed, leading to the creation of the second list. The second list was then filtered based on a requirement of $\Delta\Delta G$ values less than -2kcal/mol, generating the third list of mutants. If a hotspot was mutated to multiple amino acids, such as G295 being altered to G295V, G295M, G295W, and G295F with $\Delta\Delta G < -2$ kcal/mol, it was included in the shortlisted mutants.

Following this selection process, the mutants identified through this method were further subjected to modeling and processed using the steps outlined in the 'Common steps' methodology.

6.2.2.10 MDS + FOLDX Position Scan:

The process began by identifying residues with RMSF values exceeding 0.2nm. This initial list of hotspots was then refined by considering their proximity to the catalytic site, excluding those within <5Å. The resulting list of mutants underwent further filtering based on the $\Delta\Delta G$ value obtained from the FOLDX position scan. Only residues showing a $\Delta\Delta G$ value of <-2kcal/mol were included in the shortlisted mutants.

Following this selection process, the mutants identified through this method were subjected to modeling and processed using the steps outlined in the 'Common steps' methodology.

6.2.3 Results:

6.2.3.1 Disulfide by Design (D2D):

The web-based tool Disulfide by Design 2.0 (Craig & Dombkowski, 2013), was employed to identify specific amino acids within PVP that could be exchanged with cysteine, creating new disulfide connections. This introduction of fresh disulfide bonds aimed to enhance the conformational stability of PVP. As a result of this analysis, two mutants were shortlisted: N282C, V410C with a ΔG of -89.49 (kcal/mol), and Y394C R347C with a ΔG of -91.38 (kcal/mol) **Table 6. 1.**

Chapter 6

Table 6. 1 Disulfide by Design (D2D) was used to identify hotspots in PVP. Note: Wildtype ΔG : -87.4. The final selected mutant is highlighted in **Bold and underlined**.

Res. No.	Res. Amino acid	Res2. no.	Res2 Amino acid	Native(Y/N)	Distance > 20 amino acid	Any one of the two aa close to catalytic sites?	$\Delta G(\text{kcal/mol})$	Final selected mutants
70	LEU	117	ALA	N	Y	Y		
71	THR	74	GLY	N	N	Y		
99	CYS	130	CYS	Y	-	Y		
109	ALA	119	ALA	N	N	Y		
110	SER	115	THR	N	N	Y		
123	THR	132	VAL	N	N	Y		
190	CYS	196	CYS	Y	-	Y		
274	GLN	406	ASP	N	Y	N	-80.81	
282	ASN	410	VAL	N	Y	N	-89.49	<u>N282C/V410C</u>
323	TYR	327	ASN	N	N	Y		
347	ARG	394	TYR	N	Y	N	-91.38	<u>R347C/Y394C</u>
383	CYS	392	CYS	Y	-	Y		

6.2.3.2 Consensus Finder + Consurf + manual consensus (CCM) method:

Consensus Finder (Hua et al., 2020) aids in identifying potential stability-enhancing hotspots within PVP. The process involves generating multiple sequence alignments of similar sequences. The initial list of identified mutants comprised S293P, F238L, T283Q, G360Q, V299L, S292P, H173A, T151G, and G295V. To refine this list, Consurf (Celniker et al., 2013) was employed, resulting in a second set of potential mutants: F238L, V299L, S292P, H173A, T151G, G295V. Following this, the Manual consensus method was applied, utilizing ClustalO for aligning PVP with reference sequences such as 1DKM (*Escherichia coli* phytase), 3ZHC (*Citrobacter braakii* phytase), 4ARV (*Yersinia kristensenii* phytase), and 4ARS (*Hafnia alvei* phytase). Comparison between the second list of PVP hotspots and residues in the references led to the formation of a third list of mutants, further refined based on ΔG (kcal/mol) (Table 6. 2). Ultimately, the CCM method yielded the final shortlisted mutants: S292G, T151G, and H173A.

6.2.3.3 Literature:

This approach involved altering PVP using mutations previously conducted on *E. coli* phytase to improve its thermostability. The mutations specified in the literature were duplicated in the equivalent positions of PVP. A selection of mutants was curated based on their ΔG (kcal/mol) values, encompassing both multi-point mutations and single-point mutations, such as M1 (E85W/T97C/L178Y/N212C/R350N) and M2 (A163C, Y271C) (Table 6. 3)

6.2.3.4 Hotspot wizard

The Hotspot Wizard tool (Bendl et al., 2016) was employed to detect potential residues using Functional, Structural Flexibility, and Sequence Consensus features. None of the mutants identified through Functional hotspot analysis met the specified criteria outlined in the methodology. In the case of structural flexibility analysis, 10 potential hotspots were initially pinpointed. However, after applying the methodology's filtering criteria, only one mutant, L336F, met the requirements. The list obtained through the sequence consensus feature of stability hotspot analysis underwent evaluation using an internal mutation feature, aiding in the selection of the most stable mutants. A total of two potential hotspots emerged from this process. Among these, one mutant, S293P displayed a ΔG (kcal/mol) of -91.64, falling below the -87.4 (WT) threshold. Consequently, the final shortlisted mutant from this analysis was S293P (Table 6. 4) .

Chapter 6

Table 6. 2 Consensus Finder, Consurf, and manual consensus (CCM) method was used to identify hotspots in PVP. *Note: Wildtype ΔG : -87.4.*

The final selected mutant is highlighted in **Bold**

Consensus Finder			Consurf			Manual Consensus				Mutant (ΔG)
Residue	Position	Mutant	Buried/Exposed	Conservation Score (≤ 6)	Selected (Y or N)	1DKM_A	3ZHC_A	4ARV_A	4ARS_A	
S	293	P	E	8	N	L	L	P	P	
F	238	L	B	4	Y	M	M	M	M	F238L (-87.85)
										F238M (-87.92)
T	283	Q	B	8	N	L	M	Q	T	
G	360	Q	E	8	N	Q	D	Q	Q	
V	299	L	B	6	Y	F	F	F	F	V299L (-85.82)
										V299F (-87.56)
S	292	P	E	1	Y	P	G	R	I	S292P (-84.87)
										S292G (-89.15)
										S292R (-83.23)
										S292I (-83.70)
H	173	A	E	3	Y	D	L	E	T	H173A (-89.20)
										H173D (-86.22)
										H173L (-88.01)
										H173E (-86.39)
										H173T (-87.44)
T	151	G	E	3	Y	G	G	G	G	T151G (-79.92)
G	295	V	B	5	Y	T	V	T	N	G295V (-86.63)
										G295T (-86.92)
										G295N (-88.76)

Chapter 6

Table 6. 3 Literature-based *in-silico* identification of hotspots in PVP. Note: Wildtype ΔG : -87.4. The final selected mutant for experimental validation is highlighted in **Bold and underlined**.

Reference paper	<i>E.coli</i>		PVP		Mutant	Multiple /Point	Mutant (ΔG after repair)
	Target residue	Position	Target residue	Position			
1: (Li et al., 2019b)	S	342	R	343	T	Multiple	R343T/R384A/T386A/T64W/T65S/A66L (-87.76)
	E	383	R	384	A		
	E	384	V	385	V		
	R	385	T	386	A		
	S	80	A	102	A		
	P	41	T	64	W		
	V	42	T	65	S		
	K	43	A	66	L		
	Q	285	-	-	-		
	K	286	-	-	-		
2. (Kim and Lei, 2008b)	K	46	P	46	E	Point	P46E (-87.52)
	K	46	P	46	E	Multiple	P46E/T97M/D195G (-84.23)
	K	97	T	97	M		
	S	209	D	195	G		
3. (Chen & Tw, 2018)	A	143	A	163	C	Multiple	<u>A163C/Y271C (-89.07)</u>
	E	262	Y	271	C	Multiple	D268C/T312C (-84.88)
	R	259	D	268	C		
	G	312	T	312	C	Multiple	S213C/S266C (-84.42)
	V	205	S	213	C		
	L	257	S	266	C		

Chapter 6

	A	264	A	273	C	Multiple	A273C/M309C (-80.96)
	N	309	M	309	C		
	A	143	A	163	C	Multiple	A163C/Y271C/D268C/T312C (-84.2)
	E	262	Y	271	C		
	R	259	D	268	C		
	G	312	T	312	C		
	A	143	A	163	C	Multiple	A163C/Y271C/A273C (-81.95)
	E	262	Y	271	C		
	A	264	A	273	C		
	N	309	M	309	C		
4. (Wang et al., 2015)	W	46	E	69	E	Multiple	<u>E85W/T97C/L178Y/N212C/R350N</u> <u>(-91.19)</u>
	Q	62	E	85	W		
	A	73	DASH		P		
	K	75	T	97	C		
	S	146	DASH		E		
	R	159	L	178	Y		
	N	204	N	212	C		
	Y	255	D	264	D		
	Q	258	N	267	N		
Q	349	R	350	N			
5. (Fakhravar & Hesampour, 2018)	S	392	P	393	F	Point	P393F (-85.6)
6. (J. Zhang et al., 2016)	K	24	P	46	E	Point	P46E (-87.52)
7. (J. Zhang et al., 2016)	W	46	E	69	E	Multiple	P45E (-87.52)

Table 6. 4 Hotspot Wizard (HS) was used for identification of hotspots in PVP.

Note: Wildtype ΔG : -87.4. The final selected mutant is highlighted in **Bold and underlined**.

Stability hotspot by structural flexibility				
Hotspot		Amnio acid frequency	Mutational landscape	Mutant (ΔG)
number	residue			
36	Leu	Leu	Val	
37	Ser	Ser	Val	
335	Val	Val	Leu	
34	Val	Val	Phe	
119	Ala	Ala	Gly	
330	Pro	Pro	Gly	
334	Leu	Leu	Met	
336	Leu	Phe	Phe	L336F (-86.7)
40	Gly	Gly	Asn	
68	Gly	Gly	Cys	
Stability hotspot by Sequence				
Amino acid	Amino acid No.	Amnio acid frequency	-	Mutant (ΔG)
SER	293	Pro	-	<u>S293P (-91.64)</u>
THR	61	Pro	-	T61P (-86.67)

6.2.3.5 Fireprot

In this study, Fireprot (Musil et al., 2017) was utilized to identify potential residues based on energy, evolution, and a combination of both factors. Within the combined mutant section, 26 mutations were listed. Only two mutants were selected from this list based on their highest stability, determined by the inbuilt FoldX (kcal/mol) and ROSETTA values (kcal/mol). Among these two, only one mutant, A181Y, showcased a ΔG of -92.35, falling below the -87.4 (WT) threshold, thus securing its position in the shortlist. Furthermore, 15 mutations were shortlisted based on evolutionary aspects, termed as evolution mutants. From this selection, three mutants were chosen based on their highest values (kcal/mol) in terms of FoldX, BTC by majority, and BTC by ratio criteria. All three mutants displayed a ΔG (kcal/mol) below -87.4 kcal/mol (WT ΔG value). Consequently, the final list derived from the Fireprot online server included: A201Y, A181Y, S293P, and G360Q (**Table 6. 5**).

Table 6. 5 Fireprot was used for *in-silico* identification of hotspots in PVP:

Note: Wildtype ΔG : -87.4. The final selected mutant is highlighted in **Bold and underlined**.

Combined Mutant		
Mutant	Highest stability w.r.t FOLDX/ ROSETTA values	ΔG
<u>A201Y</u>	-2.69	-89.85
<u>A181Y</u>	-12.3	-92.35
Evolution Mutant		
Mutant	Highest stability w.r.t FOLDX/ (YES in both BTC BY MAJORITY + BY RATIO values)	ΔG
<u>S293P</u>	-2.18	-91.64
<u>G360Q</u>	Y	-90.64
T61P	Y	-86.67

6.2.3.6 Common hotspot wizard and firepot (HS-FP) method:

A fusion of Fireprot and Hotspot Wizard servers was employed to locate common hotspots identified by both servers. Through this method, a total of three potential hotspots were pinpointed. Among these, only two mutants, N235G and S293P, demonstrated a ΔG (kcal/mol) below -87.4 kcal/mol (**Table 6. 6**).

Table 6. 6 Common mutations from Hotspot Wizard and Fireprot analysis: The final selected mutant is highlighted in Bold and underlined. Note: Wildtype ΔG : -87.4.

Mutant	Mutant (ΔG)
<u>N235G</u>	-90.99
T61P	-86.67
<u>S293P</u>	-91.64

6.2.3.7 CABSFLEX (Conformational Alphabet-Based Structure Flexible Fitting) + Fireprot combination method (Jamroz et al., 2013):

Two distinct methods were combined to streamline the selection of potential mutants. Initially, CABSFLEX was employed to determine the RMSF values for each residue, which were subsequently cross-referenced with the list of mutants identified by Firepot.

The common hotspots identified from the combined mutant section and energy section (Fireprot analysis) were shortlisted. This list was compared to the RMSF values obtained from CABSFLEX. Residues with RMSF values greater than 1 were selected, resulting in a total of 7 mutants. Among these, only 4 were shortlisted based on a ΔG (kcal/mol) of less than -87.4 kcal/mol.

Likewise, the common hotspots listed in the combined and evolution sections were merged and compared with the RMSF values obtained from CABSFLEX. Residues with RMSF values exceeding 1 were identified, totaling 5 hotspots. Out of these, only two, namely T151G and N235G, displayed a ΔG (kcal/mol) lower than -87.4 kcal/mol. Furthermore, the list generated from the HS-FP method was also matched with CABSFLEX data. Two mutants, S293P and N235G, were shortlisted based on their RMSF values and ΔG (kcal/mol) (**Table 6. 7**).

Table 6. 7 CABSFLEX was used for *in-silico* identification of hotspots in PVP. Note: Wildtype ΔG : -87.4. The final selected mutant is highlighted in Bold and underlined.

fp combined+ energy		
Mutant	RMSF	ΔG
<u>Q58M</u>	1.66	-87.75
<u>H173A</u>	1.365	-89.2
<u>A201Y</u>	1.68	-89.85
N267Y	3.28	-89.35
A290P	1.826	-81.8
G295T	2.257	-86.92
N407W	1.558	-88.88
fp combined+ evolution		
T61P	1.617	-86.67
<u>T151G</u>	1.335	-79.92
<u>N235G</u>	1.13	-90.99
F238L	1.196	-87.85
I367L	1.953	-88.36

6.2.3.8 FOLDX position scan:

The PVP underwent a FOLDX position scan, generating an initial list of 28 mutants selected based on a criterion of $\Delta\Delta G < -2$. Subsequently, the amino acids from this initial list that were mutated by more than one amino acid were further screened, leading to a refined second list of mutants. This final selection comprised 17 mutants in total, as detailed in **Table 6. 8**.

Table 6. 8 FOLDX position scan used for *in-silico* identification of hotspots in PVP.

Mutants	$\Delta\Delta G < -2$
G295V	-2.43678
G295M	-3.16662
G295W	-2.69223
G295F	-3.3858
E85L	-2.71522
E85F	-2.59881
E85I	-2.26731
E85M	-2.64014
A188Y	-2.60888
A188F	-2.56215
A201Y	-2.95649
A201F	-3.28864
G360L	-2.17048
G360M	-2.41339
G360F	-2.17897
S293M	-3.67148
S293I	-2.97267

6.2.3.9 MDS + FOLDX Position Scan:

The initial list of mutants was filtered based on RMSD values greater than 0.2. Subsequently, this filtered list underwent a distance assessment from the catalytic site, resulting in a second list of shortlisted mutants. These selected mutants were then cross-referenced with the $\Delta\Delta G$

Chapter 6

values obtained from the FOLDX position scan, specifically targeting values of $\Delta\Delta G < -2$. Consequently, a total of 5 mutants were identified and shortlisted: T64Y, T151P, V152M, V152F, and T247R (**Table 6. 9**).

Table 6. 9 Common hotspots identified by using molecular dynamics simulations (MDS) + FOLDX position scan analysis:

Mutant	$\Delta\Delta G > -2$
T64Y	-4.04117
T151P	-2.15643
V152M	-2.53173
V152F	-2.35649
T247R	-2.18325

6.3 Experimental validation:

6.3.1 Introduction:

Experimental validation was centered on examining the stability of the wildtype (WT) and mutant (M1 and M2) phytases by inducing their over-expression within the *E.coli* BL21(DE3) host system, after which they underwent detailed biochemical and biophysical characterization (**Figure 6. 2**).

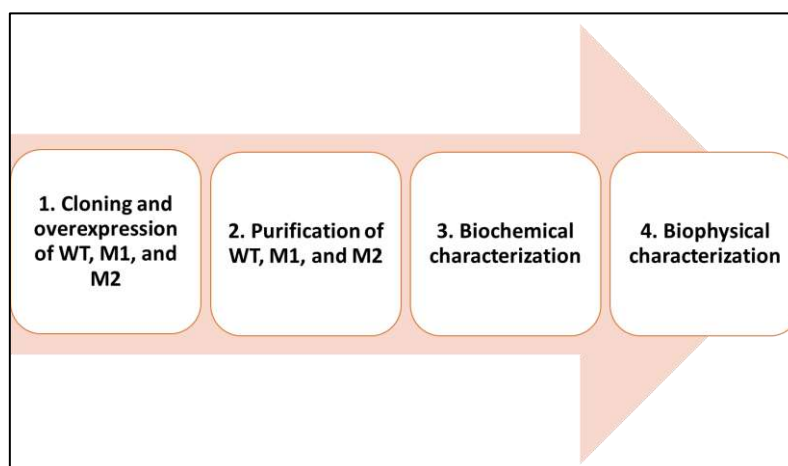


Figure 6. 2 Overview of the experimental validation of WT, M1, and M2.

6.3.2 Materials and Methods:

6.3.2.1 Materials (Chemicals, kits, reagents, solutions, plasticware):

PureLink® Genomic DNA kit (Cat. No. K182001), Primers (Eurofins), Strain: *E.coli* BL21(DE3) was used as a host for PVP over-expression, Media: Luria-Bertani (LB) broth and agar purchased from HiMedia Pvt.LTD, primary antibody: His-Tag (D3I1O) XP® Rabbit mAb #12698 purchased from Cell Signalling Technology (CST), gene synthesis from Biomatik, All chemicals were procured from either Sigma-Aldrich, MP Biomedicals, Loba Chemie, or MolyChem. All plasticware utilized in the experiments, including 15ml and 50ml falcon tubes, were obtained from SPL Life Sciences Pvt. Ltd., India.

6.3.2.2 Codon optimization, cloning, overexpression, and purification of Wildtype (WT), M1*, and M2*:

*Note:

a. M1: E85W/T97C/L178Y/N212C/R350N

b. M2: A163C/Y271C

The genetic sequences of WT, M1, and M2 were optimized for *E. coli*, synthesized chemically, and cloned into the pET29b vector between the Nco1 and Not1 restriction sites by Biomatik. These synthesized constructs (WT, M1, and M2) were lyophilized, reconstituted in TE buffer, and stored at -20°C for long-term preservation. The constructs, at a concentration of 10 ng/μl, were introduced into the *E.coli* BL21(DE3) expression host system according to the manufacturer's guidelines. The transformed cells carrying the WT, M1, and M2 constructs were cultured on LB media supplemented with 50μg/ml of Kanamycin, establishing three separate *E.coli* BL21(DE3) systems: WT, M1, and M2.

All three systems were induced with 1mM IPTG, 30°C, overnight, followed by purification of wildtype and mutant phytases (WT, M1, and M2) as per the procedure mentioned in Chapter 5 (methodology section: 5.3.2.4).

The purified WT and M2 phytase samples were separated using a 10% SDS-PAGE gel and subsequently transferred onto a polyvinylidene fluoride (PVDF) membrane from Bio-Rad, USA. A blocking buffer consisting of 5% non-fat dry milk (NFDM) and 0.1% Tween 20 in 1X Tris-buffered saline (TBST) was applied to the membrane for 1.5 hours at room temperature (RT) to block nonspecific binding. The primary antibody used was the His-Tag (D3I1O) XP®

Chapter 6

Rabbit mAb, which targeted and bound to the His-tag present on the WT and M2 phytases. Following overnight incubation at 4°C with the primary antibody, the membrane underwent three washes with TBST for 10 minutes each to remove any unbound antibodies. Subsequently, the membrane was exposed to a secondary antibody, the goat anti-rabbit HRP linked secondary antibody (GeNei™, India), for 1 hour at RT. After another three washes with TBST for 10 minutes each, the Clarity solution from Bio-Rad, USA, was applied to develop the blot. Visualization of the blot was achieved using the ChemiDoc XRS+ Gel Imaging System, also from Bio-Rad, USA.

6.3.2.3 Biochemical characterization:

Standard KH_2PO_4 reaction was set up as per the procedure mentioned in Chapter 5 (section 5.3.2.5). The standard reaction setup to estimate phytase activity of purified WT and M2 was performed as per the procedure mentioned in Chapter 5 (section 5.3.2.5). 7µl of purified WT and M2 were used in the assay.

Effect of pH on phytase activity:

To determine the optimum pH for PVP, various buffer systems were employed in the standard assay, including 0.1M Glycine-HCl (pH 2 and 3), 0.1M Sodium acetate (pH 4, 4.5, 5, 5.5), 0.1M Tris-acetate (pH 6 and 7), and 0.1M Tris-base (pH 8). The resulting graph depicted the relationship between % relative specific activity and pH.

Effect of temperature on phytase activity:

To determine the optimum temperature, the standard assay was conducted at various temperatures ranging from 30 to 70°C. The resulting graph illustrated the correlation between % relative specific activity and temperature (°C).

Effect of substrate concentration on phytase activity:

To estimate the K_m of PVP, varying concentrations of substrate (4, 2, 1, 0.5, 0.25, and 0.1 µmoles/ml) were employed in the standard assay.

6.3.2.4 Biophysical characterization:

Isothermal titration calorimetry (ITC) experiment:

ITC experiments were performed using the ITC 200 (Micro cal, UK) instrument at 25°C. ITC cell was filled with 300 µl of 10-50µM recombinant purified phytase enzyme in buffer A (20 mM Tris, pH 7.0, 150 mM NaCl). The substrate phytic acid (sigma) was also dissolved in the

Chapter 6

same buffer A. The substrate concentration used for the ITC experiment varied up to from 500 μ M to 5mM final concentration. The data were analyzed by Malvern ITC analysis software.

Circular dichroism (CD) experiment:

Circular dichroism of recombinant purified WT and M2 was done with 10-50 μ M protein concentration in a quartz cuvette of 1mm path length. The CD data was analyzed with a Jasco CD spectrometer. CD data were converted to mean residual ellipticity and corrected for differences in concentration (Greenfield, 2007). For UV 195-250 nm was used for protein folding measurements.

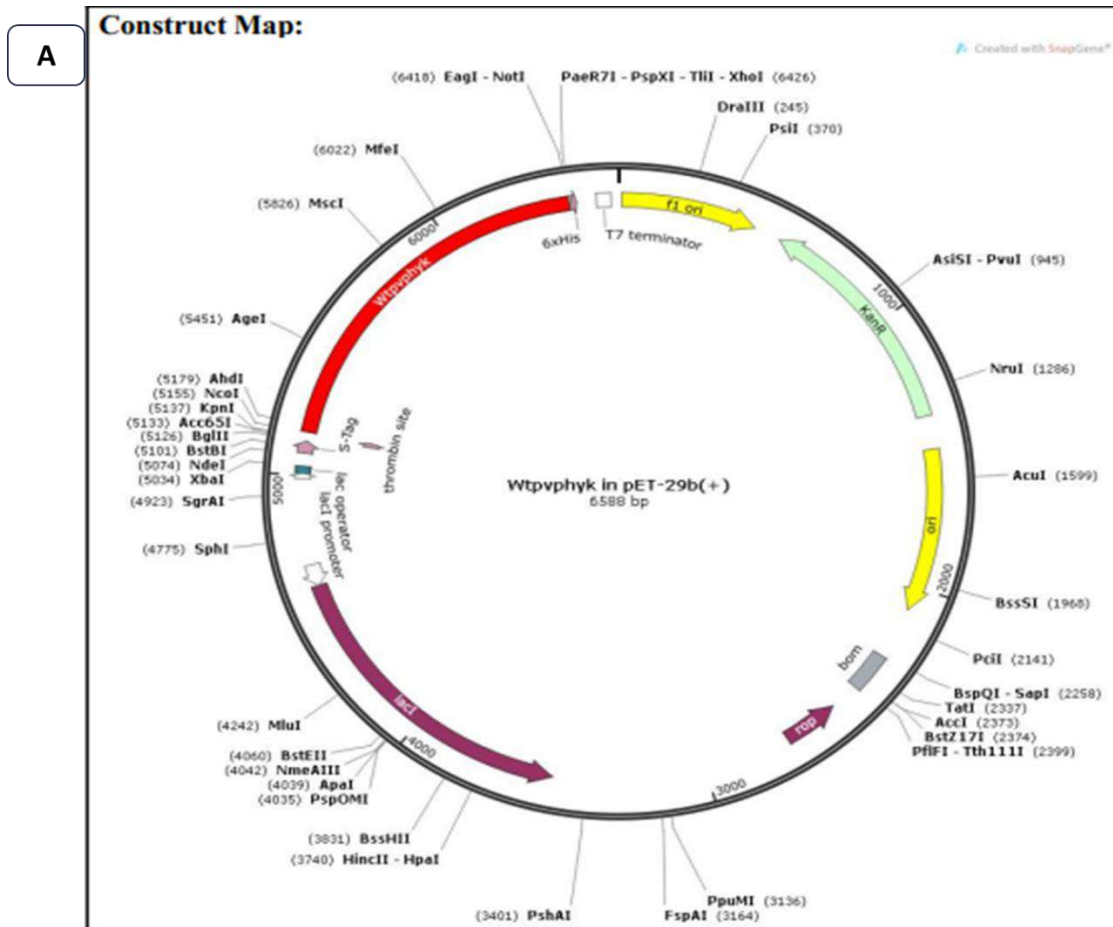
Differential scanning fluorimetry (DSF) experiment:

Differential scanning fluorimetry is also known as thermal shift assay. DSF experiment was carried out in 20 mM Tris pH 7.0 with the addition of 1-5X SYPRO Orange (Life Technologies). The assay was performed on a white 96-well plate covered with optical tape (Bio-Rad). The total reaction volume per well was 25 μ L with 1-3 μ g of purified phytase enzyme respectively, consisting of 1 μ L of SYPRO orange solution. The final ion concentration used was 1-3mM in the reaction mixture along with ligand as well. The relative fluorescence intensity (RFU) was recorded with the C1000 touch thermal cycler (Bio-Rad) with the CFX-96 real-time system. The temperature was raised from 20 to 90 $^{\circ}$ C by 0.5 $^{\circ}$ C per step. Raw data from melting curves were exported from the CFX manager software to the micro Microsoft Excel and presented using the same program. The melting temperature (T_m) for each condition was obtained from the first derivative of melting curves.

6.3.3 Result

6.3.3.1 Overexpression and purification of wildtype and shortlisted mutant (M1 and M2) phytases:

In this investigation, both wildtype and mutant genes were synthesized, undergoing codon optimization before being cloned into the pET29-b vector (depicted in **Figure 6. 3**, **Figure 6. 4**, **Figure 6. 5**).



B.

```

1   TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGGC
61  CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC
121 CTTTCTCGCC ACGTTCGCCG GCTTTCCTCCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG
181 GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCAGAAAAA CTTGATTAGG GTGATGGTTC
241 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT
301 CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACTATC AACCCATCT CGGTCTATTC
361 TTTTGATTTA TAAGGGATTT TGCCGATFTC GGCTATTTGG TTAATAAATG AGCTGATTTA
421 ACAAAAATTT AACCGGAATT TTAACAAAAT ATTAACGTTT ACAATTTTTCAG GTGGCACTTT
481 TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA
541 TCCGCTCATG AATTAATTCT TAGAAAAACT CATCGAGCAT CAAATGAAA TGCAATTTAT
601 TCATATCAGG ATTATCAATA CCATATTTT GAAAAAGCCG TTTCTGTAAT GAAGGAAAAA
661 ACTCACCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCGCACTC
721 GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAA AATAAGTTA TCAAGTGAGA
781 AATCACCATG AGTGACGACT GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC
841 AGACTTGTTT AACAGGCCAG CCATTACGCT CGTCATCAA ATCACTCGCA TCAACCAAAC
901 CGTTATTCAT TCGTGATTGC GCCTGAGCGA GACGAAATAC GCGATCGCTG TAAAAGGAC
961 AATTACAAAC AGGAATCGAA TGCAACCGGC GCAGGAACAC TGCCAGCGCA TCAACAATAT
    
```

Chapter 6

1021	TTTCACCTGA	ATCAGGATAT	TCTTCTAATA	CCTGGAATGC	TGTTTTCCCG	GGGATCGCAG
1081	TGGTGAGTAA	CCATGCATCA	TCAGGAGTAC	GGATAAAATG	CTTGATGGTC	GGAAGAGGCA
1141	TAAATTCCGT	CAGCCAGTTT	AGTCTGACCA	TCTCATCTGT	AACATCATTG	GCAACGCTAC
1201	CTTTGCCATG	TTTCAGAAAC	AACTCTGGCG	CATCGGGCTT	CCCATACAAT	CGATAGATTG
1261	TCGCACCTGA	TTGCCCCGACA	TTATCGCGAG	CCCATTTATA	CCCATATAAA	TCAGCATCCA
1321	TGTTGGAATT	TAATCGCGGC	CTAGAGCAAG	ACGTTTCCCG	TTGAATATGG	CTCATAACAC
1381	CCCTTGATTT	ACTGTTTATG	TAAGCAGACA	GTTTTATTGT	TCATGACCAA	AATCCCCTTAA
1441	CGTGAGTTTT	CGTTCCACTG	AGCGTCAGAC	CCCGTAGAAA	AGATCAAAGG	ATCTTCTTGA
1501	GATCCTTTTT	TTCTGCGCGT	AATCTGCTGC	TTGCAAACAA	AAAAACCACC	GCTACCAGCG
1561	GTGGTTTGT	TGCCGGATCA	AGAGCTACCA	ACTCTTTTTT	CGAAGGTAAC	TGGCTTCAGC
1621	AGAGCGCAGA	TACCAAATAC	TGTCCTTCTA	GTGTAGCCGT	AGTTAGGCCA	CCACTTCAAG
1681	AACTCTGTAG	CACCCCTAC	ATACCCTGCT	CTGTAATCC	TGTTACCAGT	GCTTGTGCC
1741	AGTGGCGATA	AGTCGTGTCT	TACCGGGTTG	GACTCAAGAC	GATAGTTACC	GGATAAGGCG
1801	CAGCGGTCGG	GCTGAACGGG	GGGTTCTGTC	ACACAGCCCA	GCTTGGAGCG	AACGACCTAC
1861	ACCGAACTGA	GATACCTACA	GCGTGAGCTA	TGAGAAAAGCG	CCACGCTTCC	CGAAGGGAGA
1921	AAGGCGGACA	GGTATCCGGT	AAGCGGCAGG	GTCGGAACAG	GAGAGCGCAC	GAGGGAGCTT
1981	CCAGGGGGAA	ACGCCTGGTA	TCTTTATAGT	CCTGTCGGGT	TTCGCCACCT	CTGACTTGAG
2041	CGTCGATTTT	TGTGATGCTC	GTCAGGGGGG	CGGAGCCTAT	GGAAAAACGC	CAGCAACGCG
2101	GCCTTTTTAC	GGTTCCTGGC	CTTTTGCTGG	CCTTTTGCTC	ACATGTTCTT	TCCTGCGTTA
2161	TCCCCTGATT	CTGTGGATAA	CCGTATTACC	GCCTTTGAGT	GAGCTGATAC	CGCTCGCCGC
2221	AGCCGAACGA	CCGAGCGCAG	CGAGTCAGTG	AGCGAGGAAG	CGGAAGAGCG	CCTGATGCGG
2281	TATTTTCTCC	TTACGCATCT	GTGCGGTATT	TCACACCGCA	TATATGGTGC	ACTCTCAGTA
2341	CAATCTGCTC	TGATGCCGCA	TAGTTAAGCC	AGTATACACT	CCGCTATCGC	TACGTGACTG
2401	GGTCATGGCT	GCGCCCCGAC	ACCCGCCAAC	ACCCGCTGAC	GCGCCCTGAC	GGGCTTGTCT
2461	GCTCCCGGCA	TCCGCTTACA	GACAAGCTGT	GACCGTCTCC	GGGAGCTGCA	TGTGTCAGAG
2521	GTTTTTACCG	TCATCACCGA	AACGCGCGAG	GCAGCTGCGG	TAAAGCTCAT	CAGCGTGGTC
2581	GTGAAGCGAT	TCACAGATGT	CTGCCTGTTT	ATCCCGTCC	AGCTCGTTGA	GTTTCTCCAG
2641	AAGCTTAAT	GTCTGGCTTC	TGATAAAGCG	GCCATGTTA	AGGGCGGTTT	TTTTCTGTTT
2701	AGTCACTGAT	GCCTCCGTGT	AAGGGGAGT	TCTGTTTATG	GGGGTAATGA	TACCGATGAA
2761	ACGAGAGAGG	ATGCTCACGA	TACGGGTTAC	TGATGATGAA	CATGCCCGGT	TACTGGAACG
2821	TTGTGAGGGT	AAACAACCTGG	CGGTATGGAT	GCGGCGGGAC	CAGAGAAAAA	TCACTCAGGG
2881	TCAATGCCAG	CGCTTCGTTA	ATACAGATGT	AGGTGTTCCA	CAGGGTAGCC	AGCAGCATCC
2941	TGCGATGCAG	ATCCGGAACA	TAATGGTGCA	GGGCGCTGAC	TTCCGCGTTT	CCAGACTTTA
3001	CGAAACACGG	AAACCGAAGA	CCATTCATGT	TGTTGCTCAG	GTCGCAGACG	TTTTGCAGCA
3061	GCAGTCGCTT	CACGTTGCTT	CGCGTATCGG	TGATTCATTC	TGCTAACCAG	TAAGGCAACC
3121	CCGCCAGCCT	AGCCGGGTCC	TCAACGACAG	GAGCACGATC	ATGCGCACCC	GTGGGGCCCG
3181	CATGCCGGCG	ATAATGGCCT	GCTTCTCGCC	GAAACGTTTG	GTGGCGGGAC	CAGTGACGAA
3241	GGCTTGAGCG	AGGGCGTGCA	AGATTCCGAA	TACCGCAAGC	GACAGGCCGA	TCATCGTCCG
3301	GCTCCAGCGA	AAGCGGTCCCT	CGCCGAAAAAT	GACCCAGAGC	GCTGCCGGCA	CCTGTCCCTAC
3361	GAGTTGCATG	ATAAAGAAGA	CAGTCATAAG	TGCGGCGACG	ATAGTCATGC	CCCAGCCCA
3421	CCGGAAGGAG	CTGACTGGGT	TGAAGGCTCT	CAAGGGCATC	GGTCGAGATC	CCGTGCCCCA
3481	ATGAGTGAGC	TAACCTACAT	TAATTGCGTT	GCGCTCACTG	CCCGCTTCC	AGTCGGGAAA
3541	CCTGTCTGTC	CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCGTAT
3601	TGGGCGCCAG	GGTGGTTTTT	CTTTTACCA	GTGAGACGGG	CAACAGCTGA	TTGCCCTTCA
3661	CCGCTGGCC	CTGAGAGAGT	TGCAGCAAGC	GGTCCACGCT	GGTTTGCCCC	AGCAGGCGAA
3721	AATCCTGTTT	GATGGTGGTT	AACGGCGGGA	TATAACATGA	GCTGTCTTCG	GTATCGTCTG
3781	ATCCCACTAC	CGAGATGTCC	GCACCAACGC	GCAGCCCGGA	CTCGGTAATG	GCGCGCATTG
3841	CGCCCAGCGC	CATCTGATCG	TTGGCAACCA	GCATCGCAGT	GGGAACGATG	CCCTCATTCA
3901	GCATTTGCAT	GGTTTGTTGA	AAACCGGACA	TGGCACTCCA	GTCGCCTTCC	CGTTCGCTA
3961	TCGGCTGAAT	TTGATTGCGA	GTGAGATATT	TATGCCAGCC	AGCCAGACGC	AGACGCGCCG
4021	AGACAGAACT	TAATGGGCCC	GCTAACAGCG	CGATTTGCTG	GTGACCCAAT	GCGACCAGAT
4081	GCTCCACGCC	CAGTCGCGTA	CCGTCTTCAT	GGGAGAAAAAT	AATACTGTTG	ATGGGTGTCT
4141	GGTCAGAGAC	ATCAAGAAAT	AACGCCGGAA	CATTAGTGCA	GGCAGCTTCC	ACAGCAATGG
4201	CATCCTGGTC	ATCCAGCGGA	TAGTTAATGA	TCAGCCCACT	GACGCGTTGC	GCGAGAAGAT
4261	TGTGCACCGC	CGCTTTACAG	GCTTCGACGC	CGCTTCGTTT	TACCATCGAC	ACCACCACGC
4321	TGGCACCCAG	TTGATCGGCG	CGAGATTTAA	TCGCCGCGAC	AATTTGCGAC	GGCGGTGCA

Chapter 6

```

4381 GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCGCC AGTTGTTGTG
4441 CCACGCGGTT GGGAAATGTAA TTCAGCTCCG CCATCGCCGC TTCCACTTTT TCCCGCGTTT
4501 TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGAAAC GGTCTGATAA GAGACACCGG
4561 CATACTCTGC GACATCGTAT AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT
4621 CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTGCG CCATTCGATG GTGTCCGGGA
4681 TCTCGACGCT CTCCCTTATG CGACTCCTGC ATTAGGAAGC AGCCCAGTAG TAGGTTGAGG
4741 CCGTTGAGCA CCGCCGCCGC AAGGAATGGT GCATGCAAGG AGATGGCGCC CAACAGTCCC
4801 CCGGCCACGG GGCCTGCCAC CATACCACG CCGAAACAAG CGCTCATGAG CCCGAAGTGG
4861 CGAGCCCGAT CTTCCCCATC GGTGATGTCG GCGATATAGG CGCCAGCAAC CGCACCTGTG
4921 GCGCCGGTGA TGCCGGCCAC GATGCGTCCG GCGTAGAGGA TCGAGATCGA TCTCGATCCC
4981 GCGAAATTAA TACGACTCAC TATAGGGGAA TTGTGAGCGG ATAACAATTC CCTCTAGAA
5041 ATAATTTTGT TTAACTTTAA GAAGGAGATA TACATATGAA AGAAACCGCT GCTGCTAAAT
5101 TCGAACGCCA GCACATGGAC AGCCCAGATC TGGGTACCCT GGTGCCACGC GGTTCATGG
5161 GCATGACCCT GCCGACCCTG TGTCGTTGCG CCCTGATTCT GGGTAGCCTG TGGCTGCTGA
5221 GCCCGGCCAC CCAGGCAGCA GATTATCAGC TGGAAAAAGT GGTGGAAGT AGCCGTCATG
5281 GTGTGCGTCC GCCGACCCCG GGTAATCGCA AAGAAATGGA AGCAGCAAGC CAGCAGCCGT
5341 GGACCCAGTG GACCACCGCA GATGGCGAAC TGACCGGCCA CGGTTATAGC GCCGTTGTGA
5401 ATAAGGGTCG CTGGGAAGGC GAACATTATC GCCAGCTGGG TCTGCTGGGC ACCGGTTGTC
5461 CGGATGCCGC ACAGGTGTAT GTTCGCGCCA GCCCGCTGCA GCGCACCAGA GCAACCGCAG
5521 CAGCCCTGAC CGATGGTGCA TTTCCGGGCT GCGGCGTGAC CGTTCATCAT GTTGCAGGCG
5581 ATGTGGACCC TCTGTTTTCAG GGCGAAAAAC TGACCGTTAC CCGCACCAGT CCGGCCCAGG
5641 AACTGGCAGC CAAACAGCAG AAAGCAGGCG ATCTGGCACA TCTGCAGCAG CAGCTGCAGC
5701 CGGCAATTCA GCAGCTGAAA GCAGCCGTGT GTCCGCCGGC AACCGATTGT CCGCTGTTTG
5761 AAGCCCCGTG GACCTTTCGC CAGACCCGTA ATGGTAATAG TTATGTGTAT GGTCTGAGTG
5821 TGATGGCCAG TATGGTTGAA ACCCTGCGTC TGGGCTATAG TGAAAATCTG CCGTTTGATC
5881 AGCTGGCCTG GGGTCATATT ACCACCGCCG CCCAGATTAC CAGTCTGCTG CCGCTGCTGA
5941 CCGCAAATTA TGATCTGAGC AATGATGTTT TGTATCTGGC CCAGCGTCGT GGCAGTATTC
6001 TGCTGAATAC CATGCTGAA GCAATTGCCG CCGATAGTAG CCCGGGTCGC TGGCTGTTGC
6061 TGGTTGCCCA TGATACCAAT ATTGCAATGG TTCGCACCCT GATGGATTTT AATTGGCAGC
6121 TGCCGGGTTA TAGCCGTGGC AATATTCCGC CGGGCAGCAG TCTGGTGCTG GAACGTTGGC
6181 GTGATACCCG CAGCGGTGAA CGCTTCTGTC GCCTGTATTT TCAGGCACAG AGCCTGGATG
6241 GTATTCGTCA GCTGCAGCCT ATTGATGATA AACATCCGCT GCTGCGTCAG GAATGGCATC
6301 AGCCGGATTG CCGCGTTACC GCCGTTGGCC TGCTGTGCC GTATCAGAGC ACCCTGACCC
6361 AGCTGCGCAA AAATCTGGAT AATAGCGCAG TTCTGCCGGT TAGTGTGATT CTGCCGGCGG
6421 CCGCACTCGA GCACCACCAC CACCACCACT GAGATCCGGC TGCTAACAAA GCCCGAAAGG
6481 AAGCTGAGTT GGCTGCTGCC ACCGCTGAGC AATAACTAGC ATAACCCTT GGGCCTCTA

6541 AACGGGTCTT GAGGGGTTTT TTGCTGAAAAG GAGGAACTAT ATCCGGAT

```

C.

Optimized Protein_WT

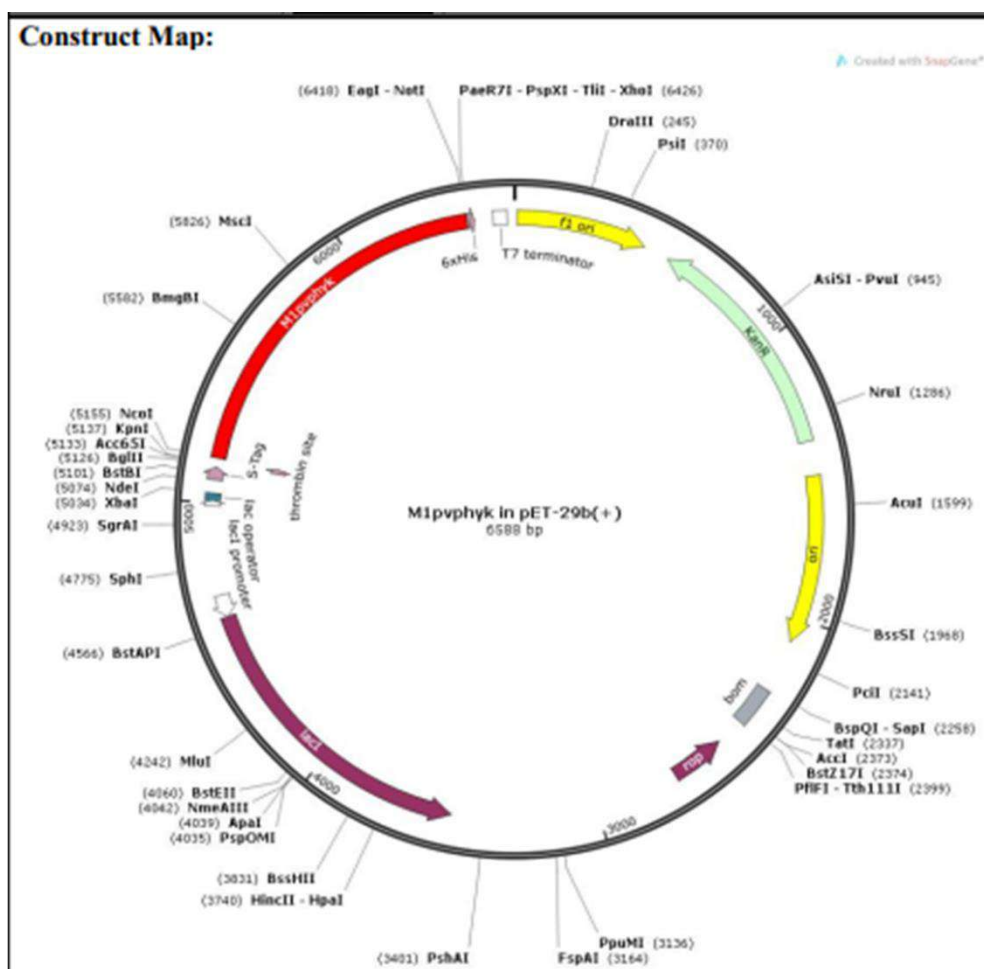
```

MTLPTLCRCALILGSLWLLSPATQAADYQLEKVVVELSRHGVRPPTPGNRKEMEAASQQP
WTQWTTADGELTGHGYSVVNKGRWEHEHYRQLGLLGTGCPDAAQVYVRASPLQRTR
ATAAALTDGAFPGCVTVHHVAGDVDPLFQGEKLVTRTDPAQELAAKQKAGDLAHL
QQQLQPAIQQLKAAVCPATDCPLFEAPWTFRQTRNGNSYVYGLSVMASMVETRLRGYS
ENLPFDQLAWGHITTAQAQITSLPLLTANYDLSNDVLYLAQRRGSILLNTMLEAIAADSSP
GRWLVLVAHDTNIAMVRTLMDFNWQLPGYSRGNIPPSSLVLERWRDTRSGERFLRLYF
QAQSLDGIRQLQPIDDKHPLLRQEWHPDCRVTAVGLLCPYQSTLTQLRKNLDNSAVLP
VSVILP

```

Figure 6. 3 A. The wildtype recombinant construct: A. codon optimized wildtype phytase gene cloned in pET29b vector. B. The complete nucleotide sequence of the wildtype recombinant construct (pET29b cloned with the optimized wildtype phytase gene (red highlighted)). C. the codon-optimized wild-type phytase protein sequence.

A



B.

```

1   TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGGC
61  CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCGCT CCTTTCGCTT TCTTCCCTTC
121 CTTTCTCGCC ACGTTCGCCG GCTTTCCTCC TCAAGCTCTA AATCGGGGGC TCCCTTTAGG
181 GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCAAAAAAA CTTGATTAGG GTGATGGTTC
241 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT
301 CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACTC AACCCTATCT CGGTCTATTC
361 TTTTGATTTA TAAGGGATTT TGCCGATTTT GGCTATTGG TTAATAAATG AGCTGATTTA
421 ACAAAAATTT AACCGGAATT TTAACAAAAT ATTAACGTTT ACAATTTT CAG GTGGCACTTT
481 TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA
541 TCCGCTCATG AATTAATTCT TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT
601 TCATATCAGG ATTATCAATA CCATATTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA
    
```

661 ACTCACCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC
 721 GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGTGAGA
 781 AATCACCATG AGTGACGACT GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC
 841 AGACTTGTTT AACAGGCCAG CCATTACGCT CGTCATCAAA ATCACTCGCA TCAACCAAAC
 901 CGTTATTTCAT TCGTGATTGC GCCTGAGCGA GACGAAATAC GCGATCGCTG TTAAGAGGAC
 961 AATTACAAAC AGGAATCGAA TGCAACCGGC GCAGGAACAC TGCCAGCGCA TCAACAATAT
 1021 TTTTCACTGA ATCAGGATAT TCTTCTAATA CCTGGAATGC TGTTTTCCCG GGGATCGCAG
 1081 TGGTGAGTAA CCATGCATCA TCAGGAGTAC GGATAAAATG CTTGATGGTC GGAAGAGGCA
 1141 TAAATTCCGT CAGCCAGTTT AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC
 1201 CTTTGCCATG TTTCAGAAAC AACTCTGGCG CACTCGGCTT CCCATACAAT CGATAGATTG
 1261 TCGCACCTGA TTGCCCGACA TTATCGCGAG CCCATTTATA CCCATATAAA CTCAGCATCCA
 1321 TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTTCCCG TTGAATATGG CTCATAACAC
 1381 CCCTTGATAT ACTGTTTATG TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA
 1441 CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA
 1501 GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAAACA AAAAACCACC GCTACCAGCG
 1561 GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTTT CGAAGGTAAC TGGCTTCAGC
 1621 AGAGCGCAGA TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG
 1681 AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC
 1741 AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG
 1801 CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC
 1861 ACCGAAGTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA
 1921 AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT
 1981 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTGCGGT TTCGCCACCT CTGACTTGAG
 2041 CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG
 2101 GCCTTTTTTAC GGTTCCTGGC CTTTTGCTGG CTTTTGCTC ACATGTTCTT TCCTGCGTTA
 2161 TCCCTTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGTCGCCCGC
 2221 AGCCGAACGA CCGAGCGCAG CGAGCTAGTG ACGGAGGAAG CGGAAGAGCG CCTGATCGCG
 2281 TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA TATATGGTGC ACTCTCAGTA
 2341 CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG
 2401 GGTCATGGCT GCGCCCCGAC ACCCGCCAAC ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT
 2461 GCTCCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC GGGAGCTGCA TGTGTCAGAG
 2521 GTTTTCACCG TCATCACCGA AACCGCGGAG GCAGCTGCGG TAAAGCTCAT CAGCGTGGTC
 2581 GTGAAGCGAT TCACAGATGT CTGCCTGTTT ATCCGCGTCC AGCTCGTTGA GTTTCTCCAG
 2641 AAGCGTTAAT GTCTGGCTTC TGATAAAGCG GGCCATGTTA AGGGCGGTTT TTTCTGTTT
 2701 GGTCACTGAT GCCTCCGTGT AAGGGGGATT TCTGTTTATG GGGGTAATGA TACCGATGAA
 2761 ACGAGAGAGG ATGCTCACGA TACGGGTTAC TGATGATGAA CATGCCCGGT TACTGGAACG
 2821 TTGTGAGGGT AAACAACCTGG CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG
 2881 TCAATGCCAG CGCTTCGTTA ATACAGATGT AGGTGTTCCA CAGGGTAGCC AGCAGCATCC
 2941 TGCGATGCAG ATCCGGAACA TAATGGTGCA GGGCGCTGAC TTCCGCGTTT CCAGACTTTA
 3001 CGAAACACGG AAACCGAAGA CCATTCATGT TGTGCTCAG GTCGAGACG TTTTGCAGCA
 3061 GCAGTCGCTT CACGTTGCTC CGCGTATCGG TGATTCATTC TGCTAACCAG TAAGGCAACC
 3121 CCGCCAGCCT AGCCGGGTCC TCAACGACAG GAGCACGATC ATGCGCACCC GTGGGGCCCG

 3181 CATGCCGGCG ATAATGGCCT GCTTCTCGCC GAAACGTTTG GTGGCGGGAC CAGTGACGAA
 3241 GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC GACAGGCCGA TCATCGTTCG
 3301 GCTCCAGCGA AAGCGGTCTT CGCCGAAAAT GACCCAGAGC GCTGCCGGCA CCTGTCTTAC
 3361 GAGTTGCATG ATAAAGAAGA CAGTCATAAG TGCGGCGACG ATAGTCATGC CCCGCGCCCA
 3421 CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTGAGATC CCGGTGCCTA
 3481 ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGTCACTG CCCGTTTCC AGTCGGGAAA
 3541 CCTGTGCTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT
 3601 TGGGCGCCAG GGTGGTTTTT CTTTTTACCA GTGAGACGGG CAACAGCTGA TTGCCCTTCA
 3661 CCGCCTGGCC CTGAGAGAGT TGCAGCAAGC GGTCCACGCT GGTTTGCCCC AGCAGGCGAA
 3721 AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCG GTATCGTCTG
 3781 ATCCCACTAC CGAGATGTCC GCACCAACGC GCAGCCCGGA CTCGGTAATG GCGCGCATTG
 3841 CGCCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT GGAACGATG CCTCATTTCA
 3901 GCATTTGCAT GGTGTTGTTG AAACCGGACA TGGCACTCCA GTCGCCTTCC CGTTCCGCTA
 3961 TCGGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC AGCCAGACGC AGACGCGCCG

4021 AGACAGA AACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT
4081 GCTCCACGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAAT AATACTGTTG ATGGGTGTCT
4141 GGTCAGAGAC ATCAAGAAAT AACGCCGGA CATTAGTGCA GGCAGCTTCC ACAGCAATGG
4201 CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT GACGCGTTGC GCGAGAAGAT
4261 TGTGCACCGC CGCTTTACAG GCTTCGACGC CGCTTCGTTT TACCATCGAC ACCACCACGC
4321 TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC AATTTGCGAC GGC GCGTGCA
4381 GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCCGCC AGTTGTTGTG
4441 CCACGCGGTT GGAATGTAA TTCAGCTCCG CCATCGCCGC TTCCACTTTT TCCCGCGTTT
4501 TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC GGTCTGATAA GAGACACCGG
4561 CATACTCTGC GACATCGTAT AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT
4621 CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTGCG CCATTCGATG GTGTCCGGGA
4681 TCTCGACGCT CTCCTTATG CGACTCCTGC ATTAGGAAAG AGCCCAGTAG TAGGTTGAGG
4741 CCGTTGAGCA CCGCCGCCGC AAGGAATGGT GCATGCAAGG AGATGGCGCC CAACAGTCCC
4801 CCGGCCACGG GGCCTGCCAC CATACCACG CCGAAACAAG CGCTCATGAG CCCGAAAGTGG
4861 CGAGCCCGAT CTTCCCATC GGTGATGTCG GCGATATAGG CGCCAGCAAC CGCCAGTGTG
4921 GCGCCGGTGA TGCCGGCCAC GATGCGTCCG CCGTAGAGGA TCGAGATCGA TCTCGATCCC
4981 GCGAAATTAA TACGACTCAC TATAGGGGAA TTGTGAGCGG ATAACAATTC CCCTCTAGAA
5041 ATAATTTTGT TTAACTTTAA GAAGGAGATA TACATATGAA AGAAACCGCT GCTGCTAAAT
5101 TCGAACGCCA GCACATGGAC AGCCCAGATC TGGGTACCCT GGTGCCACGC GGTTCATGG
5161 GCATGACCTT GCCGACCTG TGTCGTTGCG CCCTGATTCT GGGTAGTCTG TGGCTGCTGA
5221 GCCCGGCAAC CCAGGCAGCC GATTATCAGC TGGAAAAAGT GGTGAACTG AGTCGTCATG
5281 GTGTGCGCCC GCCGACCCG GGTAAATCGTA AAGAAATGGA AGCCGCCAGT CAGCAGCCGT
5341 GGACCCAGTG GACCACCGCC GATGGTGAAC TGACCGGCCA CCGTTATAGC GCAGTGGTTA
5401 ATAAGGGTCG TTGGTGGGGT GAACATTATC GTCAGCTGGG TCTGCTGGGC TGCGGCTGTC
5461 CGGATGCCGC TCAGGTTTAT GTGCGTGCAA GTCCGCTGCA GCGCACCCGC GCAACCGCAG
5521 CAGCTCTGAC CGATGGCGCC TTTCCGGGCT GCGGTGTGAC CGTTCATCAT GTGGCCGGTG
5581 ACGTGGACCC TCTGTTTTCAG GCGCAAAAAAC TGACCGTGAC CCGTACCAGT CCGGCCCAGG
5641 AACTGGCAGC CAAACAGCAG AAAGCAGGCG ATCTGGCACA TCTGCAGCAG CAGTATCAGC
5701 CGGCCATTCA GCAGCTGAAA GCCGCAGTTT GCCCGCCGGC AACCGATTGT CCGCTGTTTG
5761 AAGCCCCGTG GACCTTTCGT CAGACCCGTA ATGGCTGCAG CTATGTGTAT GGTCTGAGCG
5821 TTATGGCCAG TATGGTTGAA ACCCTGCGT TGGGCTATAG TGAAAATCTG CCGTTTGATC
5881 AGCTGGCATG GGGTCATATT ACCACCGCAG CCCAGATTAC CAGTCTGCTG CCGCTGCTGA
5941 CCGCAAATTA TGATCTGAGC AATGATGTGC TGTATCTGGC CCAGCGTCGT GGTAGTATTC
6001 TGCTGAATAC CATGCTGGAA GCCATTGCAG CCGATAGTAG TCCGGGTGCG TGGCTGGTTC
6061 TGGTGGCCCA TGATACCAAT ATTGCAATGG TTCGTACCCT GATGGATTTT AATTGGCAGC
6121 TGCCGGGTTA TAGCCGTGGT AATATTCCGC CGGGTAGTAG CCTGGTTCTG GAACGTTGGC
6181 GCGATACCCG CAGCGGTGAA CGTTTTCTGA ATCTGTATTT TCAGGCACAG AGCCTGGATG
6241 GCATTCGTCA GCTGCAGCCG ATTGATGATA AACATCCGCT GCTGCGCCAG GAATGGCATC
6301 AGCCGGATTG TCGCGTTACC GCCGTTGGTC TGCTGTGTCC GTATCAGAGC ACCCTGACCC
6361 AGCTGCGCAA AAATCTGGAT AATAGCGCAG TTCTGCCGGT GAGCGTTATT CTGCCGGCGG
6421 CCGCACTCGA GCACCACCAC CACCACCACT GAGATCCGCG TGCTAACAAA GCCCGAAAGG
6481 AAGCTGAGTT GGCTGCTGCC ACCGCTGAGC AATAACTAGC ATAACCCCTT GGGCCTCTA
6541 AACGGGTCTT GAGGGGTTTT TTGCTGAAAAG GAGGAACTAT ATCCGGAT

B.

```

1   TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGCG
61  CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC
121 CTTTCTCGCC ACGTTCGCCG GCTTTCCTCC TCAAGCTCTA AATCGGGGGC TCCCTTTAGG
181 GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCCAAAAA CTTGATTAGG GTGATGGTTC
241 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT
301 CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACACTC AACCCATATC CGGTCTATTC
361 TTTTGATTTA TAAGGGATTT TGCCGATTTT GGCCTATTGG TTAATAAATG AGCTGATTTA
421 ACAAAAATTT AACCGGAATT TTAACAAAAT ATTAACGTTT ACAATTTTCAG GTGGCACTTT
481 TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTT TAAATACATT CAAATATGTA
541 TCCGCTCATG AATTAATTCT TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT
601 TCATATCAGG ATTATCAATA CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA
661 ACTCACCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC
721 GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGTGAGA
781 AATCACCATC AGTGACGACT GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC
841 AACTTGTGTT AACAGGCCAG CCATTACGCT CGTCATCAA ATCACTCGCA TCAACCAAAC
901 CGTTATTTCAT TCGTGATTGC CCCTGAGCGA GACGAAATAC GCGATCGCTG TTAATAAGGAC
961 AATTACAAAC AGGAATCGAA TGCAACCGGC GCAGGAACAC TGCCAGCGCA TCAACAATAT
1021 TTTACACTGA ATCAGGATAT TCTTCTAATA CCTGGAATGC TGTTTTCCCG GGGATCGCAG
1081 TGGTGAGTAA CCATGCATCA TCAGGAGTAC GGATAAAATG CTTGATGGTC GGAAGAGGCA
1141 TAAATTCCGT CAGCCAGTTT AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC
1201 CTTTGCCATG TTTTCAAAAAC AACTCTGGCG CATCGGGCTT CCCATACAAT CGATAGATTG
1261 TCGCACCTGA TTGCCCGACA TTATCGCGAG CCCATTTATA CCCATATAAA TCAGCATCCA
1321 TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTCCCG TTGAATATGG CTCATAACAC
1381 CCCTTGATAT ACTGTTTATG TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA
1441 CGTGAGTTTT CGTTCCTACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA
1501 GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG
1561 GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTT CGAAGGTAAC TGGCTTCAGC
1621 AGAGCGCAGA TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG
1681 AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTACCAGT GGCTGCTGCC
1741 AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG
1801 CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCA GCTTGAGAGC AACGACCTAC
1861 ACCGACTTGA GATACCTACA GCGTAGCCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA
1921 AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT
1981 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCCGGT TTCGCCACCT CTGACTTGAG
2041 CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG
2101 GCCTTTTTTAC GGTTCCTGGC CTTTTGCTGG CTTTTGCTC ACATGTTCTT TCCTGCGTTA
2161 TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC
2221 AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG
2281 TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA TATATGGTGC ACTCTCAGTA
2341 CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG
2401 GGTCATGGCT GCGCCCCGAC ACCCGCCAAC ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT
2461 GTCCTCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC GGGAGCTGCA TGTGTCAGAG
2521 GTTTTTCACCG TCATCACCGA AACGCGCGAG GCAGCTGCGG TAAAGCTCAT CAGCGTGGTC
2581 GTGAAGCGAT TCACAGATGT CTGCCTGTTT ATCCGCTGCC AGCTCGTTGA GTTTCTCCAG
2641 AAGCGTTAAT GTCTGGCTTC TGATAAAGCG GGCCATGTTA AGGGCGGTTT TTTCTGTTT
2701 GGTCACTGAT GCCTCCGTGT AAGGGGGATT TCTGTTTATG GGGGTAATGA TACCGATGAA
2761 ACGAGAGAGG ATGCTCACGA TACGGGTAC TGATGATGAA CATGCCCGGT TACTGGAACG
2821 TTGTAGAGGT AAACAACCTG CCGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGG
2881 TCAATGCCAG GCCTTCGTTA ATACAGATGT AGGTGTTCCA CAGGGTAGCC AGCAGCATCC
2941 TGCGATGCAG ATCCGGAACA TAATGGTGCA GGGCGCTGAC TTCCGCGTTT CCAGACTTTA
3001 CGAAACACGG AAACCGAAGA CCATTCATGT TGTTGCTCAG GTCGCAGACG TTTTGCAGCA
3061 GCAGTCGCTT CACGTTGCTT CGCGTATCGG TGATTCATTC TGCTAACCCG TAAGGCAACC
3121 CCGCCAGCCT AGCCGGGTCC TCAACGACAG GAGCACGATC ATGCGCACCC GTGGGGCCCG

```

3181 CATGCCGGCG ATAATGGCCT GCTTCTCGCC GAAACGTTTG GTGGCGGGAC CAGTGACGAA
 3241 GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC GACAGGCCGA TCATCGTCGC
 3301 GCTCCAGCGA AAGCGGTCTT CGCCGAAAAAT GACCCAGAGC GCTGCCGGCA CCTGTCTTAC
 3361 GAGTTGCATG ATAAAGAAGA CAGTCATAAG TCGGGCGACG ATAGTCATGC CCCGCGCCCA
 3421 CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTGAGATC CCGGTGCCTA
 3481 ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA
 3541 CCTGTCTGTC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT
 3601 TGGGCGCCAG GGTGGTTTTT CTTTTACCA GTGAGACGGG CAACAGCTGA TTGCCCTTCA
 3661 CCGCTGGCC CTGAGAGAGT TGCAGCAAGC GGTCCACGCT GGTTCGCCCC AGCAGGCGAA
 3721 AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCG GTATCGTCGT
 3781 ATCCCACTAC CGAGATGTCC GCACCAACGC GCAGCCCGGA CTCGGTAATG GCGCGCATTG
 3841 CGCCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT GGGAACGATG CCCTCATTTCA
 3901 GCATTTGCAT GGTTCGTTGA AAACCCGACA TGGCACTCCA GTCGCCCTCC CGTTCGCTA
 3961 TCGGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC AGCCAGACGC AGACGCGCCG
 4021 AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT
 4081 GCTCCAGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAA AATACTGTTG ATGGGTGTCT
 4141 GGTACAGAGC ATCAAGAAAT AACCCGGAA CATTAGTGCA GGCAGCTTCC ACAGCAATGG
 4201 CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT GACGCGTTGC GCGAGAAGAT
 4261 TGTGCACCGC CGCTTTACAG GCTTCGACGC CGCTTCGTTT TACCATCGAC ACCACCACGC
 4321 TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC AATTTGCGAC GGCGCGTGCA
 4381 GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCGCC AGTTGTTGTG
 4441 CCACGCGGTT GGAATGTAA TTCAGCTCCG CCATCGCCGC TTCCACTTTT TCCCGCGTTT
 4501 TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC GGTCTGATAA GAGACACCGG
 4561 CATACTCTGC GACATCGTAT AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT
 4621 CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTGCG CCATTCGATG GTGTCCGGGA
 4681 TCTCGACGCT CTCCTTATG CGACTCCTGC ATTAGGAAGC AGCCAGTAG TAGGTTGAGG
 4741 CCGTTGAGCA CCGCCGCGC AAGGAATGGT GCATGCAAGG AGATGGCGCC CAACAGTCCC
 4801 CCGGCCACGG GGCCTGCCAC CATACCACG CCGAAAACAAG CGCTCATGAG CCCGAAGTGG
 4861 CGAGCCCGAT CTTCCCATC GGTGATGTCG GCGATATAGG CGCCAGCAAC CGCACCTGTG
 4921 GCGCCGGTGA TGCCGGCCAC GATGCGTCCG GCGTAGAGGA TCGAGATCGA TCTCGATCCC
 4981 GCGAAATTAA TACGACTCAC TATAGGGGAA TTGTGAGCGG ATAACAATTC CCTCTAGAA
 5041 ATAATTTTGT TTAACTTTAA GAAGGAGATA TACATATGAA AGAAACCGCT GCTGCTAAAT
 5101 TCGAACGCCA GCATGGAC AGCCAGATC TGGGTACCCT GGTGCCACGC GGTCCATGG
 5161 GCATGACCTT CCCGACCCTG TCCCGTTGTG CACTGATTCT GGGCAGCCTG TGGCTGCTGA
 5221 CCCCAGCAAC CCAGGACGCC GATTATCAGC TGGAAAAAGT TGTGGAAGT AGTCGTCATG
 5281 GCGTGCCTCC GCCGACCCCG GGTAAATCGTA AAGAAATGGA AGCAGCAAGC CAGCAGCCGT
 5341 GGACCCAGTG GACCACCGCA GATGGTGAAC TGACCCGGCA CCGTTATAGC GCCGTGGTGA
 5401 ATAAGGGCCG CTGGGAAGGC GAACATTATC GCCAGCTGGG TCTGCTGGGT ACCGGTTGCC
 5461 CGGATGCCGC CCAGGTGTAT GTTCGTGCCA GCCCCTGCA GCGCACCCGT GCAACAGCAG
 5521 CCGCACTGAC CGATGGTGCC TTTCCGGGTT GTGGTGTGAC CGTGCATCAT GTTGCAGGCG
 5581 ATGTTGATCC GCTGTTTCAG GGTGAAAAAC TGACCGTGAC CCGTACCGAT CCGGCACAGG
 5641 AACTGGCATG TAAACAGCAG AAAGCCGGTG ACCTGGCACA TCTGCAGCAG CAGCTGCAGC
 5701 CGGCAATTCA GCAGCTGAAA GCCGCCGTGT GCCCGCCGCG TACCGATTGT CCGCTGTTTG
 5761 AAGCCCCGTG GACCTTTCGC CAGACCCGCA ATGGTAATAG TTATGTGTAT GGTCTGAGCG
 5821 TTATGGCCAG CATGGTTGAA ACCCTGCGCC TGGGTATATAG CGAAAATCTG CCGTTTGATC
 5881 AGCTGGCCTG GGGCCATATT ACCACCGCCG CCCAGATTAC CAGTCTGCTG CCGCTGCTGA
 5941 CCGCAATTA TGATCTGAGC AATGATGTTT TGTGTCTGGC ACAGCGCCGT GGCAGTATTC
 6001 TGCTGAATAC CATGCTGGAA GCCATTGCCG CAGATAGCAG TCCGGGCCGC TGGCTGGTGC
 6061 TGGTTGCCCA TGATACCAAT ATTGCAATGG TTCGTACCCT GATGGATTTT AATTGGCAGC
 6121 TGCCGGGTTA TAGCCGTGGT AATATTCCG CGGGTAGTAG TCTGGTGTGCT GAACGTTGGC
 6181 GTGATACCCG CAGCGCGAA CGTTTCTGC GTCTGTATTT TCAGGCACAG AGCTGGATG
 6241 GTATTGCGCA GCTGCAGCCT ATTGATGATA AACATCCGCT GCTGCGTCAG GAATGGCAGC
 6301 AGCCGGATTG TCGCGTGACC GCAGTTGGTC TGCTGTGTCC GTATCAGAGC ACCCTGACCC
 6361 AGCTGCGCAA AAATCTGGAT AATAGTGACG TGCTGCCGGT GAGCGTGATT CTGCCGGCGG
 6421 CCGCACTCGA GCACCACCAC CACCACCACT GAGATCCGGC TGCTAACAAA GCCCGAAAGG
 6481 AAGCTGAGTT GGCTGCTGCC ACCGCTGAGC AATAACTAGC ATAACCCCTT GGGGCTCTA

6541 AACGGGTCTT GAGGGGTTTT TTGCTGAAAAG GAGGAACTAT ATCCGGAT

C.

Optimized Protein_M2

MTLPTLCRCALILGSLWLLSPATQAADYQLEKVVELSRHGVRPPTPGNRKEMEASQQP
WTQWTTADGELTGHGYSVVNKGRWEGEHYRQLGLLGTGCPDAAQVYVRASPLQRTR
ATAAALTDGAFPGCGVTVHHVAGDVDPLFQGEKLTVTRTDPAQELACKQQKAGDLAHL
QQQLQPAIQQKAAVCPATDCPLFEAPWTFRQTRNGNSYVYGLSVMASMVETLRLGYS
ENLPFDQLAWGHITTAQAITSLLPLLTANYDLSNDVLCLAQRRGSILLNTMLEAIAADSSP
GRWLVLVAHDTNIAMVRTLMDFNWQLPGYSRGNIPPGSSLVLERWRDTRSGERFLRLYF
QAQSLDGIRQLQPIDDKHPLLRQEWHPDCRVTAVGLLCPYQSTLTQLRKNLDNSAVLP
VSVILP

Figure 6. 5 A. The M2 recombinant construct: a. codon optimized M2 phytase gene cloned in pET29b vector. B. The complete nucleotide sequence of the M2 recombinant construct (pET29b cloned with the optimized M2 phytase gene (red highlighted)). C. the optimized M2 phytase protein sequence.

Subsequently, the recombinant constructs were separately introduced into the *E.coli* BL21(DE3) host system, resulting in the creation of three distinct systems: WT-pET29b (WT), M1-pET29-b (M1), and M2-pET29b (M2). Each of these *E.coli* BL21(DE3) systems was induced using IPTG (1mM) at 30°C for an overnight period, facilitating the over-expression and production of WT and M2 in substantial amounts. However, M1 was found to be unstable and failed to achieve over-expression in the host system. Therefore, only WT and M2 phytases were purified using IMAC via His60NiSF resin. The molecular weight observed for the over-expressed WT and mutant M2 phytases was approximately 43kDa (as indicated in (**Figure 6. 6**)). Western blot analysis further corroborated this data (**Figure 6. 6**).

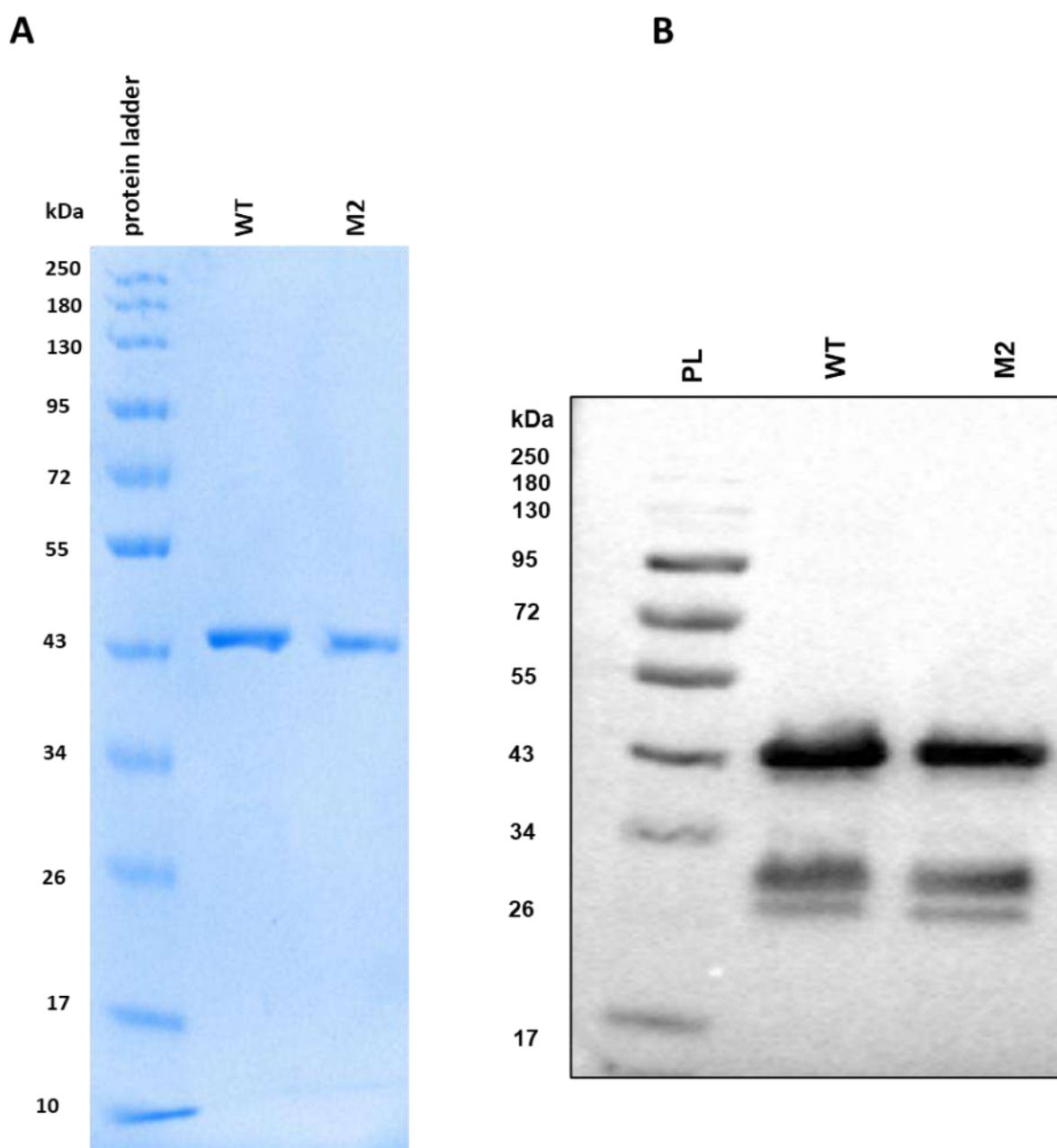


Figure 6. 6 A. Purified WT and M2 resolved on 10% SDS-PAGE gel. B. Western blot analysis of purified WT and M2 system.

6.3.3.2 Biochemical characterization of wildtype, and mutant phytases:

KH_2PO_4 was used as a standard to analyze the release of inorganic phosphate (**Figure 6. 7**).

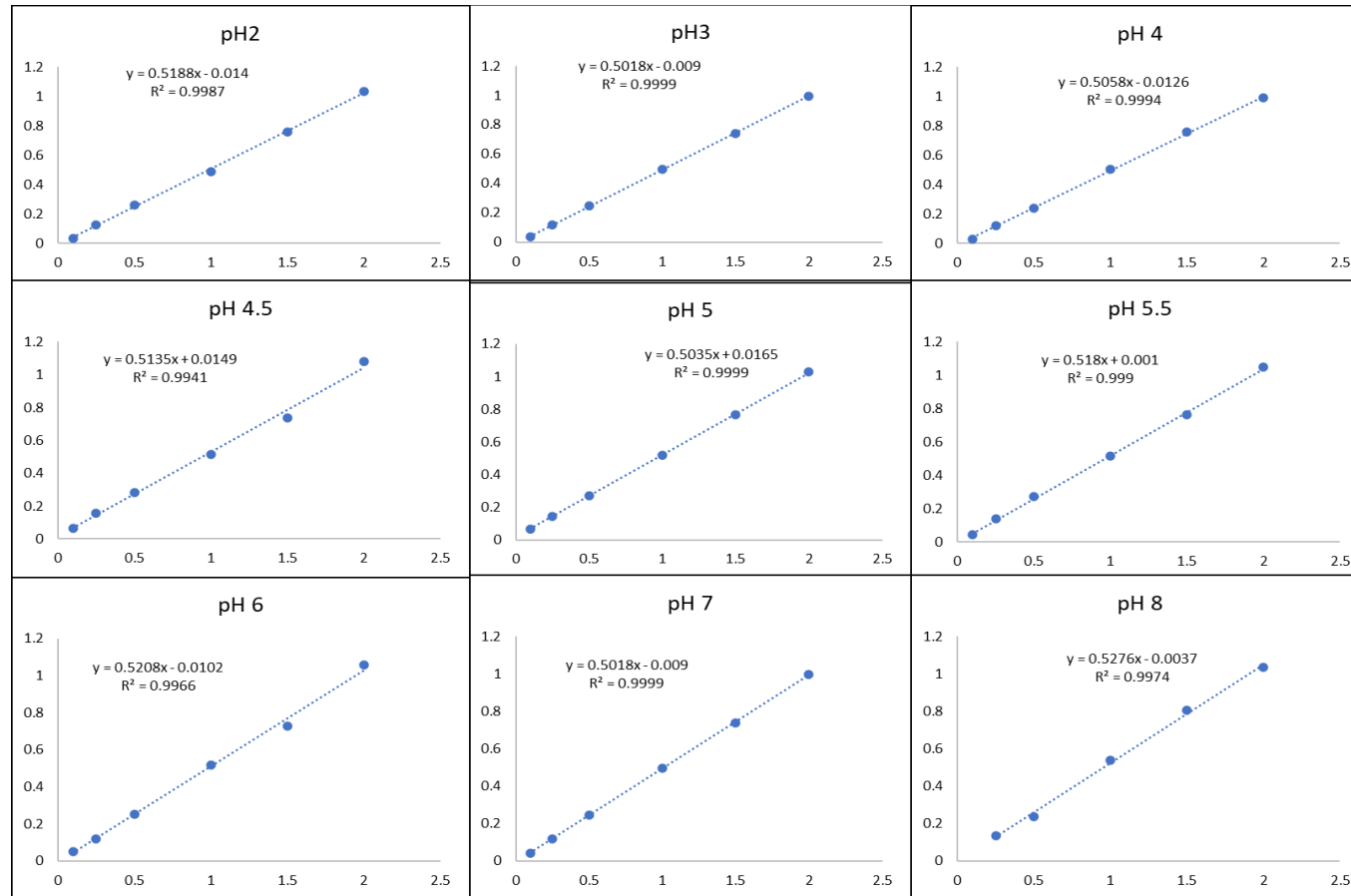


Figure 6. 7 KH₂PO₄ was utilized as a benchmark to measure the liberation of inorganic phosphate across varying pH levels (pH 2, 3, 4, 4.5, 5, 5.5, 6, 7, and 8).

Effect of pH on phytase activity:

WT and M2 phytases underwent exposure to varied pH conditions (ranging from pH 2 to pH 8), and their respective phytase activities were measured under standard conditions (37°C for 30 minutes). The results revealed that both WT and M2 displayed an optimal pH of 4 (as indicated in **Figure 6. 8**). Notably, the optimal pH for both the wild type and mutant was observed within the acidic range.

Effect of temperature on phytase activity:

Phytase activity assays were carried out for WT and M2 phytases across varying temperatures (ranging from 30 to 70°C). The results unveiled that the optimum temperature for WT was 45°C, whereas for M2, it was slightly higher at 50°C. Consequently, a noticeable 5°C difference in temperature optima was observed between the two variants (illustrated in **Figure 6. 9**).

Effect of substrate concentration on phytase activity:

K_m of WT and M2 were found to be 0.24 and 0.33 mM respectively (**Figure 6. 10**).

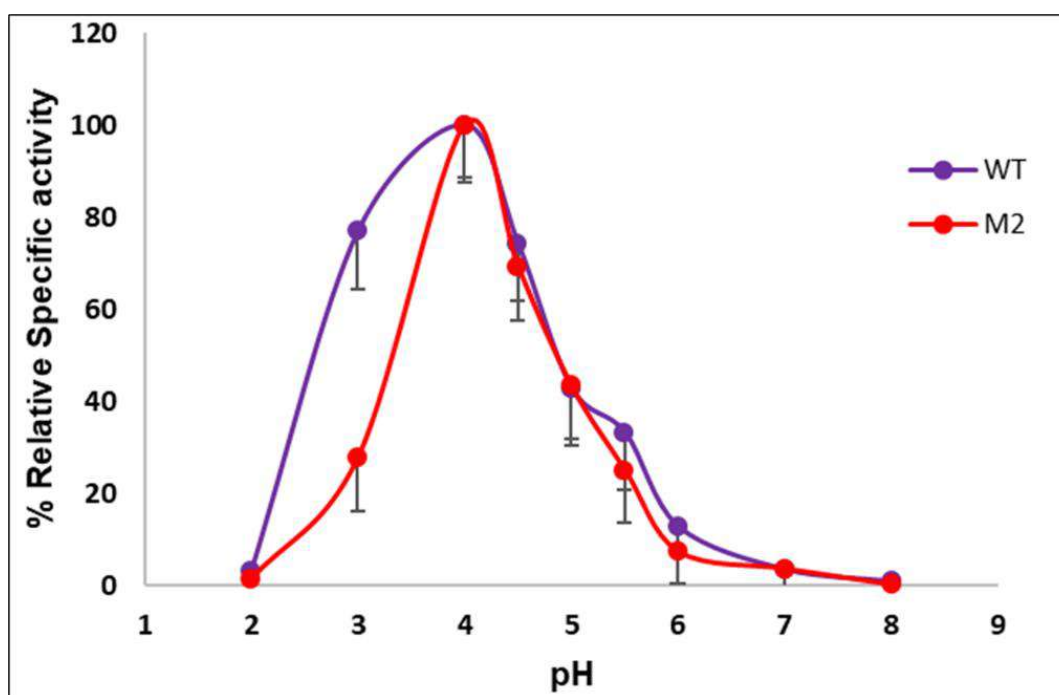


Figure 6. 8 Effect of different pH conditions on purified PVP.

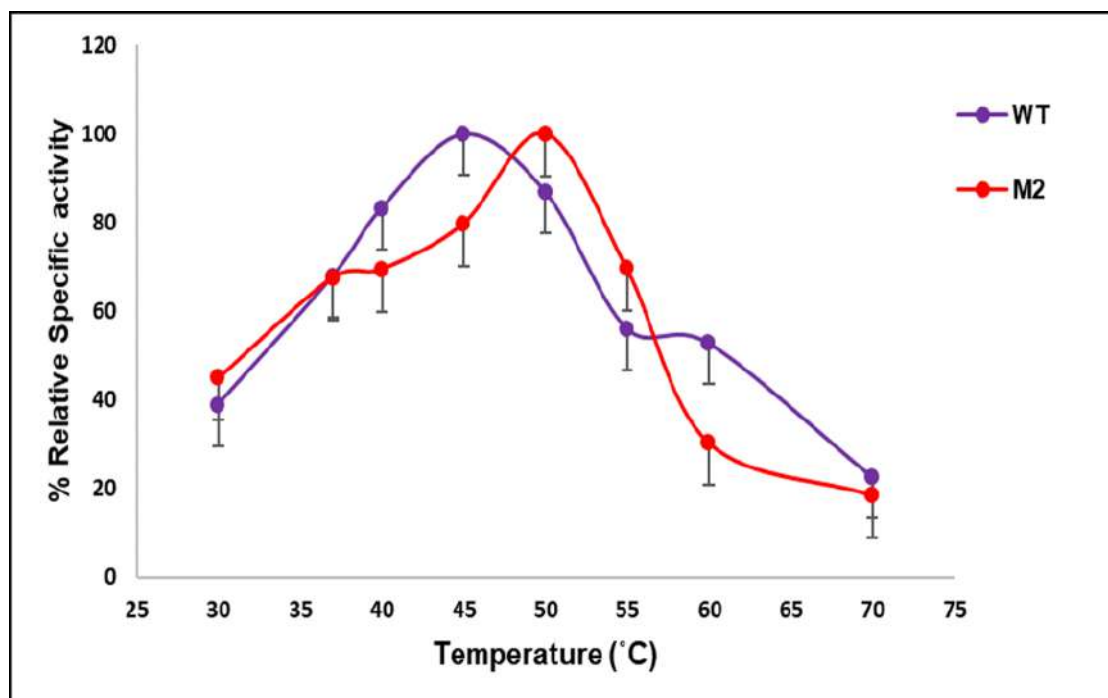


Figure 6. 9 Effect of different temperature conditions on purified PVP activity.

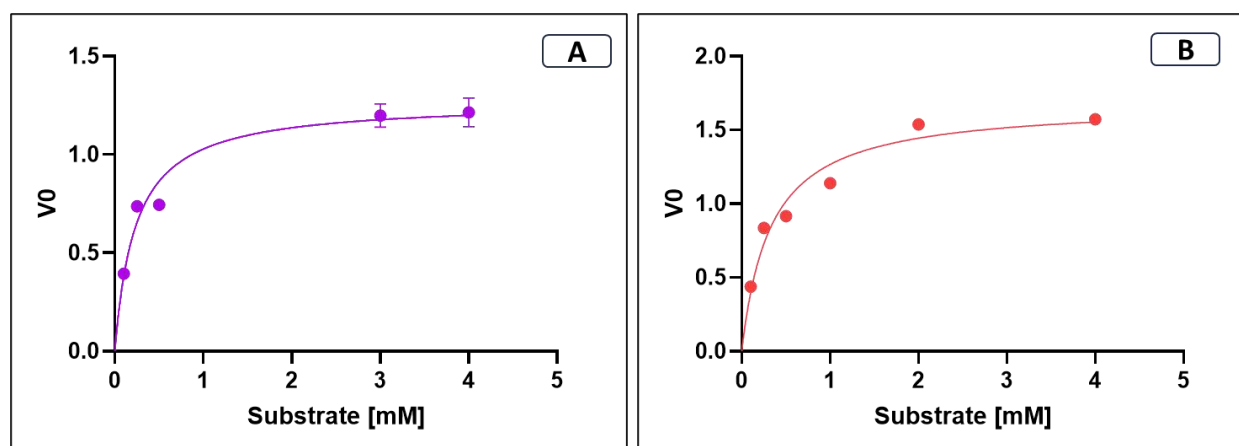


Figure 6. 10 Effect of substrate concentration on purified PVP activity: A.wildtype PVP, B. M2.

6.3.3.3 Biophysical characterization of wildtype, and mutant phytases:

Circular dichroism (CD) experiment:

The CD method was used to study protein folding of purified WT and M2. It was found that all the WT and M2 are properly folded as shown in (Figure 6. 11).

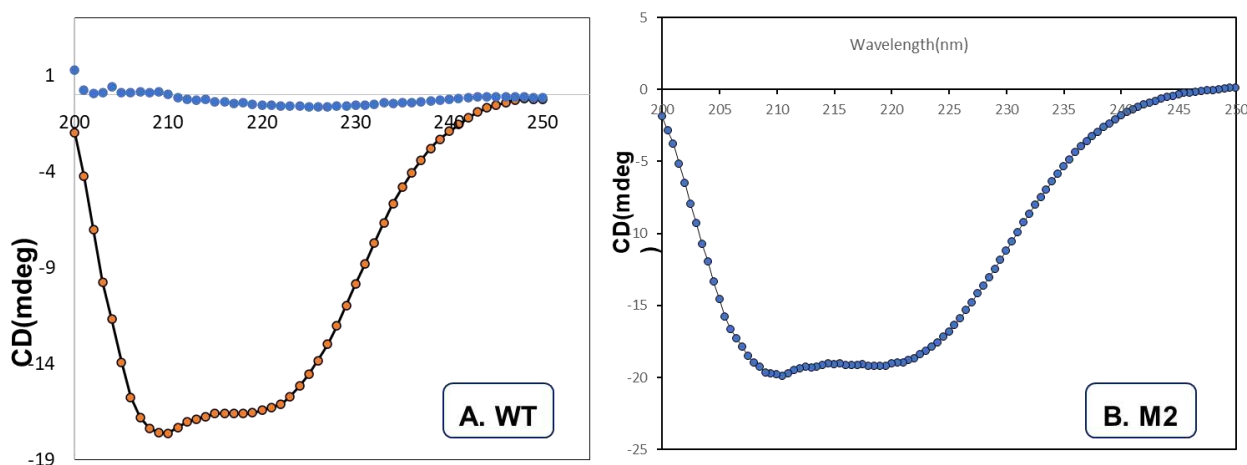


Figure 6. 11 Circular Dichroism analysis of A. WT, B.M2.

Isothermal titration calorimetry (ITC) experiment:

The ligand binding affinity of WT and M2 was measured by isothermal titration calorimetry. The binding affinity (KD) value for various phytases is in micro micromolar range as shown in **Table 6. 10**. The KD value for WT is 52 μ M whereas for M2 is also in a similar range which is 56 μ M.

Table 6. 10 Binding affinity of WT and M2 by using Isothermal titration calorimetry (ITC) experiment.

Substrate	Binding affinity (K_d) (μ M)
WT	52
M2	56

Differential scanning fluorimetry (DSF) experiment:

DSF data revealed that T_m of WT and M2 was found to be 58 and 60°C respectively (**Figure 6. 12**). This change in T_m indicated that M2 was more stable than WT.

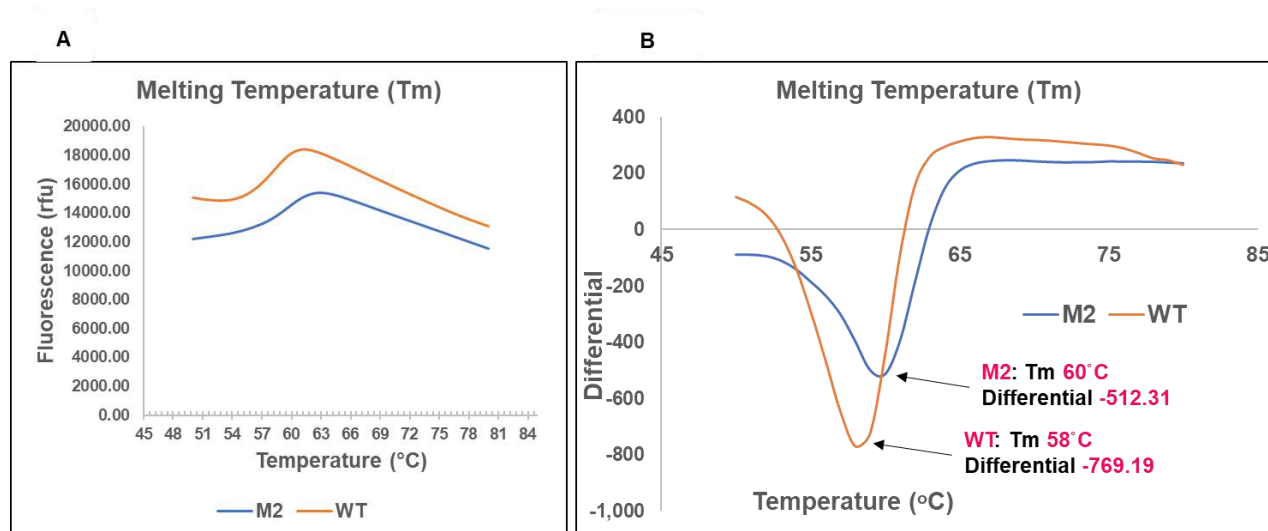


Figure 6. 12 Differential scanning fluorimetry analysis of WT and M2.

6.4 Discussion

Protein engineering methods, like rational or semi-rational approaches and directed evolution, offer ways to enhance enzyme characteristics such as thermostability, resistance to degradation, or achieving stability across various pH ranges. E.g., Shivange et al. focus on the reduction of flexibility to enhance thermostability via the adopted KeySIDE approach along with directed evolution, to improve the thermostability of *Yersinia mollaretii* phytase (Ymphytase) which involved a combination of investigating iterative key residues of the wild type and identifying substitutions using directed evolution method. M6 mutant (T77K, Q154H, G187S, and K289Q) improved the residual activity as compared to the wild type. The mutations T77K, G187S, and K289E/Q reduced the flexibility of the loops near helices which overall increased thermostability (Shivange et al., 2016). Directed evolution involves screening many colonies to find enzymes with specific traits, making it costly and time-consuming (L. Li et al., 2022).

Conversely, the rational engineering approach reduces expenses and saves time by utilizing *in-silico* hotspot identification, minimizing the number of colonies needed for screening. Fei et.al. focused on Protein flexibility as well as protein surface analysis and enhancement of salt bridges as a strategy for rational engineering of *Escherichia coli* AppA phytase. They analyzed protein flexibility with the help of a root mean square deviation (RMSF) graph obtained via

Chapter 6

Molecular Dynamic (MD) simulation. Surface thermal unstable residues were targeted based on their RMSF values which were above 2 Å and are above 4 Å away from the AppA functional sites. They demonstrated that the salt bridges and the α/β - domain of *E. coli* phytase are of utmost importance. They introduced mutation Q307D which enhanced thermostability by 9.15% at 80°C for 5 min compared to the wild type (Cao et al., 2013). Han et.al. adopted a consensus sequence rational protein engineering approach to engineer *A.niger* phytase (Anp) by structural comparison with *A.fumigatus* phytase (Afp) structure with the help of Molecular dynamics simulation. In AnP, the segments A35-P42, R163-Q168, and T248-K254 exhibited the highest main chain deviation. Hence these were targets for engineering. By hydrogen bond analysis of mutant candidate, it was found that S39 and S42 in segment-1, T165, N166, and R167 in segment-2, R248, D251, A252, and Q254 in segment-3 have the greatest contribution to the enhanced thermostability. The mutant candidate retained more activity compared to the wild type (Han et al., 2018). Similarly, *In-silico* hotspot identification, using various tools, servers, or software, has been applied to enhance the properties of enzymes like lipase (G. Li et al., 2018b), phospholipase D (L. Li et al., 2022), and phytases (Tan et al., 2016) (Wang et al., 2018).

In the present study, *In-silico* identification of hotspots focused on understanding and improving the stability of PVP. Hotspots or thermos-labile regions/ residues were identified by using computational tools/servers/ software such as Disulfide by Design2, Consensus Finder, Consurf, Hotspot Wizard, Fireprot, CABSFLEX, and based on literature. We also employed a combination of these methods to generate a comprehensive list of potential hotspots. To further refine the selection, the hotspots identified were evaluated based on the ΔG or $\Delta\Delta G$ (change in free energy) values calculated using FoldX software. The ΔG values of the mutant variants were expected to be lower than those of the wild-type (WT) enzyme, indicating improved stability.

A thorough analysis yielded a list of potential mutants. Two mutants i.e., M1 (E85W/T97C/L178Y/N212C/R350N) and M2 (A163C, Y271C) which were obtained from a Literature-based strategy were shortlisted.

According to the various individual computational tools, such as Disulfide by Design2 (D2D), consensus finder, Consurf, and FireProt, were employed to enhance the performance of a novel acidobacteria-derived phytase (Tan et al., 2016), nitrilase (Xiong et al., 2022), xylanase (Dotsenko et al., 2022), and Amylosucrases (Tian et al., 2022), respectively. However, this

study stands out due to its distinctive approach of utilizing a diverse range of computational tools, servers, and software to generate an extensive list of potential mutant enzymes, resulting in a comprehensive analysis.

After identifying potential hotspots and selecting desired mutations or combinations of mutations, experimental validation is usually carried out to confirm the predicted effects and evaluate the performance of the engineered enzymes such as lipase (G. Li et al., 2018b), amylase (Liu et al., 2014), phytases (Tan et al., 2016) and many other. In the present study, experimental validation consisted of over-expression of phytases (WT, M1, and M2) followed by biochemical and biophysical characterization. WT and M2 were found to over-express in *E.coli* BL21(D3) host system, whereas M1 was unstable. Hence, only M2 and WT were biochemically and biophysically characterized. According to the biochemical characterization, the optimum pH of M2 and WT was found to be 4 which was similar to its closest homolog *Klebsiella* (Sajidan et al., 2004) (Böhm et al., 2010). The optimum temperatures of WT and M2 were 45°C and 50°C respectively. Hence, there was a 5°C temperature shift which signifies that M2 was stable compared to that of WT and *Klebsiella* phytase (Sajidan et al., 2004) (Böhm et al., 2010). The thermostability assay revealed that the WT and M2 were stable at 50°C, however, the % residual activity decreased at 60 and 70°C. The same pattern of thermostability was observed in *Klebsiella* phytase (Sajidan et al., 2004). The effect of substrate concentration analysis revealed that Km of WT and M2 were 0.24 and 0.33mM respectively which were comparable to *Klebsiella* phytase (Greiner et al., 1997).

ITC is one of the most used methods for protein-ligand interaction thermodynamics study. It has been well well-established method for protein-protein interaction, protein-substrate binding study, protein-DNA interaction study, and for protein antibody interaction study. This method uses the binding equilibrium directly by measuring the heat upon association of substrate with its binding partner. This method has been successfully used in membrane protein kinetics. Here we also measured the binding affinity value for various recombinant expressed and purified phytase from different sources. All these phytases showed affinity in the micro molar range. Along with ITC we also successfully showed that this phytase from different sources are properly folded after purification which were used for various biophysical characterization. Along with ITC and CD we also measured the Tm values of WT and M2. By using the DSF method we could also show how mutation helps to further stabilize PVP as shown in DSF.

6.5 Summary and Conclusion:

In this investigation, the primary focus was on employing computational tools and methodologies to identify hotspots within the novel PVP, aiming to comprehend and enhance its stability. Utilizing tools like Disulfide by Design2, Consensus Finder, ConSurf, Hotspot Wizard, Fireprot, and CABSFLEX, and referencing existing literature, we generated a comprehensive list of potential hotspots in the enzyme. These identified regions were further assessed based on their calculated ΔG or $\Delta\Delta G$ values using FoldX software, where lower values compared to the wild-type enzyme indicated potential stability improvements. The meticulous analysis led to the identification of mutants, including M1 (E85W/T97C/L178Y/N212C/R350N) and M2 (A163C, Y271C), obtained through a literature-based strategy. This study's strength lies in its unique approach, amalgamating a diverse array of computational tools and software, culminating in an extensive list of potential mutant enzymes and a comprehensive analysis.

After hotspot identification and mutation selection, experimental validation typically verifies predicted effects and evaluates engineered enzyme performance. Here, the experimental validation involved over-expressing phytases (WT, M1, and M2) and conducting biochemical, biophysical characterization, and Western blot analysis. It was observed that M1 exhibited instability, leading to the focus on the characterization of only M2 and WT, which displayed successful over-expression in the *E. coli* BL21(D3) host system. In summary, this study explores varied identification approaches and sheds light on the stability and potential enhancements of PVP through both computational and experimental methodologies.

Chapter 7

**Summary and
Conclusion**

Summary and Conclusion

In this present study, three different approaches were used to identify potential phytase producers: conventional, metagenomics, and *in-silico* bioprospecting. In the case of the conventional approach by implementing preliminary screening methods, the study effectively reduced the costs and resources required for large-scale experimentation. Overall, this study offers an alternative, smaller-scale approach to identifying microbes present in soil that produce effective phytase. The metagenomic approach involved analysis of pig stag soil DNA using 16S rDNA sequencing which identified the *Xanthomonadaceae* family as a potential target for further investigation of phytase activity. *In-silico* bioprospecting methods were used to identify potential phytase producers from different databases. In this study, a total number of 17 different genera were identified, including *Pantoea vagans*(PV) and *Edwardsiella tarda* (ET). The phytases from PV and ET were further studied computationally and experimentally. Hence, the combination of these three approaches enabled us to identify a wide range of potential phytase producers.

Both *in-silico* characterization and experimental validation indicated that PVP exhibited greater stability than ETP, making it the preferred candidate for further optimization using rational engineering techniques.

In this study, we used various computational techniques to identify potential mutations in PVP that could improve its stability. By applying specific criteria and filters, we generated a list of shortlisted mutants. This study is notable for its unique methodology, which involved employing a wide array of computational tools, servers, and software to produce a comprehensive analysis and generate an extensive list of potential mutant enzymes.

Based on the experimental validation results, it was determined that M2 was more stable than WT in terms of T_m values.

Chapter 8

Significance of the study and Future prospects

Significance of the study

The significance of this study lies in its exploration of different approaches to identify potential phytase producers. By employing conventional, metagenomic, and *in-silico* bioprospecting methods, we were able to identify a wide range of potential phytase producers. The conventional approach involved preliminary screening methods that reduced costs and resources needed for large-scale experimentation. This approach allowed for the identification of microbes in soil that produce effective phytase, providing an alternative, smaller-scale method. In addition to the above points, this study stands out due to its distinctive approach, utilizing a diverse range of computational tools, servers, and software to conduct a thorough analysis and create a vast list of potential mutant enzymes.

Overall, this study contributes to the understanding of potential phytase producers, explores different approaches for their identification, and provides insights into the stability and potential improvements of these enzymes through computational and experimental techniques.

Future Prospects

This research identified promising phytase-producing candidates through *in-silico* bioprospecting. Subsequently, the isolated phytases require thorough characterization and experimental verification to assess their functionality and properties. Protein engineering methods can then be employed to optimize these enzymes, with potential mutations shortlisted for further investigation. A combination of computational characterization, molecular dynamics simulations, and wet lab validation can ultimately guide the selection of the most effective mutant.

Bibliography

Bibliography:

Abeldenov, S., Kirillov, S., Nurmagambetova, A., Kiribayeva, A., Silayev, D., & Khassenov, B. (2017). Expression, purification and biochemical characterization of recombinant phosphohydrolase appa in *Escherichia coli*. *Eurasian Journal of Applied Biotechnology*, 1–10.

Adam, N., & Perner, M. (2018). Novel hydrogenases from deep-sea hydrothermal vent metagenomes identified by a recently developed activity-based screen. *ISME Journal*, 12(5), 1225–1236. <https://doi.org/10.1038/s41396-017-0040-6>

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z>

Amol V. Shivange, Hans Wolfgang Hoeffken, Stefan Haefner, and U. S. (2016). Protein consensus based surface engineering a computer assisted method for directed protein evolution. *BioTechniques*, 61, 305–314. <https://doi.org/10.2144/000114483>

Amritha, G. K., & Venkateswaran, G. (2017). Use of Lactobacilli in Cereal-Legume Fermentation and as Potential Probiotics towards Phytate Hydrolysis. *Probiotics & Antimicro. Prot.* <https://doi.org/10.1007/s12602-017-9328-0>

Andreeva, A., Kulesha, E., Gough, J., & Murzin, A. G. (2020). The SCOP database in 2020: Expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Research*, 48(D1), D376–D382. <https://doi.org/10.1093/nar/gkz1064>

Ariaeenejad, S., Nooshi-Nedamani, S., Rahban, M., Kavousi, K., Pirbalooti, A. G., Mirghaderi, S. S., Mohammadi, M., Mirzaei, M., & Salekdeh, G. H. (2018). A Novel High Glucose-Tolerant β -Glucosidase: Targeted Computational Approach for Metagenomic Screening. *Frontiers in Bioengineering and Biotechnology*, 10. <https://doi.org/10.3389/fbioe.2020.00813>

Ariaeenejad, S., Sheykh Abdollahzadeh Mamaghani, A., Maleki, M., Kavousi, K., Foroozandeh Shahraki, M., & Hosseini Salekdeh, G. (2020). A novel high performance in-silico screened metagenome-derived alkali-thermostable endo- β -1,4-glucanase for lignocellulosic biomass hydrolysis in the harsh conditions. *BMC Biotechnology*, 20(1), 1–13. <https://doi.org/10.1186/s12896-020-00647-6>

Bibliography

Ariza, A., Moroz, O. V., Blagova, E. V., Turkenburg, J. P., Waterman, J., Roberts, S. M., Vind, J., Sjøholm, C., Lassen, S. F., de Maria, L., Glitsoe, V., Skov, L. K., & Wilson, K. S. (2013). Degradation of Phytate by the 6-Phytase from *Hafnia alvei*: A Combined Structural and Solution Study. *PLoS ONE*, 8(5), e65062. <https://doi.org/10.1371/journal.pone.0065062>

Aseri, G. K., Jain, N., & Tarafdar, J. C. (2009). Hydrolysis of Organic Phosphate Forms by Phosphatases and Phytase Producing Fungi of Arid and Semi Arid Soils of India. *American-Eurasian Journal of Agricultural and Environmental Science*, 5(4), 564–570.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, 44(W1), W344–W350. <https://doi.org/10.1093/nar/gkw408>

Ashraf, N. M., Krishnagopal, A., Hussain, A., Kastner, D., Sayed, A. M. M., Mok, Y. K., Swaminathan, K., & Zeeshan, N. (2019). Engineering of serine protease for improved thermostability and catalytic activity using rational design. *International Journal of Biological Macromolecules*, 126, 229–237. <https://doi.org/10.1016/j.ijbiomac.2018.12.218>

Attwood, T. K., Beck, M. E., Bleasby, A. J., & Parry-Smith, D. J. (1994). PRINTS - A database of protein motif fingerprints. *Nucleic Acids Research*, 22(17), 3590–3596.

Baa-Puyoulet, P., Parisot, N., Febvay, G., Huerta-Cepas, J., Vellozo, A. F., Gabaldón, T., Calevro, F., Charles, H., & Colella, S. (2016). ArthropodaCyc: A CycADS powered collection of BioCyc databases to analyse and compare metabolism of arthropods. *Database*, 2016, 1–9. <https://doi.org/10.1093/database/baw081>

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), W39–W49. <https://doi.org/10.1093/nar/gkv416>

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Research*, 28(1), 304–305. <https://doi.org/10.1093/nar/28.1.304>

Balaban, N. P., Suleimanova, A. D., Shakirov, E. V., & Sharipova, M. R. (2018). Histidine Acid Phytases of Microbial Origin. *Microbiology (Russian Federation)*, 87(6), 745–756. <https://doi.org/10.1134/S0026261718060024>

Bibliography

Balwani, I., Chakravarty, K., & Gaur, S. (2017). Role of phytase producing microorganisms towards agricultural sustainability. *Biocatalysis and Agricultural Biotechnology*, *12*, 23–29. <https://doi.org/10.1016/j.bcab.2017.08.010>

Basso, A., & Serban, S. (2019). Industrial applications of immobilized enzymes—A review. *Molecular Catalysis*, *479*(March), 110607. <https://doi.org/10.1016/j.mcat.2019.110607>

Bauer, R. A., Günther, S., Jansen, D., Heeger, C., Thaben, P. F., & Preissner, R. (2009). SuperSite: Dictionary of metabolite and drug binding sites in proteins. *Nucleic Acids Research*, *37*(SUPPL. 1), 195–200. <https://doi.org/10.1093/nar/gkn618>

Bendl, J., Stourac, J., Sebestova, E., Vavra, O., Musil, M., Brezovsky, J., & Damborsky, J. (2016). HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. *Nucleic Acids Research*, *44*(1), W479–W487. <https://doi.org/10.1093/nar/gkw416>

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343–350. <https://doi.org/10.1093/bioinformatics/btq662>

Benkert, P., Tosatto, S. C. E., & Schomburg, D. (2007). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, *71*(1), 261–277. <https://doi.org/10.1002/prot.21715>

Berman, H., Henrick, K., Nakamura, H., & Markley, J. L. (2007). The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, *35*(SUPPL. 1), 2006–2008. <https://doi.org/10.1093/nar/gkl971>

Bhagwat, P., Amobonye, A., Singh, S., & Pillai, S. (2021). A comparative analysis of GH18 chitinases and their isoforms from *Beauveria bassiana*: An in-silico approach. In *Process Biochemistry* (Vol. 100, Issue October). Elsevier Ltd. <https://doi.org/10.1016/j.procbio.2020.10.012>

Bharambe, S. P., & Peshwe, S. A. (2023). Characterizing Phytate Hydrolyzing Fungal isolate using ITS Region Profiling. *Biological Forum – An International Journal*, *15*(5a), 6–11.

Bhavsar, K., & Khire, J. M. (2014). Current research and future perspectives of phytase bioprocessing. *RSC Advances*, *4*(51), 26677. <https://doi.org/10.1039/c4ra03445g>

Bibliography

Bhavsar, K. P. (2012). *Studies on phytase from Aspergillus niger NCIM 563 under solid state fermentation and its correlation with submerged phytase I and II.*

Bienert, S., Waterhouse, A., De Beer, T. A. P., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Research*, 45(D1), D313–D319. <https://doi.org/10.1093/nar/gkw1132>

Bleasby, A. J., Akrigg, D., & Attwood, T. K. (1994). OWL - A non-redundant composite protein sequence database. *Nucleic Acids Research*, 22(17), 3574–3577.

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., ... Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 49(D1), D344–D354. <https://doi.org/10.1093/nar/gkaa977>

Boddu, R. S., Ajay Prabhakar, K., & Divakar, K. (2022). Metagenomic Bioprospecting of Uncultivable Microbial Flora in Soil Microbiome for Novel Enzymes. *Geomicrobiology Journal*, 39(2), 97–106. <https://doi.org/10.1080/01490451.2021.2017079>

Böhm, K., Herter, T., Müller, J. J., Borriss, R., & Heinemann, U. (2010a). Crystal structure of *Klebsiella* sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme. *FEBS Journal*, 277(5), 1284–1296. <https://doi.org/10.1111/j.1742-4658.2010.07559.x>

Böhm, K., Herter, T., Müller, J. J., Borriss, R., & Heinemann, U. (2010b). Crystal structure of *Klebsiella* sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme. *FEBS Journal*, 277(5), 1284–1296. <https://doi.org/10.1111/j.1742-4658.2010.07559.x>

Boteva, N., & Kambourova, M. (2018). *Thermophiles and Their Exploration for Thermostable Enzyme Production.* 167–186. https://doi.org/10.1007/978-981-13-0329-6_6

Brinch-Pedersen, H., Madsen, C. K., Holme, I. B., & Dionisio, G. (2014). Increased understanding of the cereal phytase complement for better mineral bio-availability and resource management. *Journal of Cereal Science*, 59(3), 373–381. <https://doi.org/10.1016/j.jcs.2013.10.003>

Bibliography

- Bunterngsook, B., Mhuantong, W., Kanokratana, P., Iseki, Y., Watanabe, T., & Champreda, V. (2021). Identification and characterization of a novel AA9-type lytic polysaccharide monoxygenase from a bagasse metagenome. *Applied Microbiology and Biotechnology*, *105*(1), 197–210. <https://doi.org/10.1007/s00253-020-11002-2>
- Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Christie, C. H., Dalenberg, K., Di Costanzo, L., Duarte, J. M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D. S., Ghosh, S., Green, R. K., Guranović, V., Guzenko, D., Hudson, B. P., ... Zhuravleva, M. (2021). RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*, *49*(D1), D437–D451. <https://doi.org/10.1093/nar/gkaa1038>
- Caenepeel, S., Charydczak, G., Sudarsanam, S., Hunter, T., & Manning, G. (2004). The mouse kinome: Discovery and comparative genomics of all mouse protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(32), 11707–11712. <https://doi.org/10.1073/pnas.0306880101>
- Cang, L., Wang, Y., Zhou, D., & Dong, Y. (2004). Heavy metals pollution in poultry and livestock feeds and manures under intensive farming in Jiangsu Province, China. *Journal of Environmental Sciences (China)*, *16*(3), 371–374.
- Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research*, *37*(SUPPL. 1), 233–238. <https://doi.org/10.1093/nar/gkn663>
- Carbajosa, G., Trigo, A., Valencia, A., & Cases, I. (2009). Bionemo: Molecular information on biodegradation metabolism. *Nucleic Acids Research*, *37*(SUPPL. 1), 598–602. <https://doi.org/10.1093/nar/gkn864>
- Carlos Leonardo Araújo, Iago Blanco, Luciana Souza, Sandeep Tiwari, Lino César Pereira, Preetam Ghosh, Vasco Azevedo, Artur Silva, A. F. (2020). In silico functional prediction of hypothetical proteins from the core genome of *Corynebacterium pseudotuberculosis* biovar ovis. *PeerJ*, *8*, 1–24. <https://doi.org/10.7717/peerj.9643>
-

Bibliography

Castillo Villamizar, G. A., Nacke, H., Boehning, M., Herz, K., & Daniel, R. (2019). Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. *MBio*, *10*(1), 1–15. <https://doi.org/10.1128/mBio.01966-18>

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., & Ben-Tal, N. (2013a). ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*, *53*(3–4), 199–206. <https://doi.org/10.1002/ijch.201200096>

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., & Ben-Tal, N. (2013b). ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*, *53*(3–4), 199–206. <https://doi.org/10.1002/ijch.201200096>

Chanderman, A., Puri, A. K., Permaul, K., & Singh, S. (2016). Production, characteristics and applications of phytase from a rhizosphere isolated *Enterobacter* sp. ACSS. *Bioprocess and Biosystems Engineering*, *39*(10), 1577–1587. <https://doi.org/10.1007/s00449-016-1632-7>

Chaudhary, N., Sharma, A. K., Agarwal, P., Gupta, A., & Sharma, V. K. (2015). 16S classifier: A tool for fast and accurate taxonomic classification of 16S rRNA hypervariable regions in metagenomic datasets. *PLoS ONE*, *10*(2), 1–13. <https://doi.org/10.1371/journal.pone.0116106>

Chen, C. C., & Tw, T. (2018). *PHYTASE HAVING IMPROVED THERMOSTABILITY*.

Chen, L., Jiang, K., Zhou, Y., Zhu, L., & Chen, X. (2022). Improving the Thermostability of α -Glucosidase from *Xanthomonas campestris* through Proline Substitutions Guided by Semi-rational Design. *Biotechnology and Bioprocess Engineering*, *27*(4), 631–639. <https://doi.org/10.1007/s12257-022-0129-2>

Chen, L., Yi, Z., Fang, Y., Jin, Y., Xiao, Y., Zhao, D., Luo, H., He, H., Sun, Q., & Zhao, H. (2020). Uncovering key residues responsible for the thermostability of a thermophilic 1,3(4)- β -D-glucanase from Nong flavor Daqu by rational design. *Enzyme and Microbial Technology*, *142*(24), 1–9. <https://doi.org/10.1016/j.enzmictec.2020.109672>

Chen, W., Ye, L., Guo, F., Lv, Y., & Yu, H. (2015). Enhanced activity of an alkaline phytase from *Bacillus subtilis* 168 in acidic and neutral environments by directed evolution. *Elsevier B.V.* <https://doi.org/10.1016/j.bej.2015.02.021>

Bibliography

Choi, J., Détry, N., Kim, K. T., Asiegbu, F. O., Valkonen, J. P., & Lee, Y. H. (2014). FPoxDB: Fungal peroxidase database for comparative genomics. *BMC Microbiology*, *14*(1), 1–8. <https://doi.org/10.1186/1471-2180-14-117>

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), 1511–1519. <https://doi.org/10.1002/pro.5560020916>

Craig, D. B., & Dombkowski, A. A. (2013). Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*, *14*(1), 0–6. <https://doi.org/10.1186/1471-2105-14-346>

Cruz, G., Acosta, J., Mancheño, J. M., Del Arco, J., & Fernández-Lucas, J. (2022). Rational Design of a Thermostable 2'-Deoxyribosyltransferase for Nelarabine Production by Prediction of Disulfide Bond Engineering Sites. *International Journal of Molecular Sciences*, *23*(19). <https://doi.org/10.3390/ijms231911806>

Dadwal, A., Singh, V., Sharma, S., Sahoo, A. K., & Satyanarayana, T. (2022). Structural and thermostability insights into cellobiohydrolase of a thermophilic mould *Myceliophthora thermophila*: in-silico studies. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 1–10. <https://doi.org/10.1080/07391102.2022.2133012>

Daniel, R. (2005). THE METAGENOMICS OF SOIL. *Microbiology*, *3*(June), 470–478. <https://doi.org/10.1038/nrmicro1160>

Dawson, N. L., Lewis, T. E., Das, S., Lees, J. G., Lee, D., Ashford, P., Orengo, C. A., & Sillitoe, I. (2017). CATH: An expanded resource to predict protein function through structure and sequence. *Nucleic Acids Research*, *45*(D1), D289–D295. <https://doi.org/10.1093/nar/gkw1098>

de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 362–365. <https://doi.org/10.1093/nar/gkl124>

Dev, S. S., Nisha, E. A., & Venu, A. (2016). Biochemical and molecular characterization of efficient phytase producing bacterial isolates from soil samples. *International Journal of Current Microbiology and Applied Sciences*, *5*(5), 218–226. <https://doi.org/10.20546/ijcmas.2016.505.024>

Bibliography

DiFonzo, N., & Bordia, P. (2003). PhyA, a Secreted Protein of *Xanthomonas oryzae* pv. *oryzae*, is required for optimum Virulence and Growth on Phytic Acid as a Sole Phosphate Source. *Journal of Allergy and Clinical Immunology*, *16*(11), 973–982. <http://dx.doi.org/10.1016/j.jaci.2012.05.050>

Dinkel, H., Chica, C., Via, A., Gould, C. M., Jensen, L. J., Gibson, T. J., & Diella, F. (2011). Phospho.ELM: A database of phosphorylation sites-update 2011. *Nucleic Acids Research*, *39*(SUPPL. 1), 261–267. <https://doi.org/10.1093/nar/gkq1104>

Dombkowski, A. A. (2003). Disulfide by DesignTM: A computational method for the rational design of disulfide bonds in proteins. *Bioinformatics*, *19*(14), 1852–1853. <https://doi.org/10.1093/bioinformatics/btg231>

Dotsenko, A. S., Denisenko, Y. A., Rozhkova, A. M., Zorov, I. N., Korotkova, O. G., & Sinitsyn, A. P. (2022). Enhancement of thermostability of GH10 xylanase E *Penicillium canescens* directed by $\Delta\Delta G$ calculations and structure analysis. *Enzyme and Microbial Technology*, *152*(October 2021), 109938. <https://doi.org/10.1016/j.enzmictec.2021.109938>

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, *277*, 396–404. [https://doi.org/10.1016/S0076-6879\(97\)77022-8](https://doi.org/10.1016/S0076-6879(97)77022-8)

Ekstrom, A., Taujale, R., McGinn, N., & Yin, Y. (2014). PlantCAZyme: a database for plant carbohydrate-active enzymes. *Database : The Journal of Biological Databases and Curation*, *2014*, 1–8. <https://doi.org/10.1093/database/bau079>

Elbehery, A. H. A., Leak, D. J., & Siam, R. (2017). Novel thermostable antibiotic resistance enzymes from the Atlantis II Deep Red Sea brine pool. *Microbial Biotechnology*, *10*(1), 189–202. <https://doi.org/10.1111/1751-7915.12468>

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. <https://doi.org/10.1093/nar/gky995>

Ellis, L. B., & Wackett, L. P. (2012). Use of the University of Minnesota Biocatalysis/Biodegradation Database for study of microbial degradation. *Microbial Informatics and Experimentation*, *2*(1), 1–10. <https://doi.org/10.1186/2042-5783-2-1>

Bibliography

Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, 300(4), 1005–1016. <https://doi.org/10.1006/jmbi.2000.3903>

Eric Wommack, K., Bhavsar, J., Polson, S. W., Chen, J., Dumas, M., Srinivasiah, S., Furman, M., Jamindar, S., & Nasko, D. J. (2012). VIROME: A standard operating procedure for analysis of viral metagenome sequences. *Standards in Genomic Sciences*, 6(3), 427–439. <https://doi.org/10.4056/sigs.2945050>

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics*, 1–47. <https://doi.org/10.1002/0471250953.bi0506s15.Comparative>

Eyal Akiva, Shoshana Brown, Daniel E. Almonacid, A. E. B., & Ashley F. Custer, Michael A. Hicks, Conrad C. Huang, Florian Lauck, Susan T. Mashiyama, Elaine C. Meng, David Mischel, John H. Morris, Sunil Ojha, Alexandra M. Schnoes, Doug Stryke, Jeffrey M. Yunes, Thomas E. Ferrin, Gemma L. Holliday, P. C. B. (2014). The Structure-Function Linkage Database. *Nucleic Acids Research*, 42(D1), 521–530. <https://doi.org/10.1093/nar/gkt1130>

Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L., & Hermjakob, H. (2017). Reactome pathway analysis: A high-performance in-memory approach. *BMC Bioinformatics*, 18(1), 1–9. <https://doi.org/10.1186/s12859-017-1559-2>

Faiza, Dr. M. (2020). *How to perform docking in a speci c binding site using AutoDock Vina ?*

Fakhravar, A., & Hesampour, A. (2018). Rational design-based engineering of a thermostable phytase by site-directed mutagenesis. *Molecular Biology Reports*, 45(6), 2053–2061. <https://doi.org/10.1007/s11033-018-4362-x>

Faria Fatima, Chaudhary Ira, Ali Jasarat, Rastogi Smita, P. N. (2011). Microbial DNA extraction from soil by different methods and its PCR amplification. *Biochem.Cell.Arch*, 11.

Farias, N., Almeida, I., & Meneses, C. (2018). New bacterial phytase through metagenomic prospection. *Molecules*, 23(2), 1–14. <https://doi.org/10.3390/molecules23020448>

Fei, B., Xu, H., Cao, Y., Ma, S., Guo, H., Song, T., Qiao, D., & Cao, Y. (2013). A multi-factors rational design strategy for enhancing the thermostability of Escherichia coli AppA phytase.

Bibliography

Journal of Industrial Microbiology and Biotechnology, 40(5), 457–464.
<https://doi.org/10.1007/s10295-013-1260-z>

Fei, B., Xu, H., Zhang, F., Li, X., Ma, S., Cao, Y., Xie, J., Qiao, D., & Cao, Y. (2013). Relationship between *Escherichia coli* AppA phytase's thermostability and salt bridges. *Journal of Bioscience and Bioengineering*, 115(6), 623–627.
<https://doi.org/10.1016/j.jbiosc.2012.12.010>

Fischer, M., & Pleiss, J. (2003). The Lipase Engineering Database: A navigation and analysis tool for protein families. *Nucleic Acids Research*, 31(1), 319–321.
<https://doi.org/10.1093/nar/gkg015>

Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K. B., Bairoch, A., Schomburg, D., Tipton, K. F., & Apweiler, R. (2004). IntEnz, the integrated relational enzyme database. *Nucleic Acids Research*, 32(DATABASE ISS.), 434–437.
<https://doi.org/10.1093/nar/gkh119>

Foong, C. P., Lakshmanan, M., Abe, H., Taylor, T. D., Foong, S. Y., & Sudesh, K. (2018). A novel and wide substrate specific polyhydroxyalkanoate (PHA) synthase from unculturable bacteria found in mangrove soil. *Journal of Polymer Research*, 25(1), 23.
<https://doi.org/10.1007/s10965-017-1403-4>

Forster, S. C., Browne, H. P., Kumar, N., Hunt, M., Denise, H., Mitchell, A., Finn, R. D., & Lawley, T. D. (2016). HPMCD: The database of human microbial communities from metagenomic datasets and microbial reference genomes. *Nucleic Acids Research*, 44(D1), D604–D609. <https://doi.org/10.1093/nar/gkv1216>

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607. <https://doi.org/10.1385/1592598900>

Gauchan, D. P., Pandey, S., Pokhrel, B., Bogati, N., Thapa, P., Acharya, A., KC, B. M., & Lamichhane, J. (2023). Growth Promoting Role of Phytase Producing Bacteria Isolated from *Bambusa tulda* Roxb. Rhizosphere in Maize Seedlings Under Pot Conditions. *Journal of Nepal Biotechnology Association*, 4(1), 17–26. <https://doi.org/10.3126/jnba.v4i1.53442>

Bibliography

Geourjon, C., & Deléage, G. (1995). Sopma: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, *11*(6), 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681>

Gharechahi, J., Vahidi, M. F., Sharifi, G., Ariaeenejad, S., Ding, X. Z., Han, J. L., & Salekdeh, G. H. (2023). Lignocellulose degradation by rumen bacterial communities: New insights from metagenome analyses. *Environmental Research*, *229*(April), 115925. <https://doi.org/10.1016/j.envres.2023.115925>

Gocheva, Y., Engibarov, S., Lazarkevich, I., & Eneva, R. (2023). Phytases - Types, Sources, and Factors Affecting Their Activity. *Acta Microbiologica Bulgarica*, *39*(3), 249–263. <https://doi.org/10.59393/amb23390305>

Góngora-Castillo, E., López-Ochoa, L. A., Apolinar-Hernández, M. M., Caamal-Pech, A. M., Contreras-de la Rosa, P. A., Quiroz-Moreno, A., Ramírez-Prado, J. H., & O'Connor-Sánchez, A. (2020). Data mining of metagenomes to find novel enzymes: a non-computationally intensive method. *3 Biotech*, *10*(2), 1–8. <https://doi.org/10.1007/s13205-019-2044-6>

Gontia-Mishra, I., Kumar Singh, V., Tripathi, N., Sasidharan, S., & Tiwari, S. (2014). Computational identification, homology modelling and docking analysis of phytase protein from *Fusarium oxysporum*. *Biologia*, *69*(10). <https://doi.org/10.2478/s11756-014-0447-8>

González-Torres, I., Perez-Rueda, E., Evangelista-Martínez, Z., Zárata-Romero, A., Moreno-Enríquez, A., & Huerta-Saquero, A. (2020). Identification of L-asparaginases from *Streptomyces* strains with competitive activity and immunogenic profiles: A bioinformatic approach. *PeerJ*, *8*, 1–23. <https://doi.org/10.7717/peerj.10276>

Gordeeva, T. L., Borshchevskaya, L. N., Kalinina, A. N., Sineoky, S. P., Kashirskaya, M. D., & Voronin, S. P. (2019). Increase in the Thermal Stability of Phytase from *Citrobacter freundii* by Site-Directed Saturation Mutagenesis. *Applied Biochemistry and Microbiology*, *55*(8), 788–796. <https://doi.org/10.1134/S0003683819080052>

Gouet, P., Courcelle, E., Stuart, D. I., & Métoz, F. (1999). ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics*, *15*(4), 305–308. <https://doi.org/10.1093/bioinformatics/15.4.305>

Goutami Banerjee, Khin Oo, XIyun Zhang, Jie Yang, Y. Z. (2017). *PHYTASES AND USES THEREOF*.

Bibliography

Greiner, R., Haller, E., Konietzny, U., & Jany, K. (1997). Purification and Characterization of a Phytase from *Klebsiella terrigena*. *ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS*, *341*(2), 201–206.

Gromiha, M. M., An, J., Kono, H., Oobatake, M., Uedaira, H., Prabakaran, P., & Sarai, A. (2000). ProTherm, version 2.0: Thermodynamic database for proteins and mutants. *Nucleic Acids Research*, *28*(1), 283–285. <https://doi.org/10.1093/nar/28.1.283>

Guha, S., Das, S., & Ganguli, S. (2020). A comparative genomics pipeline for in silico characterization and functional annotation of short hypothetical proteins. *Journal of Tropical Life Science*, *10*(2), 141–148. <https://doi.org/10.11594/jtls.10.02.06>

Gujar, P. D. (2014). *Studies on acidic phytase from Aspergillus niger mutants*.

Gupta, S., Singh, Y., Kumar, H., Raj, U., Rao, A. R., & Varadwaj, P. K. (2018). Identification of Novel Abiotic Stress Proteins in *Triticum aestivum* Through Functional Annotation of Hypothetical Proteins. *Interdisciplinary Sciences: Computational Life Sciences*, *10*(1), 205–220. <https://doi.org/10.1007/s12539-016-0178-3>

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Research*, *41*(D1), 387–395. <https://doi.org/10.1093/nar/gks1234>

Han, N., Miao, H., Yu, T., Xu, B., Yang, Y., Wu, Q., Zhang, R., & Huang, Z. (2018). Enhancing thermal tolerance of *Aspergillus niger* PhyA phytase directed by structural comparison and computational simulation. *BMC Biotechnology*, *18*(1), 1–8. <https://doi.org/10.1186/s12896-018-0445-y>

Heinegård, D., & Paulsson, M. (1987). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Methods in Enzymology*, *145*(C), 336–363. [https://doi.org/10.1016/0076-6879\(87\)45020-9](https://doi.org/10.1016/0076-6879(87)45020-9)

Hemmati, S. A., & Mehrabadi, M. (2020). Structural ensemble-based computational analysis of trypsin enzyme genes discovered highly conserved peptide motifs in insects. *Archives of Phytopathology and Plant Protection*, *53*(7–8), 335–354. <https://doi.org/10.1080/03235408.2020.1744978>

Hua, Y., Lyu, C., Liu, C., Wang, H., Hu, S., Zhao, W., Mei, J., Huang, J., & Mei, L. (2020). Improving the Thermostability of Glutamate Decarboxylase from *Lactobacillus brevis* by

Bibliography

Consensus Mutagenesis. *Applied Biochemistry and Biotechnology*, 191(4), 1456–1469. <https://doi.org/10.1007/s12010-020-03283-0>

Huang, L., Ma, J., Sang, J., Wang, N., Wang, S., Wang, C., Kang, H., Liu, F., Lu, F., & Liu, Y. (2020). Enhancing the thermostability of phospholipase D from *Streptomyces halstedii* by directed evolution and elucidating the mechanism of a key amino acid residue using molecular dynamics simulation. *International Journal of Biological Macromolecules*, 164, 3065–3074. <https://doi.org/10.1016/j.ijbiomac.2020.08.160>

Igarashi, Y., Heureux, E., Doctor, K. S., Talwar, P., Gramatikova, S., Gramatikoff, K., Zhang, Y., Blinov, M., Ibragimova, S. S., Boyd, S., Ratnikov, B., Cieplak, P., Godzik, A., Smith, J. W., Osterman, A. L., & Eroshkin, A. M. (2009). PMAP: Databases for analyzing proteolytic events and pathways. *Nucleic Acids Research*, 37(SUPPL. 1), 611–618. <https://doi.org/10.1093/nar/gkn683>

Ihsein Rokia Amine-Khodja, Ryn Maougal, A. D. (2023). Bacteria Mineralizing Phytate In The Bean Rhizosphere In An Algerian Agro-Ecosystem. *International Conference on Pioneer and Innovative Studies*, 1, 122–126. <https://doi.org/10.59287/icpis.816>

Industrial Enzymes Market - Global Forecast by 2022. (2017). <https://www.marketsandmarkets.com/PressReleases/industrial-enzymes.asp>

Iqbal, F., Govender, A., Permaul, K., Singh, S., & Bisetty, K. (2015). Thermostable chitinase II from *Thermomyces lanuginosus* SSBP: Cloning, structure prediction and molecular dynamics simulations. *Journal of Theoretical Biology*, 374, 107–114. <https://doi.org/10.1016/j.jtbi.2015.03.035>

Jaiwal, P. K., Chhillar, A. K., Chaudhary, D., & Jaiwal, R. (2019). *Nutritional Quality Improvement in Plants*. <https://doi.org/10.1007/978-3-319-95354-0>

Jamroz, M., Kolinski, A., & Kmiecik, S. (2013). CABS-flex: Server for fast simulation of protein structure fluctuations. *Nucleic Acids Research*, 41(Web Server issue), 427–431. <https://doi.org/10.1093/nar/gkt332>

Jorquera, M. A., Gabler, S., Inostroza, N. G., Acuña, J. J., Campos, M. A., Menezes-Blackburn, D., & Greiner, R. (2017). Screening and Characterization of Phytases from Bacteria Isolated from Chilean Hydrothermal Environments. *Microbial Ecology*, 1–13. <https://doi.org/10.1007/s00248-017-1057-0>

Bibliography

Kalathur, R. K. R., Pinto, J. P., Hernández-Prieto, M. A., MacHado, R. S. R., Almeida, D., Chaurasia, G., & Futschik, M. E. (2014). UniHI 7: An enhanced database for retrieval and interactive analysis of human molecular interaction networks. *Nucleic Acids Research*, *42*(D1), 408–414. <https://doi.org/10.1093/nar/gkt1100>

Kalsi, H. K., Singh, R., Dhaliwal, H. S., & Kumar, V. (2016). Phytases from *Enterobacter* and *Serratia* species with desirable characteristics for food and feed applications. *3 Biotech*, *6*(1), 1–13. <https://doi.org/10.1007/s13205-016-0378-x>

Kamble, A. D., & Singh, H. (2022). Finding novel enzymes by in silico bioprospecting approach. In *Value-Addition in Food Products and Processing Through Enzyme Technology* (pp. 347–364). Academic Press. <https://doi.org/10.1016/B978-0-323-89929-1.00028-7>

Kamble, A., Sawant, S., & Singh, H. (2020). 16S Ribosomal RNA Gene - Based Metagenomics: A Review. *Biomedical Research Journal*, *7*(5), 5–11. <https://doi.org/10.4103/BMRJ.BMRJ>

Kamble, A., & Singh, H. (2020). Different Methods of Soil DNA Extraction. *Bio-Protocol*, *10*(2), 1–23. <https://doi.org/10.21769/bioprotoc.3521>

Kamble, A., Srinivasan, S., & Singh, H. (2019a). In-Silico Bioprospecting: Finding Better Enzymes. *Molecular Biotechnology*, *61*(1), 53–59. <https://doi.org/10.1007/s12033-018-0132-1>

Kamble, A., Srinivasan, S., & Singh, H. (2019b). In-Silico Bioprospecting: Finding Better Enzymes. *Molecular Biotechnology*, *61*(1), 53–59. <https://doi.org/10.1007/s12033-018-0132-1>

Kamble Asmita, & Singh Harinder. (2020). Different Methods of Soil DNA Extraction. *Bio-101*, *e3521*, 1–23. <https://doi.org/10.21769/BioProtoc.3521>

Karp, P. D., Riley, M., Paley, S. M., & Pellegrini-Toole, A. (2002). The MetaCyc database. *Nucleic Acids Research*, *30*(1), 59–61. <https://doi.org/10.1093/nar/30.1.59>

Kaur, P., Vohra, A., & Satyanarayana, T. (2021). Multifarious applications of fungal phytases. In *Encyclopedia of Mycology* (Issue May). Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-819990-9.00028-7>

Bibliography

Keegan, K. P., Glass, E. M., & Meyer, F. (2016). MG-RAST, a metagenomics service for analysis of microbial community structure and function. In *Methods in Molecular Biology* (Vol. 1399, pp. 207–233). Humana Press Inc. https://doi.org/10.1007/978-1-4939-3369-3_13

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2016). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. <https://doi.org/10.1038/nprot.2015-053>

Kim, M. S., & Lei, X. G. (2008a). Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR. *Applied Microbiology and Biotechnology*, *79*(1), 69–75. <https://doi.org/10.1007/s00253-008-1412-7>

Kim, M. S., & Lei, X. G. (2008b). Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR. *Applied Microbiology and Biotechnology*, *79*(1), 69–75. <https://doi.org/10.1007/s00253-008-1412-7>

Kimura, N., & Kamagata, Y. (2016). A Thermostable Bilirubin-Oxidizing Enzyme from Activated Sludge Isolated by a Metagenomic Approach. *Microbes and Environments*. <https://doi.org/10.1264/jsme2.ME16106>

Klimke, W., Agarwala, R., Badretdin, A., Chetvernin, S., Ciuffo, S., Fedorov, B., Kiryutin, B., O'Neill, K., Resch, W., Resenchuk, S., Schafer, S., Tolstoy, I., & Tatusova, T. (2009). The National Center for Biotechnology Information's Protein Clusters Database. *Nucleic Acids Research*, *37*, 216–223. <https://doi.org/10.1093/nar/gkn734>

Knudsen, M., & Wiuf, C. (2010). The CATH database. *Human Genomics*, *4*(3), 207–212. <https://doi.org/10.1186/1479-7364-4-3-207>

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. <https://doi.org/10.1006/jmbi.2000.4315>

Kuang, B., Xiao, R., Hu, Y., Wang, Y., Zhang, L., Wei, Z., Bai, J., Zhang, K., Acuña, J. J., Jorquera, M. A., & Pan, W. (2023). Metagenomics reveals biogeochemical processes carried out by sediment microbial communities in a shallow eutrophic freshwater lake. *Frontiers in Microbiology*, *13*(January), 1–13. <https://doi.org/10.3389/fmicb.2022.1112669>

Bibliography

Kuhn, M., von Mering, C., Campillos, M., Jensen, L. J., & Bork, P. (2008). STITCH: Interaction networks of chemicals and proteins. *Nucleic Acids Research*, 36(SUPPL. 1), 684–688. <https://doi.org/10.1093/nar/gkm795>

Kumar, K., Patel, K., Agrawal, D. C., & Khire, J. M. (2015). Insights into the unfolding pathway and identification of thermally sensitive regions of phytase from *Aspergillus niger* by molecular dynamics simulations. *Journal of Molecular Modeling*, 21(6). <https://doi.org/10.1007/s00894-015-2696-z>

Kumar, R. M., & Raja, S. S. S. (2019). *Isolation , Screening and Identification of Potential Thermo Stable Bacterial Enzyme Producers in Sangameshwar , Tural Hot Spring*. 9(4), 510–517.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/MOLBEV/MSW054>

Kumar, S., Tsai, C. J., & Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Engineering*, 13(3), 179–191. <https://doi.org/10.1093/protein/13.3.179>

Kumar, V., & Agrawal, S. (2014). An insight into protein sequences of PTP-like cysteine phytases. *Nusantara Bioscience*, 6(1), 102–106. <https://doi.org/10.13057/nusbiosci/n060116>

Kumar, V., Singh, G., Verma, A. K., & Agrawal, S. (2012). In silico characterization of histidine acid phytase sequences. *Enzyme Research*. <https://doi.org/10.1155/2012/845465>

Kumar, V., Singh, P., Jorquera, M. A., Sangwan, P., Kumar, P., Verma, A. K., & Agrawal, S. (2013). Isolation of phytase-producing bacteria from Himalayan soils and their effect on growth and phosphorus uptake of Indian mustard (*Brassica juncea*). *World Journal of Microbiology and Biotechnology*, 29(8), 1361–1369. <https://doi.org/10.1007/s11274-013-1299-z>

Kumar, V., & Sinha, A. K. (2018). General aspects of phytases. In *Enzymes in Human and Animal Nutrition: Principles and Perspectives*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-805419-2.00003-4>

Land, H., & Humble, M. S. (2018). YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations Henrik. In *Protein Engineering: Methods and Protocols* (Vol. 1685, pp. 43–67). <https://doi.org/10.1007/978-1-4939-7366-8>

Bibliography

Lang, M., Stelzer, M., & Schomburg, D. (2011). BKM-react, an integrated biochemical reaction database. *BMC Biochemistry*, *12*(1). <https://doi.org/10.1186/1471-2091-12-42>

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018a). PDBsum: Structural summaries of PDB entries. *Protein Science*, *27*(1), 129–134. <https://doi.org/10.1002/pro.3289>

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018b). PDBsum: Structural summaries of PDB entries. *Protein Science*, *27*(1), 129–134. <https://doi.org/10.1002/pro.3289>

Le, Q. A. T., Joo, J. C., Yoo, Y. J., & Kim, Y. H. (2012). Development of thermostable *Candida antarctica* lipase B through novel in silico design of disulfide bridge. *Biotechnology and Bioengineering*, *109*(4), 867–876. <https://doi.org/10.1002/bit.24371>

Lei, X. (2010). *Phytase with improved thermal stability*.

Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., & Chatonnet, A. (2013). ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: Tools to explore diversity of functions. *Nucleic Acids Research*, *41*(D1), 423–429. <https://doi.org/10.1093/nar/gks1154>

Lespinet, O., & Labedan, B. (2006). ORENZA: A web resource for studying ORphan ENZYme activities. *BMC Bioinformatics*, *7*, 1–11. <https://doi.org/10.1186/1471-2105-7-436>

Li, G., Chen, Y., Fang, X., Su, F., Xu, L., & Yan, Y. (2018a). Identification of a hot-spot to enhance: *Candida rugosa* lipase thermostability by rational design methods. *RSC Advances*, *8*(4), 1948–1957. <https://doi.org/10.1039/c7ra11679a>

Li, G., Chen, Y., Fang, X., Su, F., Xu, L., & Yan, Y. (2018b). Identification of a hot-spot to enhance: *Candida rugosa* lipase thermostability by rational design methods. *RSC Advances*, *8*(4), 1948–1957. <https://doi.org/10.1039/c7ra11679a>

Li, J., Jiang, L., Cao, X., Wu, Y., Lu, F., Liu, F., Li, Y., & Liu, Y. (2021). Improving the activity and stability of *Bacillus clausii* alkaline protease using directed evolution and molecular dynamics simulation. *Enzyme and Microbial Technology*, *147*(29). <https://doi.org/10.1016/j.enzmictec.2021.109787>

Bibliography

Li, J., Li, X., Gai, Y., Sun, Y., & Zhang, D. (2019a). Evolution of e. coli phytase for increased thermostability guided by rational parameters. *Journal of Microbiology and Biotechnology*, 29(3), 419–428. <https://doi.org/10.4014/JMB.1811.11017>

Li, J., Li, X., Gai, Y., Sun, Y., & Zhang, D. (2019b). Evolution of E. coli Phytase for Increased Thermostability Guided by Rational Parameters. *Journal of Microbiology and Biotechnology*, 29(3), 419–428. <https://doi.org/10.4014/jmb.1811.11017>

Li, L., Mao, X., Deng, F., Wang, Y., & Wang, F. (2022). Improving Both the Thermostability and Catalytic Efficiency of Phospholipase D from *Moritella* sp. JT01 through Disulfide Bond Engineering Strategy. *International Journal of Molecular Sciences*, 23(19), 1–13. <https://doi.org/10.3390/ijms231911319>

Li, Q., Yang, X., Li, J., Li, M., Li, C., & Yao, T. (2023). In-depth characterization of phytase-producing plant growth promotion bacteria isolated in alpine grassland of Qinghai-Tibetan Plateau. *Frontiers in Microbiology*, 13(January), 1–18. <https://doi.org/10.3389/fmicb.2022.1019383>

Li, S., Yang, X., Yang, S., Zhu, M., & Wang, X. (2012). Technology prospecting on enzymes: Application, marketing and engineering. *Computational and Structural Biotechnology Journal*, 2(3), e201209017. <https://doi.org/10.5936/csbj.201209017>

Liew, K. J., Lim, C. C., Chan, C. S., Wei, K. Y., Salleh, M. M., Sani, R. K., Chan, K. G., & Goh, K. M. (2017). Direct Cellulase Gene Amplification From Hot Spring Using the Guidance of 16S rRNA Amplicon Metagenomics. In *Metagenomics: Perspectives, Methods, and Applications*. Elsevier Inc. <https://doi.org/10.1016/B978-0-08-102268-9.00016-1>

Lim, D., Golovan, S., Forsberg, C. W., & Jia, Z. (2000). Crystal structures of *Escherichia coli* phytase and its complex with phytate. *Nature Structural Biology*, 7(2), 108–113. <https://doi.org/10.1038/72371>

Lingner, T., Aßhauer, K. P., Schreiber, F., & Meinicke, P. (2011). CoMet - A web server for comparative functional profiling of metagenomes. *Nucleic Acids Research*, 39(SUPPL. 2), 518–523. <https://doi.org/10.1093/nar/gkr388>

Liu, L., Deng, Z., Yang, H., Li, J., Shin, H. D., Chen, R. R., Du, G., & Chen, J. (2014). In silico rational design and systems engineering of disulfide bridges in the catalytic domain of an

Bibliography

alkaline α -amylase from *Alkalimonas amylolytica* to improve thermostability. *Applied and Environmental Microbiology*, *80*(3), 798–807. <https://doi.org/10.1128/AEM.03045-13>

Liu, Y., Li, Z. Y., Guo, C., Cui, C., Lin, H., & Wu, Z. L. (2021). Enhancing the thermal stability of ketoreductase ChKRED12 using the FireProt web server. *Process Biochemistry*, *101*, 207–212. <https://doi.org/10.1016/j.procbio.2020.11.018>

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. <https://doi.org/10.1093/nar/gkz991>

Luo, X., Wang, Y., Zheng, W., Sun, X., Hu, G., Yin, L., Zhang, Y., Yin, F., & Fu, Y. (2022). Simultaneous improvement of the thermostability and activity of lactic dehydrogenase from *Lactobacillus rossiae* through rational design. *RSC Advances*, *12*(51), 33251–33259. <https://doi.org/10.1039/d2ra05599f>

Madej, T., Address, K. J., Fong, J. H., Geer, L. Y., Geer, R. C., Lanczycki, C. J., Liu, C., Lu, S., Marchler-Bauer, A., Panchenko, A. R., Chen, J., Thiessen, P. A., Wang, Y., Zhang, D., & Bryant, S. H. (2012). MMDB: 3D structures and macromolecular interactions. *Nucleic Acids Research*, *40*(D1), 461–464. <https://doi.org/10.1093/nar/gkr1162>

Maenpuen, S., Pongsupasa, V., Pensook, W., Anuwat, P., Kraivisitkul, N., Pinthong, C., Phonbuppha, J., Luanloet, T., Wijma, H. J., Fraaije, M. W., Lawan, N., Chaiyen, P., & Wongnate, T. (2020). Rational-design engineering to create the flavin reductase variants with thermostable and solvent-tolerant properties. *ChemBioChem*, *21*(10), 1481–1491. <https://doi.org/10.1002/cbic.201900737>

Magrane, M., & Consortium, U. P. (2011). UniProt Knowledgebase: A hub of integrated protein data. *Database*, *2011*, 1–13. <https://doi.org/10.1093/database/bar009>

Makolomakwa, M., Puri, A. K., Permaul, K., & Singh, S. (2017). Thermo-acid-stable phytase-mediated enhancement of bioethanol production using *Colocasia esculenta*. *Bioresource Technology*, *235*, 396–404. <https://doi.org/10.1016/j.biortech.2017.03.157>

Maleki, M., Shahraki, M. F., Kavousi, K., Ariaeenejad, S., & Hosseini Salekdeh, G. (2020). A novel thermostable cellulase cocktail enhances lignocellulosic bioconversion and biorefining

Bibliography

in a broad range of pH. *International Journal of Biological Macromolecules*, 154, 349–360. <https://doi.org/10.1016/j.ijbiomac.2020.03.100>

Marco Alexander Fraatz, M. R. and H. Z. (2014). Food and Feed Enzymes. *Advances in Biochemical Engineering/Biotechnology*, 143, 229–256. <https://doi.org/10.1007/10>

Maulana, H., Widyastuti, Y., Herlina, N., Hasbuna, A., Syiarudin, A., Al, H., Triratna, L., & Mayasari, N. (2023). Bioinformatics study of phytase from *Aspergillus niger* for use as feed additive in livestock feed. *Journal of Genetic Engineering and Biotechnology*. <https://doi.org/10.1186/s43141-023-00600-y>

McDonald, A. G., Boyce, S., & Tipton, K. F. (2009). ExplorEnz: The primary source of the IUBMB enzyme list. *Nucleic Acids Research*, 37(SUPPL. 1), 593–597. <https://doi.org/10.1093/nar/gkn582>

McGuffin, L. J., Adiyaman, R., Maghrabi, A. H. A., Shuid, A. N., Brackenridge, D. A., Nealon, J. O., & Philomina, L. S. (2019). IntFOLD: An integrated web resource for high performance protein structure and function prediction. *Nucleic Acids Research*, 47(W1), W408–W413. <https://doi.org/10.1093/nar/gkz322>

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404–405. <https://doi.org/10.1093/bioinformatics/16.4.404>

Menezes-Blackburn, D., Jorquera, M. A., Greiner, R., Gianfreda, L., & De La Luz Mora, M. (2013). Phytases and phytase-labile organic phosphorus in manures and soils. *Critical Reviews in Environmental Science and Technology*, 43(9), 916–954. <https://doi.org/10.1080/10643389.2011.627019>

Meng, Z., Ma, J., Sun, Z., Yang, C., Leng, J., Zhu, W., & Cheng, Y. (2023). Characterization of a novel bifunctional enzyme from buffalo rumen metagenome and its effect on in vitro ruminal fermentation and microbial community composition. *Animal Nutrition*, 13, 137–149. <https://doi.org/10.1016/j.aninu.2023.01.004>

Mitchell, A. L., Scheremetjew, M., Denise, H., Potter, S., Tarkowska, A., Qureshi, M., Salazar, G. A., Pesseat, S., Boland, M. A., Hunter, F. M. I., Ten Hoopen, P., Alako, B., Amid, C., Wilkinson, D. J., Curtis, T. P., Cochrane, G., & Finn, R. D. (2018). EBI Metagenomics in 2017: Enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Research*, 46(D1), D726–D735. <https://doi.org/10.1093/nar/gkx967>

Bibliography

Montgomerie, S., Cruz, J. A., Shrivastava, S., Arndt, D., Berjanskii, M., & Wishart, D. S. (2008). PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation. *Nucleic Acids Research*, *36*(Web Server issue), 202–209. <https://doi.org/10.1093/nar/gkn255>

Mosca, R., Céol, A., Stein, A., Olivella, R., & Aloy, P. (2014). 3did: A catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Research*, *42*(D1), 374–379. <https://doi.org/10.1093/nar/gkt887>

Motahar, S. F. S., Khatibi, A., Salami, M., Ariaeenejad, S., Emam-Djomeh, Z., Nedaei, H., Kavousi, K., Sheykhabdolahzadeh Mamaghani, A., & Salekdeh, G. H. (2020). A novel metagenome-derived thermostable and poultry feed compatible α -amylase with enhanced biodegradation properties. *International Journal of Biological Macromolecules*, *164*, 2124–2133. <https://doi.org/10.1016/j.ijbiomac.2020.08.064>

Motamedi, H. (2016). Screening cabbage rhizosphere as a habitat for isolation of phosphate-solubilizing bacteria. *Environmental and Experimental Biology*, *14*(4), 173–181. <https://doi.org/10.22364/eeb.14.24>

Moushree Pal Roy, Subhabrata Datta, S. G. (2017). A novel extracellular low-temperature active phytase from *Bacillus aryabhatai* RS1 with potential application in plant growth. *Biotechnology Progress*, 1–26. <https://doi.org/10.1002/btpr>.

Mulder, N., & Apweiler, R. (2007). InterPro and InterProScan. *Comparative Genomics*, *2*, 59–70. https://doi.org/10.1007/978-1-59745-515-2_5

Musil, M., Stourac, J., Bendl, J., Brezovsky, J., Prokop, Z., Zendulka, J., Martinek, T., Bednar, D., & Damborsky, J. (2017). FireProt: web server for automated design of thermostable proteins. *Nucleic Acids Research*, *45*(W1), W393–W399. <https://doi.org/10.1093/nar/gkx285>

Musumeci, M. A., Lozada, M., Rial, D. V., Cormack, W. P. M., Jansson, J. K., Sjöling, S., Carroll, J. L., & Dionisi, H. M. (2017). Prospecting biotechnologically-relevant monooxygenases from cold sediment metagenomes: An in silico approach. *Marine Drugs*, *15*(114), 2–19. <https://doi.org/10.3390/md15040114>

Muwawa, E. M., Obieze, C. C., Makonde, H. M., Jefwa, J. M., Kahindi, J. H. P., & Khasa, D. P. (2021). 16S rRNA gene amplicon-based metagenomic analysis of bacterial communities in

Bibliography

the rhizospheres of selected mangrove species from Mida Creek and Gazi Bay, Kenya. *PLoS ONE*, *16*(3), 1–22. <https://doi.org/10.1371/journal.pone.0248485>

Nagano, N. (2005). EzCatDB: The Enzyme Catalytic-mechanism Database. *Nucleic Acids Research*, *33*(DATABASE ISS.), 407–412. <https://doi.org/10.1093/nar/gki080>

Nagar, A., Kamble, A., & Singh, H. (2021). Preliminary screening, isolation and identification of microbial phytase producers from soil. *Environmental and Experimental Biology*, *19*(1), 11–22. <https://doi.org/10.22364/eeb.19.03>

Narayan, N. R., Weinmaier, T., Laserna-Mendieta, E. J., Claesson, M. J., Shanahan, F., Dabbagh, K., Iwai, S., & Desantis, T. Z. (2020). Piphillin predicts metagenomic composition and dynamics from DADA2- corrected 16S rDNA sequences (BMC Genomics (2020) 21 (56) DOI: 10.1186/s12864-019-6427-1). *BMC Genomics*, *21*(1), 1–12. <https://doi.org/10.1186/s12864-020-6537-9>

Nasko, D. J., Wommack, K. E., Ferrell, B. D., & Polson, S. W. (2018). Fast and sensitive protein sequence homology searches using hierarchical cluster BLAST. *BioRxiv*, *19711*(302), 1–23. <https://doi.org/10.1101/426098>

Nevskaya, E. V., Tyurina, I. A., Borodulin, D. M., Shulbaeva, M. T., & Sumina, A. V. (2021). Prospects for the use of phytase to keep the freshness of bread. *IOP Conference Series: Earth and Environmental Science*, *640*(2). <https://doi.org/10.1088/1755-1315/640/2/022075>

Nezhad, N. G., Noor, R., Raja, Z., Rahman, A., Normi, Y. M., Oslan, S. N., Mohd, F., & Leow, T. C. (2020). Integrative Structural and Computational Biology of Phytases for the Animal Feed Industry. *Catalysts*, *10*(844), 1–24.

Nezhad, N. G., Rahman, R. N. Z. R. A., Normi, Y. M., Oslan, S. N., Shariff, F. M., & Leow, T. C. (2023). Isolation, screening and molecular characterization of phytase-producing microorganisms to discover the novel phytase. *Biologia*, *78*, 2527–2537. <https://doi.org/10.1007/s11756-023-01391-w>

Nirwan, S., Chahal, V., & Kakkar, R. (2020). Structure-based virtual screening, free energy of binding and molecular dynamics simulations to propose novel inhibitors of Mtb-MurB oxidoreductase enzyme. *Journal of Biomolecular Structure and Dynamics*. <https://doi.org/10.1080/07391102.2020.1712258>

Bibliography

Niu, C., Zhu, L., Xu, X., & Li, Q. (2016). Rational design of disulfide bonds increases thermostability of a mesophilic 1,3-1,4- β -glucanase from *Bacillus terquilensis*. *PLoS ONE*, *11*(4). <https://doi.org/10.1371/journal.pone.0154036>

Ojha, B. K., Singh, P. K., & Shrivastava, N. (2018). Enzymes in the animal feed industry. *Enzymes in Food Biotechnology: Production, Applications, and Future Prospects*, 93–109. <https://doi.org/10.1016/B978-0-12-813280-7.00007-4>

Outchkourov, N., & Petkov, S. (2019). Phytases for feed applications. In *Industrial Enzyme Applications* (pp. 255–285). https://doi.org/10.1002/9783527813780.ch3_3

Oyewusi, H. A., Akinyede, K. A., Abdul Wahab, R., & Huyop, F. (2023). In silico analysis of a putative dehalogenase from the genome of halophilic bacterium *Halomonas smyrnensis* AAD6T. *Journal of Biomolecular Structure and Dynamics*, *41*(1), 319–335. <https://doi.org/10.1080/07391102.2021.2006085>

Parra-Cruz, R., Jäger, C. M., Lau, P. L., Gomes, R. L., & Pordea, A. (2018). Rational Design of Thermostable Carbonic Anhydrase Mutants Using Molecular Dynamics Simulations. *Journal of Physical Chemistry B*, *122*(36), 8526–8536. <https://doi.org/10.1021/acs.jpcc.8b05926>

Passardi, F., Theiler, G., Zamocky, M., Cosio, C., Rouhier, N., Teixeira, F., Margis-Pinheiro, M., Ioannidis, V., Penel, C., Falquet, L., & Dunand, C. (2007). PeroxiBase: The peroxidase database. *Phytochemistry*, *68*(12), 1605–1611. <https://doi.org/10.1016/j.phytochem.2007.04.005>

Perez Rojo, F., Pillow, J. J., & Kaur, P. (2023). Bioprospecting microbes and enzymes for the production of pterocarpan and coumestans. *Frontiers in Bioengineering and Biotechnology*, *11*, 1–16. <https://doi.org/10.3389/fbioe.2023.1154779>

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A Visualization System for Exploratory Research. *Journal of Computational Chemistry*, *25*(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>

Pongpamorn, P., Wathaisong, P., Pimviriyakul, P., Jaruwat, A., Lawan, N., Chitnumsub, P., & Chaiyen, P. (2019). Identification of a Hotspot Residue for Improving the Thermostability of a Flavin-Dependent Monooxygenase. *ChemBioChem*, *20*(24), 3020–3031. <https://doi.org/10.1002/cbic.201900413>

Bibliography

Porollo, A. A., Adamczak, R., & Meller, J. (2004). POLYVIEW: A flexible visualization tool for structural and functional annotations of proteins. *Bioinformatics*, *20*(15), 2460–2462. <https://doi.org/10.1093/bioinformatics/bth248>

Powar, V. K., & Jagannathan, V. (1982). Purification and properties of phytate-specific phosphatase from *Bacillus subtilis*. *Journal of Bacteriology*.

Pramanik, K., Kumar, P., Ray, S., Sarkar, A., Mitra, S., & Kanti, T. (2017). An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of *Pseudomonas aeruginosa*. *Journal of Genetic Engineering and Biotechnology*, 1–11. <https://doi.org/10.1016/j.jgeb.2017.05.003>

Pramanik, K., Kundu, S., Banerjee, S., Ghosh, P. K., & Maiti, T. K. (2018). Computational-based structural, functional and phylogenetic analysis of *Enterobacter* phytases. *3 Biotech*, *8*(6), 1–12. <https://doi.org/10.1007/s13205-018-1287-y>

Pronk, S. (2013). *Molecular Simulation Methods with Gromacs*. 1–14. <http://www.gromacs.org/@api/deki/files/198/=gmx-tutorial.pdf>

Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Research*, *40*(D1), 130–135. <https://doi.org/10.1093/nar/gkr1079>

Pucci, F., Kwasigroch, J. M., & Rooman, M. (2017). SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics (Oxford, England)*, *33*(21), 3415–3422. <https://doi.org/10.1093/bioinformatics/btx417>

Puppala, K. R., Bhavsar, K., Sonalkar, V., Khire, J. M., & Dharne, M. S. (2019). Characterization of novel acidic and thermostable phytase secreting *Streptomyces* sp. (NCIM 5533) for plant growth promoting characteristics. *Biocatalysis and Agricultural Biotechnology*, *18*(Ncim 5533), 101020. <https://doi.org/10.1016/j.bcab.2019.101020>

Rao, D. E. C. S., Rao, K. V., Reddy, T. P., & Reddy, V. D. (2009). Molecular characterization, physicochemical properties, known and potential applications of phytases: An overview. *Critical Reviews in Biotechnology*, *29*(2), 182–198. <https://doi.org/10.1080/07388550902919571>

Bibliography

Rappoport, N., Karsenty, S., Stern, A., Linial, N., & Linial, M. (2012). ProtoNet 6.0: Organizing 10 million protein sequences in a compact hierarchical family tree. *Nucleic Acids Research*, *40*(D1), 313–320. <https://doi.org/10.1093/nar/gkr1027>

Rathour, R., Patel, D. H., Madamwar, D., & Desai, C. (2023). Metagenomic analysis of a bioreactor biofilm treating raw textile effluent and biochemical characterization of a novel azoreductase gene. *Biocatalysis and Agricultural Biotechnology*, *53*(July), 102876. <https://doi.org/10.1016/j.bcab.2023.102876>

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, *46*(D1), D624–D632. <https://doi.org/10.1093/nar/gkx1134>

Rebello, S., Jose, L., Sindhu, R., & Aneesh, E. M. (2017). Molecular advancements in the development of thermostable phytases. In *Applied Microbiology and Biotechnology* (Vol. 101, Issue 7, pp. 2677–2689). Applied Microbiology and Biotechnology. <https://doi.org/10.1007/s00253-017-8195-7>

Ribeiro, A. J. M., Holliday, G. L., Furnham, N., Tyzack, J. D., Ferris, K., & Thornton, J. M. (2018). Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Research*, *46*(D1), D618–D623. <https://doi.org/10.1093/nar/gkx1012>

Robe, P., Nalin, R., Capellano, C., Vogel, T. M., & Simonet, P. (2003). Extraction of DNA from soil. *European Journal of Soil Biology*, *39*(4), 183–190. [https://doi.org/10.1016/S1164-5563\(03\)00033-5](https://doi.org/10.1016/S1164-5563(03)00033-5)

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, *43*(D1), D298–D299. <https://doi.org/10.1093/nar/gku1046>

Roche, D. B., Tetchner, S. J., & McGuffin, L. J. (2011). FunFOLD: An improved automated method for the prediction of ligand binding residues using 3D models of proteins. *BMC Bioinformatics*, *12*(1), 160. <https://doi.org/10.1186/1471-2105-12-160>

Bibliography

Rolf, J., Siedentop, R., Lütz, S., & Rosenthal, K. (2020). Screening and identification of novel cGAS homologues using a combination of in vitro and in vivo protein synthesis. *International Journal of Molecular Sciences*, *21*(1). <https://doi.org/10.3390/ijms21010105>

ROMAN A. LASKOWSKI, MALCOLM W. MACARTHUR, DAVID S. MOSS, J. M. T. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, *26*. <https://doi.org/10.1107/S0021889892009944>

Rosen, G. L., Reichenberger, E. R., & Rosenfeld, A. M. (2011). NBC: The naïve Bayes classification tool webserver for taxonomic classification of metagenomic reads. *Bioinformatics*, *27*(1), 127–129. <https://doi.org/10.1093/bioinformatics/btq619>

Rui, Wang, S., Xu, Y., & Yu, X. (2020). Enhancing the thermostability of *Rhizopus chinensis* lipase by rational design and MD simulations. *International Journal of Biological Macromolecules*, *160*, 1189–1200. <https://doi.org/10.1016/j.ijbiomac.2020.05.243>

Sadeghian, I., Rezaie, Z., Rahmatabadi, S. S., & Hemmati, S. (2020). Biochemical insights into a novel thermo/organo tolerant bilirubin oxidase from *Thermosediminibacter oceani* and its application in dye decolorization. *Process Biochemistry*, *88*, 38–50. <https://doi.org/10.1016/j.procbio.2019.09.030>

Sahoo, R. K., Kumar, M., Sukla, L. B., & Subudhi, E. (2017). Bioprospecting hot spring metagenome: lipase for the production of biodiesel. *Environmental Science and Pollution Research*, *24*(4), 3802–3809. <https://doi.org/10.1007/s11356-016-8118-7>

Sahoo, S., Mahapatra, S. R., Das, N., Parida, B. K., Rath, S., Misra, N., & Suar, M. (2020). Functional elucidation of hypothetical proteins associated with lipid accumulation: Prioritizing genetic engineering targets for improved algal biofuel production. *Algal Research*, *47*(February), 101887. <https://doi.org/10.1016/j.algal.2020.101887>

Şahutoğlu, A. S., Duman, H., Frese, S. A., & Karav, S. (2020). Structural insights of two novel N-acetyl-glucosaminidase enzymes through in silico methods. *Turkish Journal of Chemistry*, *44*(6), 1703–1712. <https://doi.org/10.3906/kim-2006-19>

Sajidan, A., Farouk, A., Greiner, R., Jungblut, P., Müller, E. C., & Borriss, R. (2004). Molecular and physiological characterisation of a 3-phytase from soil bacterium *Klebsiella* sp. ASR1. *Applied Microbiology and Biotechnology*, *65*(1), 110–118. <https://doi.org/10.1007/s00253-003-1530-1>

Bibliography

Sajidan, Wulandari, R., Sari, E. N., Ratriyanto, A., Weldekiros, H., & Greiner, R. (2015). Phytase-Producing Bacteria from Extreme Regions in Indonesia. *Brazilian Archives of Biology and Technology*, 58(5), 711–717. <https://doi.org/10.1590/S1516-89132015050173>

Sanchez, S., & Demain, A. L. (2011). Enzymes and bioconversions of industrial, pharmaceutical, and biotechnological significance. *Organic Process Research and Development*, 15(1), 224–230. <https://doi.org/10.1021/op100302x>

Sanchez-Romero, I., Ariza, A., Wilson, K. S., Skjøt, M., Vind, J., de Maria, L., Skov, L. K., & Sanchez-Ruiz, J. M. (2013). Mechanism of Protein Kinetic Stabilization by Engineered Disulfide Crosslinks. *PLoS ONE*, 8(7), 1–9. <https://doi.org/10.1371/journal.pone.0070013>

Sarmiento, F., Peralta, R., & Blamey, J. M. (2015). Cold and Hot Extremozymes: Industrial Relevance and Current Trends. *Frontiers in Bioengineering and Biotechnology*. <https://doi.org/10.3389/fbioe.2015.00148>

Satpathy, R., Konkimalla, V. B., & Ratha, J. (2015). In-silico Rational Protein Engineering and Design Approach to Improve Thermostability of a Haloalkane Dehalogenase Enzyme. *American Journal of Bioinformatics*, 4(2), 34–46. <https://doi.org/10.3844/ajbsp.2015.34.46>

Satyanarayana, S. D. V. O. of high-yielding protocol for D. extraction from the forest rhizosphere microbes, Krishna, M. S. R., & Kumar, P. P. (2017). Optimization of high-yielding protocol for DNA extraction from the forest rhizosphere microbes. *3 Biotech*, 7(2), 1–9. <https://doi.org/10.1007/s13205-017-0737-2>

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., & Schomburg, D. (2004). BRENDA, the enzyme database: Updates and major new developments. *Nucleic Acids Research*, 32(DATABASE ISS.), 431–433. <https://doi.org/10.1093/nar/gkh081>

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research*, 33(SUPPL. 2), 382–388. <https://doi.org/10.1093/nar/gki387>

Shakeel, T., Gupta, M., Fatma, Z., Kumar, R., Kumar, R., Singh, R., Sharma, M., Jade, D., Gupta, D., Fatma, T., & Yazdani, S. S. (2018). A consensus-guided approach yields a heat-stable alkane-producing enzyme and identifies residues promoting thermostability. *The Journal of Biological Chemistry*, 1–30. <https://doi.org/10.1074/jbc.RA117.000639>

Bibliography

Shamim, K., Sharma, J., Mutnale, M., Dubey, S. K., & Mujawar, S. (2018). Characterization of a metagenomic serine metalloprotease and molecular docking studies. In *Process Biochemistry* (Vol. 71). Elsevier Ltd. <https://doi.org/10.1016/j.procbio.2018.05.020>

Sharma, N., Thakur, N., Raj, T., Savitri, & Bhalla, T. C. (2017). Mining of microbial genomes for the novel sources of nitrilases. *BioMed Research International*, 1–14. <https://doi.org/10.1155/2017/7039245>

Sharma, A. D. E. V. (2020). MOLECULAR MODELING AND 3D ANALYSIS OF WATER STRESS RESPONSIVE TAPASE PHOSPHATASE ENCODING GENE IN WHEAT (TRITICUM AESTIVUM). *Agriculture*, 4(3).

Sharma, V. K., Kumar, N., Prakash, T., & Taylor, T. D. (2009). MetaBioME: A database to explore commercially useful enzymes in metagenomic datasets. *Nucleic Acids Research*, 38(SUPPL.1), 468–472. <https://doi.org/10.1093/nar/gkp1001>

Shivange, A. V., Dennig, A., & Schwaneberg, U. (2014). Multi-site saturation by OmniChange yields a pH- and thermally improved phytase. *Journal of Biotechnology*, 170, 68–72. <https://doi.org/10.1016/j.jbiotec.2013.11.014>

Shivange, A. V., Roccatano, D., & Schwaneberg, U. (2016). Iterative key-residues interrogation of a phytase with thermostability increasing substitutions identified in directed evolution. *Applied Microbiology and Biotechnology*, 100(1), 227–242. <https://doi.org/10.1007/s00253-015-6959-5>

Shivange, A. V., Schwaneberg, U., & Roccatano, D. (2010). Conformational dynamics of active site loop in Escherichia coli phytase. *Biopolymers*, 93(11), 994–1002. <https://doi.org/10.1002/bip.21513>

Shivange, A. V., Serwe, A., Dennig, A., Roccatano, D., Haefner, S., & Schwaneberg, U. (2012). Directed evolution of a highly active Yersinia mollaretii phytase. *Applied Microbiology and Biotechnology*, 95(2), 405–418. <https://doi.org/10.1007/s00253-011-3756-7>

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(539). <https://doi.org/10.1038/msb.2011.75>

Bibliography

Singhania, R. R., Patel, A. K., Thomas, L., Goswami, M., Giri, B. S., & Pandey, A. (2015). Industrial Enzymes. In *Industrial Biorefineries and White Biotechnology*. Elsevier B.V. <https://doi.org/10.1016/B978-0-444-63453-5.00015-X>

Smith, A. A., & Caruso, A. (2013). Characterization and Homology Modeling of a Cyanobacterial Phosphoenolpyruvate Carboxykinase Enzyme. *Hindawi*. <https://doi.org/10.1155/2013/370820>

Smith, R. D., Clark, J. J., Ahmed, A., Orban, Z. J., Dunbar, J. B., & Carlson, H. A. (2019). Updates to Binding MOAD (Mother of All Databases): Polypharmacology Tools and Their Utility in Drug Repurposing. *Journal of Molecular Biology*, 431(13), 2423–2433. <https://doi.org/10.1016/j.jmb.2019.05.024>

Soh, L. M. J., Mak, W. S., Lin, P. P., Mi, L., Chen, F. Y. H., Damoiseaux, R., Siegel, J. B., & Liao, J. C. (2017). Engineering a Thermostable Keto Acid Decarboxylase Using Directed Evolution and Computationally Directed Protein Design. *ACS Synthetic Biology*, 6(4), 610–618. <https://doi.org/10.1021/acssynbio.6b00240>

Song, H. Y., El Sheikha, A. F., & Hu, D. M. (2018). The positive impacts of microbial phytase on its nutritional applications. *Trends in Food Science and Technology*, 553–562. <https://doi.org/10.1016/j.tifs.2018.12.001>

Song, J., Tan, H., Perry, A. J., Akutsu, T., Webb, G. I., Whisstock, J. C., & Pike, R. N. (2012). PROSPER: An Integrated Feature-Based Tool for Predicting Protease Substrate Cleavage Sites. *PLoS ONE*, 7(11). <https://doi.org/10.1371/journal.pone.0050300>

Stephane Blesa, Helene Chautard, Delcourt, Laurent Mesta, B. W. (2015). *Enhanced phytase variants*.

Stewart, E. J. (2012). Growing unculturable bacteria. *Journal of Bacteriology*, 194(16), 4151–4160. <https://doi.org/10.1128/JB.00345-12>

Suleimanova, A. D., Beinhauer, A., Valeeva, L. R., Chastukhina, I. B., Balaban, N. P., Shakirov, E. V., Greiner, R., & Sharipova, M. R. (2015). Novel Glucose-1-Phosphatase with High Phytase Activity and Unusual Metal Ion Activation from Soil Bacterium *Pantoea* sp. Strain 3.5.1. *Applied and Environmental Microbiology*, 81(19), 6790–6799. <https://doi.org/10.1128/aem.01384-15>

Bibliography

Sussman, J. L. (1995). *The Swiss-3DImage collection and PDB-Browser on the World-Wide Web COMPUTER. February*, 82–84.

Suzuki, M., Date, M., Kashiwagi, T., Suzuki, E., & Yokoyama, K. (2022). Rational design of a disulfide bridge increases the thermostability of microbial transglutaminase. *Applied Microbiology and Biotechnology*, 106(12), 4553–4562. <https://doi.org/10.1007/s00253-022-12024-8>

Tan, H., Miao, R., Liu, T., Cao, X., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2016). Enhancing the thermal resistance of a novel acidobacteria-derived phytase by engineering of disulfide bridges. *Journal of Microbiology and Biotechnology*, 26(10), 1717–1722. <https://doi.org/10.4014/jmb.1604.04051>

Tan, H., Mooij, M. J., & , Matthieu Barret , Pdraig M. Hegarty , Catriona Harrington , Alan D.W. Dobson, and F. O. (2014). Identification of Novel Phytase Genes from an Agricultural Soil-Derived Metagenome. *J.Microbiol.Biotechnol*, 24(1), 113–118.

Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2015). Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Applied Microbiology and Biotechnology*, 100(5), 2225–2241. <https://doi.org/10.1007/s00253-015-7097-9>

Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2016). A Novel Phytase Derived from an Acidic Peat-Soil Microbiome Showing High Stability under Acidic Plus Pepsin Conditions. *Journal of Molecular Microbiology and Biotechnology*, 26(4), 291–301. <https://doi.org/10.1159/000446567>

Tang, Z., Jin, W., Sun, R., Liao, Y., Zhen, T., Chen, H., Wu, Q., Gou, L., & Li, C. (2018). Improved thermostability and enzyme activity of a recombinant phyA mutant phytase from *Aspergillus niger* N25 by directed evolution and site-directed mutagenesis. *Enzyme and Microbial Technology*, 108(May 2017), 74–81. <https://doi.org/10.1016/j.enzmictec.2017.09.010>

Thebti, W., Riahi, Y., Gharsalli, R., & Belhadj, O. (2016). Screening and characterization of thermo-active enzymes of biotechnological interest produced by thermophilic *Bacillus* isolated from hot springs in Tunisia. *Acta Biochimica Polonica*. https://doi.org/10.18388/abp.2016_1271

Bibliography

Thornbury, M., Sicheri, J., Guinard, C., Mahoney, D., Routledge, F., Curry, M., Elaghil, M., Boudreau, N., Tsai, A., Slaine, P., Getz, L., Cook, J., Rohde, J., McCormick, C., Sicheri, C. J., Thornbury, M., Guinard, C., Slaine, S. P., Getz, L., ... Getz, L. (2018). Discovery and Characterization of Novel Lignocellulose-Degrading Enzymes from the Porcupine Microbiome. *BioRxiv*.

Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Research*, *46*(W1), W363–W367. <https://doi.org/10.1093/nar/gky473>

Tian, Y., Hou, X., Ni, D., Xu, W., Guang, C., Zhang, W., Chen, Q., Rao, Y., & Mu, W. (2022). Structure-based interface engineering methodology in designing a thermostable amylose-forming transglucosylase. *Journal of Biological Chemistry*, *298*(7), 102074. <https://doi.org/10.1016/j.jbc.2022.102074>

Tina, K. G., Bhadra, R., & Srinivasan, N. (2007). PIC: Protein Interactions Calculator. *Nucleic Acids Research*, *35*(SUPPL.2), 473–476. <https://doi.org/10.1093/nar/gkm423>

Touw, W. G., Baakman, C., Black, J., Te Beek, T. A. H., Krieger, E., Joosten, R. P., & Vriend, G. (2015). A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, *43*(D1), D364–D368. <https://doi.org/10.1093/nar/gku1028>

Toyama, D., de Morais, M. A. B., Ramos, F. C., Zanthorlin, L. M., Tonoli, C. C. C., Balula, A. F., de Miranda, F. P., Almeida, V. M., Marana, S. R., Ruller, R., Murakami, M. T., & Henrique-Silva, F. (2018). A novel β -glucosidase isolated from the microbial metagenome of Lake Poraquê (Amazon, Brazil). *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1866*(4), 569–579. <https://doi.org/10.1016/j.bbapap.2018.02.001>

Tran, T. T., Mamo, G., Mattiasson, B., & Hatti-Kaul, R. (2010a). A thermostable phytase from *Bacillus* sp. MD2: Cloning, expression and high-level production in *Escherichia coli*. *Journal of Industrial Microbiology and Biotechnology*. <https://doi.org/10.1007/s10295-009-0671-3>

Tran, T. T., Mamo, G., Mattiasson, B., & Hatti-Kaul, R. (2010b). A thermostable phytase from *Bacillus* sp. MD2: Cloning, expression and high-level production in *Escherichia coli*. *Journal of Industrial Microbiology and Biotechnology*, *37*(3), 279–287. <https://doi.org/10.1007/s10295-009-0671-3>

Bibliography

Tripathi, P., Garg, S., Panwar, D., & Kaira, G. S. (2016). Phytase from *Citrobacter koseri* PM-7 : Cost-Effective Production Using Agro-Industrial Residues , Biochemical Characterization and Application in de-Phytinization. *Waste and Biomass Valorization*. <https://doi.org/10.1007/s12649-016-9662-6>

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), NA-NA. <https://doi.org/10.1002/jcc.21334>

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., & Markley, J. L. (2008). BioMagResBank. *Nucleic Acids Research*, *36*(SUPPL. 1), 402–408. <https://doi.org/10.1093/nar/gkm957>

Vaquero, M. E., De Eugenio, L. I., Martínez, M. J., & Barriuso, J. (2015). A novel CalB-type lipase discovered by fungal genomes mining. *PLoS ONE*, *10*(4), 1–11. <https://doi.org/10.1371/journal.pone.0124882>

Varol, A., Albayrak, S., Ozkan, H., Demir, Y., Taskin, M., & Adiguzel, A. (2023). Production, purification and characterization of novel fibrinolytic enzyme from *Bacillus atrophaeus* V4. *Biologia*, *78*(2), 591–600. <https://doi.org/10.1007/s11756-022-01281-7>

Vashishth, A., Ram, S., & Beniwal, V. (2017). Cereal phytases and their importance in improvement of micronutrients bioavailability. In *3 Biotech* (Vol. 7, Issue 1, pp. 1–7). Springer Berlin Heidelberg. <https://doi.org/10.1007/s13205-017-0698-5>

Verma, N. K. V. and V. (2014). Sequence and Structure Analysis of Pyranose Dehydrogenase in *Agaricus campestris* through Insilico methods. *Journal of Advanced Bioinformatics Applications and Research*, *5*(3), 197–205.

Victorino da Silva Amatto, I., Gonsales da Rosa-Garzon, N., Antônio de Oliveira Simões, F., Santiago, F., Pereira da Silva Leite, N., Raspante Martins, J., & Cabral, H. (2022). Enzyme engineering and its industrial applications. *Biotechnology and Applied Biochemistry*, *69*(2), 389–409. <https://doi.org/10.1002/bab.2117>

Voß, H., Heck, A., & Schallmeyer, M. (2020). Database Mining for Novel Bacterial beta-Etherases, Glutathione- Dependent Lignin-Degrading Enzymes. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *86*(2), 1–15.

Bibliography

Wang, C. Y., Chang, P. M., Ary, M. L., Allen, B. D., Chica, R. A., Mayo, S. L., & Olafson, B. D. (2018). ProtaBank: A repository for protein design and engineering data. *Protein Science*, 27(6), 1113–1124. <https://doi.org/10.1002/pro.3406>

Wang, Q., Liu, S., Li, K., Xing, R., & Chen, X. (2023). A Computational Biology Study on the Structure and Dynamics Determinants of Thermal Stability of the Chitosanase from *Aspergillus fumigatus*. *Int. J. Mol. Sci.*, 24.

Wang, X., Yang, H., Ruan, L., Liu, X., Li, F., & Xu, X. (2008). Cloning and characterization of a thermostable superoxide dismutase from the thermophilic bacterium *Rhodothermus* sp. XMH10. *Journal of Industrial Microbiology and Biotechnology*. <https://doi.org/10.1007/s10295-007-0274-9>

Wang, X., Yao, M., Yang, B., Fu, Y., Hu, F., & Liang, A. (2015). Enzymology and thermal stability of phytase appA mutants. *RSC Advances*, 5(54), 43863–43872. <https://doi.org/10.1039/c5ra02199e>

Wiederstein, M., & Sippl, M. J. (2007a). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(SUPPL.2), 407–410. <https://doi.org/10.1093/nar/gkm290>

Wiederstein, M., & Sippl, M. J. (2007b). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(SUPPL.2), 407–410. <https://doi.org/10.1093/nar/gkm290>

Wittig, U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J., & Rojas, I. (2006). SABIO-RK: Integration and curation of reaction kinetics data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4075, 94–103. https://doi.org/10.1007/11799511_9

Wu, C. H., Huang, H., Nikolskaya, A., Hu, Z., & Barker, W. C. (2004). The iProClass integrated database for protein functional analysis. *Computational Biology and Chemistry*, 28(1), 87–96. <https://doi.org/10.1016/j.compbiolchem.2003.10.003>

Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: A customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, 12(Cd). <https://doi.org/10.1186/1471-2164-12-444>

Bibliography

Wu, T., Chen, C., Cheng, Y., Ko, T., Lin, C., Lai, H., Huang, T., Liu, J., & Guo, R. (2014). Improving specific activity and thermostability of Escherichia coli phytase by structure-based rational design. *Journal of Biotechnology*, *175*, 1–6. <https://doi.org/10.1016/j.jbiotec.2014.01.034>

Xi Wang & Jun Du & Zhi-yun Zhang & Yue-jun Fu & Wen-ming Wang & Ai-Hua Liang. (2018). A rational design to enhance the resistance of Escherichia coli phytase appA to trypsin. *Applied Microbiology and Biotechnology*, *102*(22), 9647–9656. <https://doi.org/10.1007/s00253-018-9327-4>

Xing, H., Wang, P., Yan, X., Yang, Y., Li, X., Liu, R., & Zhou, Z. (2023). Thermostability enhancement of Escherichia coli phytase by error-prone polymerase chain reaction (epPCR) and site-directed mutagenesis. *Frontiers in Bioengineering and Biotechnology*, *11*(March), 1–10. <https://doi.org/10.3389/fbioe.2023.1167530>

Xiong, N., Lv, P. J., Song, J. W., Shen, Q., Xue, Y. P., & Zheng, Y. G. (2022). Engineering of a nitrilase through consensus sequence analysis and conserved site substitution to improve its thermostability and activity. *Biochemical Engineering Journal*, *184*(May), 108475. <https://doi.org/10.1016/j.bej.2022.108475>

Xu, D., Jaroszewski, L., Li, Z., & Godzik, A. (2014). FFAS-3D: Improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics*, *30*(5), 660–667. <https://doi.org/10.1093/bioinformatics/btt578>

Xu, Z., Cen, Y. K., Zou, S. P., Xue, Y. P., & Zheng, Y. G. (2019). Recent advances in the improvement of enzyme thermostability by structure modification. *Critical Reviews in Biotechnology*, *40*(1), 83–98. <https://doi.org/10.1080/07388551.2019.1682963>

Yan, W., Li, F., Wang, L., Zhu, Y., Dong, Z., & Bai, L. (2017). Discovery and characterization of a novel lipase with transesterification activity from hot spring metagenomic library. *Biotechnology Reports*, *14*, 27–33. <https://doi.org/10.1016/j.btre.2016.12.007>

Yang, H. M., Yao, B., Meng, K., Wang, Y. R., Bai, Y. G., & Wu, N. F. (2007). Introduction of a disulfide bridge enhances the thermostability of a Streptomyces olivaceoviridis xylanase mutant. *Journal of Industrial Microbiology and Biotechnology*, *34*(3), 213–218. <https://doi.org/10.1007/s10295-006-0188-y>

Bibliography

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2014). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. <https://doi.org/10.1038/nmeth.3213>

Yang, W., Sun, L., Dong, P., Chen, Y., Zhang, H., Huang, X., Wu, L., Chen, L., Jing, D., & Wu, Y. (2022). Structure-guided rational design of the *Geobacillus thermoglucosidasius* feruloyl esterase GthFAE to improve its thermostability. *Biochemical and Biophysical Research Communications*, *600*, 117–122. <https://doi.org/10.1016/j.bbrc.2022.02.074>

Yang, Y., Sun, J., Wu, J., Zhang, L., Du, L., Matsukawa, S., Xie, J., & Wei, D. (2016). Characterization of a Novel α -L-Arabinofuranosidase from *Ruminococcus albus* 7 and Rational Design for Its Thermostability. *Journal of Agricultural and Food Chemistry*. <https://doi.org/10.1021/acs.jafc.6b02482>

Yata, V. K., Biswas, A. D., Deb, A., Sanjeev, A., & Mattaparthi, V. S. K. (2021). Identification of *cucumis sativus* urease as a potential urea binding enzyme by computational methods. *Biointerface Research in Applied Chemistry*, *11*(2), 9184–9200. <https://doi.org/10.33263/BRIAC112.91849200>

Yellaboina, S., Tasneem, A., Zaykin, D. V., Raghavachari, B., & Jothi, R. (2011). DOMINE: A comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Research*, *39*(SUPPL. 1), 730–735. <https://doi.org/10.1093/nar/gkq1229>

Yi, Y., Fang, Y., Wu, K., Liu, Y., & Zhang, W. (2020). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Oncology Letters*, *19*(4), 3316–3332. <https://doi.org/10.3892/ol.2020.11439>

Yin, X., Hu, D., Li, J. F., He, Y., Zhu, T. Di, & Wu, M. C. (2015). Contribution of disulfide bridges to the thermostability of a type a feruloyl esterase from *Aspergillus usamii*. *PLoS ONE*, *10*(5), 1–16. <https://doi.org/10.1371/journal.pone.0126864>

Yinglan Wang, Caiming Li, Xiaofeng Ban, Zhengbiao Gu, Yan Hong, L. C. and Z. L. (2022). *Disulfide Bond Engineering for Enhancing the Thermostability of the Maltotetraose-Forming Amylase from Pseudomonas*.

Yu, C., Zavaljevski, N., Desai, V., & Reifman, J. (2008). Genome-wide enzyme annotation with precision control: Catalytic families (CatFam) databases. *Proteins: Structure, Function, and Bioinformatics*, *74*(2), 449–460. <https://doi.org/10.1002/prot.22167>

Bibliography

Zhang, G., Liu, P., Zhang, L., Wei, W., Wang, X., Wei, D., & Wang, W. (2016). Bioprospecting metagenomics of a microbial community on cotton degradation: Mining for new glycoside hydrolases. *Journal of Biotechnology*. <https://doi.org/10.1016/j.jbiotec.2016.07.017>

Zhang, H., Sang, J., Zhang, Y., Sun, T., Liu, H., Yue, R., Zhang, J., Wang, H., Dai, Y., Lu, F., & Liu, F. (2019). Rational design of a *Yarrowia lipolytica* derived lipase for improved thermostability. *International Journal of Biological Macromolecules*, *137*, 1190–1198. <https://doi.org/10.1016/j.ijbiomac.2019.07.070>

Zhang, J., Liu, Y., Gao, S., Zhu, L., Li, W., Tian, X., & Liu, Y. (2016). Site-directed mutagenesis and thermal stability analysis of phytase from *Escherichia coli*. *Bioscience Biotechnology Research Communications*, *9*(3), 357–365. <https://doi.org/10.21786/bbrc/9.3/4>

ZHANG Jian-zhi, FU Li-hao, TANG Ting, ZHANG Song-ya, ZHU Jing, LI Tuo, WANG Zi-ning, S. T. (2020). Scalable Enzyme Mining via Synthetic Biology. *Synthetic Biology Journal*, 1–18.

Zhang, X., Li, W., Pan, L., Yang, L., Li, H., Ji, F., Zhang, Y., Tang, H., & Yang, D. (2022). Improving the thermostability of alginate lyase FlAlyA with high expression by computer-aided rational design for industrial preparation of alginate oligosaccharides. *Frontiers in Bioengineering and Biotechnology*, *10*(September), 1–11. <https://doi.org/10.3389/fbioe.2022.1011273>

Zhang, Y., Aryee, A. N., & Simpson, B. K. (2020). Current role of in silico approaches for food enzymes. *Current Opinion in Food Science*, *31*, 63–70. <https://doi.org/10.1016/j.cofs.2019.11.003>

Zhao, C., Chu, Y., Li, Y., Yang, C., Chen, Y., Wang, X., & Liu, B. (2017). High-throughput pyrosequencing used for the discovery of a novel cellulase from a thermophilic cellulose-degrading microbial consortium. *Biotechnology Letters*, *39*(1), 123–131. <https://doi.org/10.1007/s10529-016-2224-y>

Zhao, J., Chen, J., Wang, H., Guo, Y., Li, K., & Liu, J. (2021). Enhanced thermostability of d-psicose 3-epimerase from *Clostridium boltae* through rational design and engineering of new disulfide bridges. *International Journal of Molecular Sciences*, *22*(18). <https://doi.org/10.3390/ijms221810007>

Bibliography

Zhou, X., Xu, Z., Li, Y., He, J., & Zhu, H. (2022). Improvement of the Stability and Activity of an LPMO Through Rational Disulfide Bonds Design. *Frontiers in Bioengineering and Biotechnology*, 9(January), 1–9. <https://doi.org/10.3389/fbioe.2021.815990>

Zhu, F., Li, G., Wei, P., Song, C., Xu, Q., Ma, M., Ma, J., Song, P., & Zhang, S. (2023). Rational engineering of a metalloprotease to enhance thermostability and activity. *Enzyme and Microbial Technology*, 162(September 2022), 110123. <https://doi.org/10.1016/j.enzmictec.2022.110123>

Zhu, W., Qin, L., Xu, Y., Lu, H., Wu, Q., Li, W., Zhang, C., & Li, X. (2023). Three Molecular Modification Strategies to Improve the Thermostability of Xylanase XynA from *Streptomyces rameus* L2001. *Foods*, 12(4). <https://doi.org/10.3390/foods12040879>

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, 430(15), 2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>

Zong, Z., Gao, L., Cai, W., Yu, L., Cui, C., Chen, S., & Zhang, D. (2015). Computer-Assisted Rational Modifications to Improve the Thermostability of β -Glucosidase from *Penicillium piceum* H16. *Bioenergy Research*, 8(3), 1384–1390. <https://doi.org/10.1007/s12155-015-9603-4>

Appendix

Appendix

Appendix

- **Phytase screening broth (PSB) composition:**

Components	gm/L of Distilled water and respective (%)
D-glucose	10 (1%)
Sodium phytate (Himedia)	4 (0.4%)
Calcium chloride (CaCl ₂)	2 (0.2%)
Ammonium nitrate (NH ₄ NO ₃)	5 (0.5%)
Potassium chloride (KCl)	0.5 (0.05%)
Magnesium sulfate heptahydrate (MgSO ₄ .7H ₂ O)	0.5 (0.05%)
Ferrous sulfate heptahydrate (FeSO ₄ .7H ₂ O)	0.01 (0.001%)
Manganese sulfate heptahydrate (MnSO ₄ .7H ₂ O)	0.01 (0.001%)

Note:

1. *D-glucose and Sodium phytate should be filtered sterilized and added after autoclaving.*
2. *Add 3% agar in PSB to prepare phytase screening agar (PSA).*

- **50X Tris-acetate Buffer:**

Reagent	Weight
2M Tris	121.14g
1M Glacial acetic acid	28.57ml
50mM EDTA	9.306gm

Mix the above chemicals in 450ml water. Adjust pH 8. Make up the volume to 500ml with distilled water.

Appendix

- **0.1M Glycine-HCl buffer (pH 2,3, and 3.6):**

Reagent	pH 2	pH 3	pH 3.6
Glycine (75.07 g/mol)	0.38g	0.38g	0.38g
HCl (36.46 g/mol)	0.37ml	0.09ml	0.045ml

Dissolve Glycine in 45ml of distilled water and adjust pH at 2, 3, and 3.6 by slowly adding HCl (as mentioned in the table). Do not add the mentioned HCl quantity at once. Add gently as per the requirement. Make up the volume to 50ml with distilled water.

- **0.1M Sodium acetate buffer (pH 3.6, 4, 4.5, 5, 5.5, 6):**

Reagent	pH 3.6	pH 4	pH 4.5	pH 5	pH 5.5	pH 6
Sodium acetate (mg)	19.791 (0.004825 M)	93.047 (0.02269M)	184.617 (0.04501M)	276.187 (0.06734M)	367.757 (0.08966M)	386.071 (0.09413M)
Acetic acid (ml)	0.285 (0.09517 M)	0.232 (0.07731M)	0.165 (0.0549M)	0.098 (0.03266M)	0.031 (0.01034M)	0.017 (0.005871M)

Dissolve Sodium acetate in 40ml of distilled water and adjust pH 3.6, 4, 4.5, 5, 5.5, and 6 with glacial acetic acid as mentioned in the table. Do not add the mentioned glacial acetic acid at once. Add as per the requirement. Make up the volume to 50ml with distilled water.

- **0.1M Tris-acetic acid buffer (pH 6 and 7):**

Dissolve 0.788g of Tris-HCl in 40 ml of distilled water and adjust pH 6 and 7 with glacial acetic acid. Make up the volume to 50ml with distilled water.

Appendix

- **0.1M Tris-Base (pH 7, 8, and 9):**

Dissolve 0.605g of Tris-HCl in 40 ml of distilled water and adjust pH 7, 8, and 9 with HCl. Make up the volume to 50ml with distilled water.

- **Phytate stock solution (44mM):**

Dissolve 145mg of phytate in 5 ml of respective pH buffers used in the phytase assay.

- **Pottasium dihydrogen phosphate (KH₂PO₄) (50mM):**

Dissolve 68mg of *dried KH₂PO₄ in 5 ml of distilled water.

*Note: *Keep some KH₂PO₄ powder in a hot air oven at 60°C overnight before using it for assay.*

- **5N H₂SO₄:**

Add 6.86ml of concentrated H₂SO₄ dropwise in 43.14ml of distilled water.

- **5% Ammonium molybdate:**

Add 5g of ammonium molybdate in 45ml of distilled water. Keep it in a water bath at 50°C. Let it dissolve. Make up the volume to 50ml by using distilled water.

- **Lysis buffer (200ml, pH 7):**

Stock Reagent	Working Reagent	Volume required to prepare working reagent
1M Tris base	25mM Tris base	5ml
1M NaCl	300mM NaCl	60ml
Dissolve the above chemicals and adjust pH 7.		
Distilled water	-	135ml
Total volume		200ml

Appendix

- **Kanamycin sulfate (50 mg/ml):**

Stock preparation: Weigh 500mg of Kanamycin sulfate in 8 ml of distilled water. Vortex till it is soluble completely. Make up the volume to 10ml. Filter sterilized solution by using a 0.2 μ m filter. Aliquot and store it in -20°C.

- **1% Agarose gel:**

Weigh 1gm of agarose powder. Add 100ml of 1XTAE buffer. Dissolve it by exposing it to heat by using the oven. Pour it into the gel casting apparatus.

- **1M CaCl₂.2H₂O: for competent cells preparation:**

Weigh 14.7g CaCl₂.2H₂O and dissolve it in sterile distilled water via vortexing. Filter sterilizes solution by using a 0.2 μ m filter. Store at 4°C for competent cell preparation.

- **0.1M CaCl₂.2H₂O: competent cells preparation:**

Add 1 ml of the above filter and sterilize 1M CaCl₂.2H₂O in 9 ml of distilled water. Store at 4°C for competent cell preparation.

- **Freezing Mixture for competent cell storage:**

Reagent	Volume
Sterile 0.1M CaCl ₂ .2H ₂ O	0.6ml
Sterile 15% glycerol	3ml of sterile 30% glycerol stock
Sterile Distilled water	2.4ml
Total volume	6ml

- **1M Isopropyl β -D-1-thiogalactopyranoside (IPTG):**

Weigh 2.38g IPTG powder and dissolve it in 8 ml of sterile distilled water. Make up the final volume of 10ml. Filter sterilize it by using a 0.2 μ m filter. Store at -20°C for up to one year.

Appendix

- **4X SDS-PAGE gel loading dye was prepared according to the following:**

Reagents	Quantity
100% Glycerol	4 ml
1 M Tris (pH 6.8)	2.4ml
SDS	0.8 g
Bromophenol blue	4 mg
* β -mercaptoethanol	0.5 ml
Distilled water	3.1 ml
Total volume	10ml

*Note: Add all the above reagents except β -mercaptoethanol. This is labeled as incomplete 4X SDS-PAGE gel loading dye. To prepare complete 4X SDS-PAGE gel loading dye: aliquot required incomplete 4X SDS-PAGE gel loading dye and *add required β -mercaptoethanol to make complete 4X SDS-PAGE gel loading dye.*

- **Resolving Gel Buffer Solution (1.5M Tris, pH 8.8):**

18.17 g Tris-Base was dissolved in 70 mL of deionized water. The pH was adjusted to 8.8 with HCl and the volume was made up to 100 mL with distilled water.

- **Stacking Gel Buffer Solution (0.5M Tris, pH 6.8):**

6.055 g Tris base was dissolved in 70 mL of DI water. The pH was adjusted to 6.8 with HCl and the volume was made up to 100 mL with distilled water.

- **30% Acrylamide solution:**

Prepare the solution in dark conditions. Add 29.2g of acrylamide powder and 0.8g of Bis-acrylamide powder. Mix it by using a magnetic stirrer. Store at 4°C.

Appendix

- **10% SDS-PAGE resolving gel:**

Composition of resolving gel	Quantity (ml)
Distilled water	4
30% Acrylamide	3.3
1.5 M Tris (pH 8.8)	2.5
10% SDS	0.1
10% Ammonium persulphate	0.1
TEMED	0.004
Total volume	10

- **5% stacking gel:**

Composition of resolving gel	Quantity (ml)
Distilled water	3.4
30% Acrylamide	0.83
1 M Tris (pH 6.8)	0.63
10% SDS	0.05
10% Ammonium persulphate	0.05
TEMED	0.005
Total volume	5

Appendix

- **5X SDS-PAGE Running Buffer (Tris-Glycine SDS buffer, pH – 8.3):**

Reagent	Volume
Tris-Base	15.1g
Glycine	94g
10% SDS	50ml
Distilled water	1000ml

- **Coomassie staining solution:**

Reagent	Volume
0.1% Coomassie dye	0.1g
10% acetic acid	10ml
40% methanol	40ml
Make up a Volume of 50 by adding Distilled water	

- **Destaining solution:**

Reagent	Volume
45% methanol	45ml
10% Glacial acetic acid	10ml
45% water	45ml
Total Volume	100ml

- **1M Tris-Base (100ml, pH7):**

Dissolve 12.1g of Tris-Base in 90ml of distilled water. Adjust pH7 with HCl. Make up the volume to 100ml with distilled water. Autoclave it and store at RT.

Appendix

- **1M NaCl (100ml):**

Dissolve 5.844g of NaCl in 95ml of distilled water. Make up the volume to 100ml with distilled water.

- **1M Imidazole (100ml, pH7):**

Dissolve 6.808g of imidazole in 90ml of distilled water. Adjust pH 7. Make up the volume to 100ml with distilled water. Autoclave it and store at RT.

- **Equilibration buffer (pH7):**

Stock	Working Concentration	Volume to make working concentration (ml)
1M Tris-Base	50mM Tris-Base	0.5
1M NaCl	300mM NaCl	3
1M Imidazole	10mM Imidazole	0.1
Distilled water		6.4
Total volume		10

- **Wash buffer (pH7):**

Components	Concentration	Volume to make working concentration (ml)
1M Tris-Base	50mM Tris-Base	0.5
1M NaCl	300mM NaCl	3
1M Imidazole	40mM Imidazole	0.4
30% Glycerol	5% Glycerol	1.7
Distilled water		4.4
Total volume		10

Appendix

- **Elution buffer (pH7):**

Components	Concentration	Volume to make working concentration (ml)
1M Tris-Base	50mM Tris-Base	0.5
1M NaCl	300mM NaCl	3
1M Imidazole	450mM Imidazole	4.5
30% Glycerol	5% Glycerol	1.7
Distilled water		0.3
Total volume		10

- **Dialysis buffer (pH7):**

Components	Concentration	Volume to make working concentration (ml)
1M Tris Base	20mM Tris Base	20
1M NaCl	150mM NaCl	150
30% Glycerol	3% Glycerol	100
Sterile water		730
Total volume		1000

List of Grant, Conferences, Workshops, Awards

List of Grant, Conferences, Workshops, Awards

Grant

- NMIMS seed grant: 2018 (Grant no.: 401420)
- NMIMS seed grant: 2020 (Grant no.: 402260)

Conference(s) and poster presentation:

- International conference: Advances in Materials Science & Applied Biology (AMSAB) held on 8th – 10th January 2019 at NMIMS university. Title of the Poster presented in AMSAB: Computational analysis of putative phytase enzymes.
- International conference: International Conference on Advances in Biotechnology: Current Discoveries and Future Perspectives (ICAB-2023) held on 17-18th October, 2023 at Amity University. Title of the poster presented in ICAB: Exploring novel enzymes: *in-silico* bioprospecting and experimental validation.

Workshop:

- ‘Bioinformatics Workshop: RNAseq data analysis’, Indian Women Scientist Association (IWSA) (1st-4th December 2017).
- ‘Bioinformatics Workshop: Molecular Dynamics’, IWSA (23rd -25th December 2018).
- ‘Molecular Docking in Rational drug design’, Ramnarain Ruia College, Mumbai (Feb 4-6th, 2021).
- ‘Next Generation Sequencing, Proteomics and Bioinformatics’, Guru Nanak Khalsa College of Arts Science & Commerce (March 2020).
- ‘Docking of Ritonavir with SARS-CoV-2 main protease to understand their binding affinity’, Department of Bioinformatics, GN Khalsa college.
- ‘International workshop in the area of “Bioinformatics, Genomics, Text Mining and NGS data analysis’, Nextgenhelper (September 12 - 30, 2022)

Grant/Conferences/Workshops/Awards

- ‘Basics to advanced Multiomics data analysis’, Nextgenhelper (May 15th – June 15th, 2023).

Award:

- ‘Late Dr. Suresh Mahajan - IWSA Scholarship’ for the year 2022-2023.

List of Publications

List of Publications

Review/Research paper

1. **Kamble, A.**, Srinivasan, S. and Singh, H. (2019) ‘In-Silico Bioprospecting: Finding Better Enzymes’, *Molecular Biotechnology*, 61(1), pp. 53–59. DOI: [10.1007/s12033-018-0132-1](https://doi.org/10.1007/s12033-018-0132-1). (Impact factor: 2.860)
2. **Kamble, A.**, Sawant, S. and Singh, H. (2020) ‘16S Ribosomal RNA Gene - Based Metagenomics : A Review’, *BMRJ*, pp. 5–11. DOI: 10.4103/BMRJ.BMRJ_4_20.
3. **Kamble A** and Singh H (2020) ‘Different Methods of Soil DNA Extraction’, *Bio-protocol*, e3521, pp. 1–23. (*Web of Science*: Emerging Sources Citation Index (ESCI) indexes). DOI: [10.21769/BioProtoc.3521](https://doi.org/10.21769/BioProtoc.3521).
4. Nagar, A., **Kamble, A.** and Singh, H. (2021) ‘Preliminary screening, isolation and identification of microbial phytase producers from soil’, *Environmental and Experimental Biology*, 19, pp. 11–22. (Additional *Web of Science* Indexes). DOI: <https://doi.org/10.22364/eeb.19.03>.

Book chapter

- **Kamble, A.D.** and Singh, H. Ch26 - Finding novel enzymes by in silico bioprospecting approach. Ed(s): M Kuddus, CN Aguilar. In: *Value-Addition in Food Products and Processing Through Enzyme Technology*, Academic Press, 2022, Pages 347-364, ISBN 9780323899291. DOI: [10.1007/s12033-018-0132-1](https://doi.org/10.1007/s12033-018-0132-1)



In-Silico Bioprospecting: Finding Better Enzymes

Asmita Kamble¹ · Sumana Srinivasan² · Harinder Singh¹

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Enzymes are essential biological macromolecules, which catalyse chemical reactions and have impacted the human civilization tremendously. The importance of enzymes as biocatalyst was realized more than a century ago by eminent scientists like Kuhne, Buchner, Payen, Sumner, and the last three decades has seen exponential growth in enzyme industry, mainly due to the revolution in tools and techniques in molecular biology, biochemistry and production. This has resulted in high demand of enzymes in various applications like food, agriculture, chemicals, pharmaceuticals, cosmetics, environment and research sector. The cut-throat competition also pushes the enzyme industry to constantly discover newer and better enzymes regularly. The conventional methods to discover enzymes are generally costly, time consuming and have low success rate. Exploring the exponentially growing biological databases with the help of various computational tools can increase the discovering process, with less resource consumption and higher success rate. Present review discusses this approach, known as in-silico bioprospecting, which broadly involves computational searching of gene/protein databases to find novel enzymes.

Keywords In-silico · Bioprospecting · Enzyme

Introduction

Enzymes play an important role in our daily lives and are used in variety of industries and sectors like food, detergent and medicine [1]. The demand of certain enzymes has increased exponentially, like lipases, proteases, hydrolases and polymerases. Research laboratories and industries are extensively working to find newer and better candidates. Major enzyme industries are regularly introducing new enzymes in the market. In the past two decades, several patents on enzymes have been filed and issued. Apart from this, there are ongoing efforts to substitute chemical reaction processes in industries with enzymatic processes, as they are greener and environment friendly alternatives. It has been widely accepted that a cleaner chemical synthesis process should be practiced to prevent pollution and avoid generation of toxic wastes [2]. Enzymatic synthesis of chemical

compounds has emerged as a simple, better and competitive route in comparison to chemical methods. Also, a high substrate specificity and better conversion rate with formation of low or no by-products makes enzyme a robust and efficient choice. Recently, Merck and Codexis developed a greener process for the synthesis of Sitagliptin, a drug used in diabetes treatment [3]. In the recent years, advancement in recombinant DNA technology has resulted in successful approaches to overexpress an enzyme in variety of host cells, which can help in producing the biocatalyst in high amount. To obtain an efficient enzyme candidate, stringent selection criteria are required to achieve high activity, specificity, and stability. In an industrial processes, the substrate, solvent, reaction conditions are important and an enzyme chosen should be able to withstand these components and conditions. It is actually difficult to find a natural enzyme with all the properties desired in an industrial process. To fulfil the massive enzyme demand, various approaches are practiced to constantly explore different resources to obtain new and better enzymes. Among these, in-silico bioprospecting has come up as an efficient, cost and time effective approach to discover new enzyme candidates. Although this approach has been practiced at various laboratories [4–6], it has not been reviewed or discussed.

✉ Harinder Singh
Harinder.Singh@nmims.edu

¹ Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS Deemed to be University, Vile Parle (W), Mumbai 400056, India

² Biosystems Engineering Lab, Department of Chemical Engineering, IIT Bombay, Powai, Mumbai 400076, India

In-Silico Bioprospecting

New enzyme discovery can be accomplished using various conventional and contemporary methods as mentioned in Fig. 1. Common methods of screening to identify novel enzymes are performed by exploring natural sources like industrial waste or soil, but they require an established protocol for screening assay or selection method based on the desired properties of the enzyme. This process involves biochemical screening and isolating the organism on selective media, which is usually time and resource consuming and may or may not result in a novel candidate. From these screening assays, the selected organism further needs to be identified, followed by the identification of gene sequence which is coding for the desired enzyme and function. One approach is to perform random mutagenesis to create enzyme mutant, and then sequence the DNA region. Another way is to perform targeted or whole genome sequencing to identify the desired enzyme gene sequence. As an alternative, amplification of target gene can be performed using degenerate primers [7]. There are challenges involved in primer designing, which affects the success rate. The process is followed by PCR library cloning and screening for prospective candidates with desired properties, which again demands a well-established protocol for screening positive candidates. After selecting the desired clone, the responsible gene can be sequenced, cloned and expressed.

The direct screening and identification methods are preferred where molecular biology resources are inadequate. These experimental approaches are used commonly, but they are time and resource consuming, with low success rate. However, in-silico bioprospecting is a simple, straightforward and promising approach to identify novel enzyme candidates with better enzymatic properties. A compilation of recent reports, where in-silico bioprospecting approach has been used to find novel enzymes, is given in Table 1. The current fast paced, high-throughput whole genome/metagenome sequencing has tremendously increased the biological database and thus the enzyme diversity. This diversity in turn has increased the complexity and difficulty of finding a novel candidate. The in-silico bioprospecting process can be broadly divided into two steps: (i) Searching databases (ii) Using Bioinformatics tools to screen, analyse and shortlist prospective candidates.

Step 1: Searching Databases

This can be performed by exploring databases using various search tools based on homology, conserved motif, consensus guided approach, or simply keyword search. The search result can be further screened using filters, such as percentage identity, query coverage, *e*-value. For example, a keyword search in NCBI protein database can be performed, followed by filtering the results to show candidates between 30 and 80% identity with query coverage > 95%. Gupta et al. [11] used keywords such as 'Hypothetical Protein of

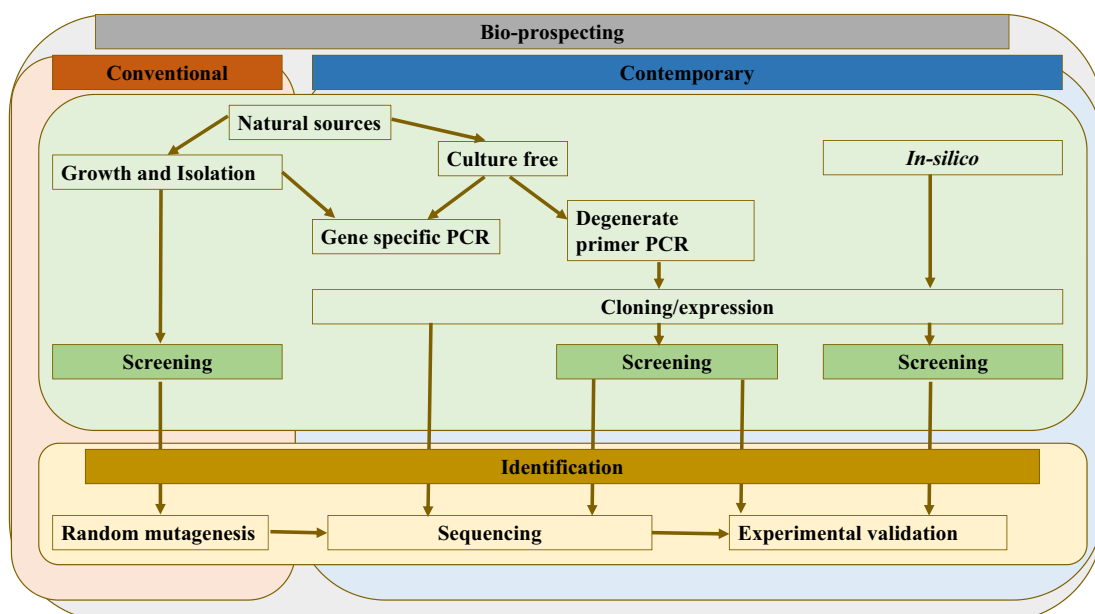


Fig. 1 Methods of enzyme bioprospecting

Table 1 In-silico bioprospecting approach used to find novel enzymes

Sr. no.	Enzyme	Approach/method	Reference
1	Histidine acid phytase (HAP)	Pfam identifier for the conserved domain Pfam00328 was used screen novel HAP metagenomes of the acidic peatland soil in Schlöppnerbrunnen, in the Fichtelgebirge Mountains, Germany Sequences were further manually analysed to examine the presence of the conserved motifs (RHGXRRP and HD) which were present in HAP phytase Further Phylogenetic analysis and taxonomic affiliations given by using MAFFT online service, nBLAST, PhyloPythiaS online tool Followed by Comparative homology modelling by using SWISS-MODEL Experimental validation	[8]
2	Aldehyde-deformylating oxygenase (ADO)	<i>Synechococcus elongatus</i> PCC7942 used as the query sequence to screen for the presence of <i>ado</i> gene in the set of sequences of organisms belonging to hot spring IMG/MER database Structural and Functional analysis performed by using various Bioinformatics tools such as ExPASy translate tool for translating sequence, MEGA software for phylogenetic analysis Adopted structure based protein engineering approach to improve thermostability Experimental validation	[9]
3	Nitrilases	Homology and motif based approach adopted for Genome screening Insilico analysis of putative sequences Experimental validation	[10]
4	Baeyer–Villiger and CYP153Monooxygenases	Metagenomic dataset created by isolating DNA sample from 23 sediment samples from (i) Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway [NOR]; (ii) Port Värtahamnen, Stockholm, Baltic Sea, Sweden [SWE]; (iii) Ushuaia Bay, Tierra del Fuego Island, Argentina [ARG]; and (iv) Potter Cove, 25 de Mayo (King George) Island, Antarctica [ANT] was further submitted to IMG database Pfam domains (PF00067 for Cytochrome P450 or PF00743 for Flavine-binding monooxygenase-like) was used to screen this metagenomic dataset Further Blastp analysis performed using well characterized enzyme and/or crystallized sequences as reference Insilico characterization of putative sequences	[5]
5	Hypothetical protein	Hypothetical Protein was searched using NCBI database using keywords ‘Hypothetical Protein of <i>Triticum aestivum</i> ’ or ‘Hypothetical Proteins of wheat’ Functional annotation of identified novel abiotic stress proteins of <i>Triticum aestivum</i>	[11]
6	Cellulose-and/or hemicellulose-degrading enzymes	Functional metagenomics Porcupine Microbiome metagenomic data created Relevant protein sequences of interest were selected on the basis of domain conservation, low <i>e</i> -values- pHMMER & the Research Collaboratory for Structural Bioinformatics Protein Data Bank Experimental validation of putative candidates	[12]
7	β -Glucosidase	Metagenomic DNA library constructed Search in the existing databases Experimental validation	[13]
8	b-(1,3) Galactosyltransferases	20 putative b-(1,3)-GalT genes identified by performing tblastn using three Homo sapiens b-(1,3)-GalT sequences: GalT1 (Q9Y5Z6), GalT2 (O43825), GalT4 (O96024) against <i>Arabidopsis thaliana</i> sequence database at NCBI Insilico characterization and experimental validation performed	[14]
9	Polyhydroxyalkanoate (PHA) synthase	Complete or nearly full-length PHA synthase genes retrieved from MG-RAST database mangrove soil metagenomic data Experimental Validation of novel and wide substrate specific PHA	[15]

T. aestivum', 'Hypothetical Proteins of wheat' in NCBI database followed by manual screening to get unique protein candidates. After removing redundant entries, unique candidates were further subjected to physicochemical, localization, function and domain analysis. In another database search, keywords such hydroxybutyrate, hydroxyalkanoate, hydroxyalkanoic, PHA and PHB were used as input [15]. Another common approach practiced by researcher is to search biological databases using a known candidate enzyme sequence. While choosing a potential enzyme gene sequence, it is of utmost importance to select a full length protein sequence having conserved domains, as many incomplete sequences annotated in database do not code for a functional protein, when checked experimentally. Also, in the search result, the selected candidate's sequence similarity should not be very high with known sequence. This is to ensure that a novel candidate is shortlisted and not a close homologue of a known sequence. In the similarity search result, the hits with > 90 identity are very closely related, sources like different species of same family, and it is more likely that they are very similar. But, the hits with ~80% identity or lower are those candidates who are different from the query candidates, not closely related, but do have conserved sequences similar to known candidates. This ensures that novel candidates are chosen, which is predicted to retain the enzyme activity but is different from the search query. There have been reports where researchers had selected candidates with sequence similarity as low as 40 percent. Sharma et al. [10] searched novel sources of nitrilases from microbial genomes by adopting homology-based approach and selected sequences which exhibited > 30% and < 80% identity. The shortlisted search results need to be confirmed for a complete coding sequence or sequences. For example, shortlisted candidates of nitrilase were checked by GenMark S tool to verify complete coding sequences or sequences [10]. Since the protein length information is available for the input sequence, the search results should be restricted to length closer to the input sequence length. In case of nitrilases, sequences with less than 100 amino acids were considered as false positive and were discarded [10]. In another instance, sequences less than 250 amino acids were excluded to find novel BVMO (Bayer-Villiger Monooxygenases) enzyme [5]. For PHA synthase, sequences with ~120 to 260 bp were considered as prospective candidate in a database search [15]. These search filters along with others like *e*-value, can aid in gathering positive sequences which could code for functional enzyme of appropriate length and reduces the chance of false discovery or random or irrelevant search result.

In certain cases, designing motif from selected protein sequences [e.g. by using MAST (Motif Alignment and Search Tool) at MEME suite] can be used to search bacterial genome. For example, Homology-based approach and motif

search resulted in the identification of 138 putative/hypothetical protein sequences which had potential to code for nitrilase [10]. Vaquero et al. [16] also adopted homology-based strategy to screen for novel CalB-type lipase in fungal genomes using blastp algorithm, against JGI and NCBI databases, with *e*-value cut-off as 10^{-2} . In the same study, conserved motif approach failed to identify putative lipase gene due to absence of conserved sequence motif generated by MEME software. Therefore, different individual strategies or combinations should be implemented in the process of finding novel putative enzymes. Consensus-guided approach, using Pfam domain, can also be used to search databases for the presence of particular enzyme family. Consensus-guided approach was adopted by Shakeel et al. [9] to obtain heat stable alkane-producing enzymes, using *ado* gene from *Synechococcus elongatus* PCC7942 as a query to search IMG/MER hot spring database. A consensus sequence was generated from the list of homologous sequences using Bioinformatics tools, which was further validated computationally and experimentally.

Specific datasets like metagenomes from various ecosystems can also be searched for obtaining novel enzymes. Around 264 putative monooxygenases were obtained when Pfam domain and blastp search were used to search BVMO [5] from ~14 million protein-coding sequences present in metagenomic dataset of cold marine sediments [5]. Metagenome data of mangrove soil were explored to find polyhydroxyalkanoate (PHA) synthase genes [15]. Adam et al. [17] reported a novel activity-based approach to screen H₂-uptake enzyme from hydrothermal Metagenome. Toyama et al. [13] reported a novel β-glucosidase from microbial Metagenome of a lake in Amazon. Tan et al. [6] reported a novel thermostable phytase using bioinformatics approach which was screened from Metagenome database. Various steps and approaches used in gene mining from Metagenome data have been discussed and reviewed recently and reader is referred to these articles and reviews [18, 19] for details.

The steps of in-silico bioprospecting can be modified as per the desired property of enzyme. For example, if a thermostable enzyme is desirable, but the known enzyme reported is not thermostable, the similarity searches in thermophiles will be useful to find putative thermostable enzymes. It has been commonly observed that the thermostable enzyme sequences are different from their mesophilic counterpart. The putative thermophilic candidates searched this way should be further analysed (discussed in Step 2) to make sure that residues important for structure and functions are conserved.

Step 2: Using Bioinformatics Tools to Screen, Analyse and Shortlist Prospective Candidates

Once the primary list has been generated using various database search approaches, the next step will be to

analyse their physiochemical, phylogenetic and functional properties using different bioinformatics tools. ProtParam software using ExPASy server is widely used to access physiochemical properties (such as the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) of putative candidates [10, 11, 20]). Predicted values of all parameters of putative enzyme(s) are compared to the well characterized enzyme which affects the confidence level to study the putative enzyme(s) experimentally. For example, ProtParam predicted physiochemical properties of 138 putative nitrilases with in the range of well-characterized nitrilases [10]. All the parameters are based on protein sequence *i.e.* sequence-dependent analysis; therefore, it is necessary to get complete or nearly complete sequence for accurate analysis and prediction of various physiochemical properties.

Phylogenetic analysis can be performed using tools like Molecular Evolutionary Genetics Analysis (MEGA) [11, 15, 16, 21]. For example, phylogenetic analysis of selected putative candidates belonging to CalB-family grouped putative lipases in to different clusters of known lipases depending upon its evolutionary closeness [16], thus helping in deciding on novel and unique candidates. Structural modelling of putative candidates can be performed using SWISS-MODEL server or MODELLER v9.15 software [21]. Vaquoro et al. [16] used CalB as template to model PlicB, which exhibits 30% sequence identity and 44% similarity. The information about structure and residue conservation prediction is only possible if structural data of protein homologues are available through crystal structures. Hence, persistent exploration and enrichment of databases are necessary for in-silico bioprospecting of novel enzymes.

There are other tools which can predict structural information such as signal peptide (e.g. Signal P) or disulphide linkages (e.g. DiANNA). DiANNA 1.1 web server predicted two disulphide bonds in PlicB whereas CalB and Uml2 lacks disulfide bonds [16]. Protein functional domains and families are studied by comparing list of putative enzyme(s) against databases like Pfam, CATH, SVM-Prot, CDART, SMART. In one study, hypothetical proteins (HPs) were explored using tools based on domain architecture and profiles [11]. Out of 124 HPs, 77 sequences were annotated with high confidence by using Pfam, CATH, SVM-Prot, CDART, SMART and ProtoNet, and among them, 16 were predicted as enzymes. Functional protein network provides information about the association of hypothetical/putative protein(s) with the known functional protein, which can be generated by STRING database. In the study conducted by Gupta et al. [11], it was found that the predicted HPs such as HAV22 (Q7XAP6) and F-box protein (D0QEJ9) were interacting

with other proteins of the STRING database such as protein 4,345,793 of *Oryza sativa* subsp. *Japonica*.

Analysing the putative candidates using bioinformatic tools provides clarity and help in selecting those candidates which are structurally and functionally more suitable, novel and unique. Following the sequence selection, candidates are validated for desired properties by cloning and expressing them in artificial expression systems followed by physiochemical characterization of enzyme [6, 13]. Apart from in-silico bioprospecting, enzymes with desired properties such as high activity [22, 23], substrate specificity [24] and stability [25, 26] can also be obtained by modifying the existing enzyme using mutagenesis via directed evolution, rational or semi-rational approaches [27–35]. Random mutagenesis of a single gene can be done by chemical, error prone-PCR or saturation mutagenesis, or by using mutator strains. On the other hand, gene recombination approach can be applied with more than one related gene sequences, using tools like DNA shuffling, Random Chimeragenesis on Transient Templates (RACHITT), Exon shuffling, incremental truncation for the creation of hybrid enzymes (ITCHY), Sequence Homology-Independent Protein Recombination (SHIPREC). The reader is referred to review by Rubin-pital et al. [31] for details about these processes, their advantages and drawbacks. Recent developments along with additions of rational component have resulted in faster selection methods and maximized qualities of libraries with more relevant mutations [36]. Rational mutagenesis to improve enzyme property has been attempted in recent years to obtain the desired property; however, the phenotype of certain mutations is still beyond the current understanding of enzyme structure and function.

Conclusion

In the past few years, enzyme production and research have taken a major leap and a vast number of potential enzymes are available in market and are produced at industrial scale. Reports are being continuously published related to the screening and finding newer and better enzymes. However, it is generally observed that wild-type enzymes are not directly applicable for an industrial process. In the coming years, it is expected that more industrially important enzymes will be discovered or engineered that can satisfy the ever-growing demand of enzymes. The availability of various expression vectors, host and systems has increased the possibility of expressing a gene artificially in a host of our choice. However, protein expression even in bacterial host like *E. coli* can be challenging many times [37, 38]. The diversity of enzymes present in databases indicates that the present knowledge of structure and function is vast but far from complete. The last two decades has seen

tremendous growth in protein structural information, and expression systems and tools have enriched in large, but we still require more information to understand and utilize it to its full potential. With the rise in molecular techniques, enzyme improvement by protein engineering has taken a big leap [35]. Drastic improvement in enzymatic properties like activity and stability has been witnessed by using methods of directed evolution or rational mutagenesis. With the current knowledge of enzyme structure and function, it is still a challenging task to pursue a rational approach of enzyme engineering in every case to improve their properties. Efforts should be more focussed towards solving enzyme crystal structures and expanding our knowledge and understanding of enzyme function and properties. The pace of structure information cannot be compared with the way new genes or proteins are being discovered, but attempts can be made to improve it further. Generating and analysing diverse crystallographic data will help in understanding the enzymes in greater details, and also, will help in rational engineering of the enzyme for improved properties. There is an urgent demand for developing new tools and pipelines which can handle and analyse the exponentially growing database, and related experimental literature, with minimal manual intervention. This will help in discovering novel and better enzymes comparatively faster with high success rate.

Funding The manuscript is a review article and was not supported by any funding agency.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent For this type of study, formal consent is not required.

References

- Coker, J. A. (2016). Extremophiles and biotechnology: Current uses and prospects. *F1000Research*, 5, 396.
- Anastas, P. T., & Warner, J. C. (1998). *Green chemistry: Theory and practice*. Oxford University Press: New York.
- Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., ... Hughes, G. J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, 329(5989), 305–309.
- Chamoli, M., & Pant, K. (2014). In-silico bioprospecting of the enzymes involved in the biosynthetic pathway of the alkaloid berberrine and its distance study Through R. *International Journal of Advanced Technology in Engineering and Science*, 2(9), 165–178.
- Musumeci, M. A., Lozada, M., Rial, D. V., Cormack, W. P. M., Jansson, J. K., Sjöling, S., ... Dionisi, H. M. (2017). Prospecting biotechnologically-relevant monooxygenases from cold sediment metagenomes: An in silico approach. *Marine Drugs*, 15(4).
- Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2016). Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Applied Microbiology and Biotechnology*, 100(5), 2225–2241.
- Berón, C. M., Curatti, L., & Salerno, G. L. (2005). New strategy for identification of novel cry-type genes from bacillus thuringiensis strains. *Applied and Environmental Microbiology*, 71(2), 761–765.
- Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2016). A novel phytase derived from an acidic peat-soil microbiome showing high stability under acidic plus pepsin conditions. *Journal of Molecular Microbiology and Biotechnology*, 26(4), 291–301.
- Shakeel, T., Gupta, M., Fatma, Z., Kumar, R., Kumar, R., Singh, R., ... Yazdani, S. S. (2018). A consensus-guided approach yields a heat-stable alkane-producing enzyme and identifies residues promoting thermostability. *The Journal of Biological Chemistry*, 1–30.
- Sharma, N., Thakur, N., Raj, T., Savitri, & Bhalla, T. C. (2017). Mining of microbial genomes for the novel sources of nitrilases. *BioMed Research International*, 2017.
- Gupta, S., Singh, Y., Kumar, H., Raj, U., Rao, A. R., & Varadwaj, P. K. (2018). Identification of novel abiotic stress proteins in triticum aestivum through functional annotation of hypothetical proteins. *Interdisciplinary Sciences: Computational Life Sciences*, 10(1), 205–220.
- Thornbury, M., Sicheri, J., Guinard, C., Mahoney, D., Routledge, F., Curry, M., ... Getz, L. (2018). Discovery and Characterization of Novel Lignocellulose-Degrading Enzymes from the Porcupine Microbiome. *bioRxiv*, (February).
- Toyama, D., de Moraes, M. A. B., Ramos, F. C., Zanphorlin, L. M., Tonoli, C. C. C., Balula, A. F., et al. (2018). A novel β -glucosidase isolated from the microbial metagenome of Lake Poraquê (Amazon, Brazil). *Biochimica et Biophysica Acta*, 1866(4), 569–579.
- Qu, Y., Egelund, J., Gilson, P. R., Houghton, F., Gleeson, P. A., Schultz, C. J., & Bacic, A. (2008). Identification of a novel group of putative Arabidopsis thaliana β -(1,3)-galactosyltransferases. *Plant Molecular Biology*, 68(1–2), 43–59.
- Foong, C. P., Lakshmanan, M., Abe, H., Taylor, T. D., Foong, S. Y., & Sudesh, K. (2018). A novel and wide substrate specific polyhydroxyalkanoate (PHA) synthase from unculturable bacteria found in mangrove soil. *Journal of Polymer Research*, 25(1), 23.
- Vaquero, M. E., De Eugenio, L. I., Martínez, M. J., & Barriuso, J. (2015). A novel CalB-type lipase discovered by fungal genomes mining. *PLoS ONE*, 10(4), 1–11.
- Adam, N., & Perner, M. (2018). Novel hydrogenases from deep-sea hydrothermal vent metagenomes identified by a recently developed activity-based screen. *ISME Journal*, 12(5), 1225–1236.
- Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., & Golyshin, P. N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. *Microbial Biotechnology*, 9(1), 22–34.
- Uria, A. R., & Zilda, D. S. (2016). Metagenomics-guided mining of commercially useful biocatalysts from marine microorganisms. In *Advances in Food and nutrition research*.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the Expasy server. *The Proteomics Protocols Handbook*, 571–607.
- Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., & Watson, M. (2017). A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Frontiers in Genetics*.

22. Machielsen, R., Leferink, N. G. H., Hendriks, A., Brouns, S. J. J., Hennemann, H. G., Daußmann, T., & Van Der Oost, J. (2008). Laboratory evolution of *Pyrococcus furiosus* alcohol dehydrogenase to improve the production of (2S,5S)-hexanediol at moderate temperatures. *Extremophiles*, *12*(4), 587–594.
23. Wang, N. Q., Sun, J., Huang, J., & Wang, P. (2014). Cloning, expression, and directed evolution of carbonyl reductase from *Leifsonia xyli* HS0904 with enhanced catalytic efficiency. *Applied Microbiology and Biotechnology*, *98*(20), 8591–8601.
24. Jakoblinert, A., Wachtmeister, J., Schukur, L., Shivange, A. V., Bocola, M., Ansoerge-Schumacher, M. B., & Schwaneberg, U. (2013). Reengineered carbonyl reductase for reducing methyl-substituted cyclohexanones. *Protein Engineering, Design and Selection*, *26*(4), 291–298.
25. Hoelsch, K., Sührer, I., Heusel, M., & Weuster-Botz, D. (2013). Engineering of formate dehydrogenase: Synergistic effect of mutations affecting cofactor specificity and chemical stability. *Applied Microbiology and Biotechnology*, *97*(6), 2473–2481.
26. Koudelakova, T., Chaloupkova, R., Brezovsky, J., Prokop, Z., Sebestova, E., Hesseler, M., ... Damborsky, J. (2013). Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angewandte Chemie - International Edition*, *52*(7), 1959–1963.
27. Buller, A. R., Brinkmann-Chen, S., Romney, D. K., Herger, M., Murciano-Calles, J., & Arnold, F. H. (2015). Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proceedings of the National Academy of Sciences*, *112*(47), 14599–14604.
28. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., ... Arnold, F. H. (2013). General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences*, *110*(27), 10946–10951.
29. Fox, R. J., & Huisman, G. W. (2008). Enzyme optimization: moving from blind evolution to statistical exploration of sequence-function space. *Trends in Biotechnology*, *26*(3), 132–138.
30. Reetz, M. T., Rentsch, M., Pletsch, A., Maywald, M., Maiwald, P., Peyralans, J. J. P., ... Taglieber, A. (2007). Directed evolution of enantioselective hybrid catalysts: a novel concept in asymmetric catalysis. *Tetrahedron*, *63*(28), 6404–6414.
31. Rubin-Pitel, S. B., Cho, C. M. H., Chen, W., & Zhao, H. (2007). Directed evolution tools in bioproduct and bioprocess development. *Bioprocessing for Value-Added Products from Renewable Resources*, 49–72.
32. Li, Y., & Cirino, P. C. (2014). Recent advances in engineering proteins for biocatalysis. *Biotechnology and Bioengineering*, *111*(7), 1273–1287.
33. Wang, M., Si, T., & Zhao, H. (2012). Biocatalyst development by directed evolution. *Bioresource Technology*, *40*(6), 1301–1315.
34. Woodley, J. M. (2013). Protein engineering of enzymes for process applications. *Current Opinion in Chemical Biology*, *17*(2), 310–316.
35. Zheng, G. W., & Xu, J. H. (2011). New opportunities for biocatalysis: Driving the synthesis of chiral chemicals. *Current Opinion in Biotechnology*, *22*(6), 784–792.
36. Lane, M. D., & Seelig, B. (2014). Advances in the directed evolution of proteins. *Current Opinion in Chemical Biology*, *22*, 129–136.
37. Kaur, J., Kumar, A., & Kaur, J. (2018). Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *International Journal of Biological Macromolecules*, *106*, 803–822.
38. Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*.

Preliminary screening, isolation and identification of microbial phytase producers from soil

Akshita Nagar, Asmita Kamble, Harinder Singh*

Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS Deemed to be University, Mumbai 400056, India

*Corresponding author, E-mail: Harinder.Singh@nmims.edu



ISSN 2255-9582



UNIVERSITY OF LATVIA

Abstract

Phytate is a widely found form of phosphate in plant seeds that can be hydrolysed by phytase enzyme to release phosphate and myo-inositol intermediates. Phytase has been mainly used in feed and aquaculture to reduce the anti-nutritional effect and eutrophication caused by phytate. The present study focused on preliminary screening, isolation and characterization of phytase producers from soil. A total of eleven soil samples were collected and screened by plate assay on phytase screening media followed by isolation of phytase producers to obtain pure cultures. Biochemical assays, 16s rDNA restriction fragment length polymorphism and 16s rDNA sequencing were performed for identification of phytase producers. The intracellular and extracellular phytase activity was measured by a modified Heinonen method at pH 2.5, 5.0 and 7.5. The activity of bacterial phytase producers (PSD5TSB, CD5, VD5 FD5O, FD2 and FD5T) was compared with the wild type *Escherichia coli* phytase. The optimum pH for enzyme activity from the soil isolates was observed at pH 5. From soil samples from eleven different sources, presence of bacterial phytase producers was confirmed in four soil samples by this preliminary screening method. The simple plate screening and activity assay helped in isolating phytase producers from soil, which can be used as potential candidates for phytase production.

Key words: phytase, phytase activity, phytase producers, soil.

Abbreviations: LB, Luria-Bertani; PSA, phytase screening agar; PSB, phytase screening broth; RFLP, restriction fragment length polymorphism; TS, tryptic soy.

Introduction

Phytic acid is a type of organic phosphorous with the chemical name myo-inositol hexakisdi-hydrogen phosphate (Dahiya 2016). It is found in seeds of crops such as wheat, barley, soybean, rice, maize, groundnut, legumes, and nuts. It is primary source of inositol and the storage form of phosphorous as well as minerals in plant seeds (Mittal et al. 2011; Kim et al. 2015; Baruah et al. 2017). Phosphorous is an essential macronutrient, utilized in plant metabolism processes such as photosynthesis, respiration and cell division (Karpagam, Nagalakshmi 2014; Motamedi 2016).

Cereals and oilseeds are used as feed for monogastric animals like chicken, pigs and fish, in which phytate has an anti-nutritional effect (Baruah et al. 2017). This is due to the lack of phytase producing intestinal microbiota, which leads to excretion of non-utilized phosphorous, causing environmental pollution (Bohn et al. 2008). On the other hand, phytate has the ability to chelate mineral cations that reduces their solubility (Baruah et al. 2017). Phytate can also form complexes with amino acids and proteins, lipids, starch and vitamins, resulting in reduced absorption of these nutrients (Costenaro-Ferreira, Della Flora 2017;

Savita et al. 2017). However, moderate consumption of phytate can be beneficial. For instance, it has been shown to have antioxidant and anticancer effect, prevent renal lithiasis, lower the glycaemic index and balance glucose and cholesterol levels (Grases, Costa-Bauza 2019).

Myo-inositol hexakisphosphate hydrolase, also known as phytase, is a phosphatase enzyme found in animals, plants and microorganisms (Savita et al. 2017). Phytase cleaves phytic acid or its salt derivatives by hydrolysis to produce free inorganic phosphorous and inositol intermediates with lesser phosphate (Dvořáková 1998). Phytase first hydrolyses all of the hexaphosphate to penta-esters and then moves forward for further dephosphorylation to tetra-esters and so on (Dersjant-Li et al. 2015). The action of the enzyme decreases the affinity of substrate to cations and provides a free source of phosphorous (Bohn et al. 2008).

There are broadly two classes of phytase, first depending on the position of dephosphorylation on the inositol ring and second based on the catalytic mechanism (Dvořáková 1998; Mullaney, Ullah 2003; Greiner et al. 2007; Nasrabadi et al. 2018). The first class includes 3-phytase (EC 3.1.3.8), a class produced by microorganisms; 4/6-phytase (EC 3.1.3.26), mainly produced by plants; and 5-phytase (EC

3.1.3.72), found in several legumes such as *Pisum sativum*, *Phaseolus vulgaris* and *Medicago sativa* (Greiner, Carlsson 2006; Bhavsar, Khire 2014). In the second class, there are acidic (EC 3.1.3.2) and alkaline (EC 3.1.3.8) phytase enzymes. Phy A-3-phytase, Phy B-3-phytase, Phy C-6-phytase belong to a subclass of histidine acid phosphatase. Histidine acid phosphatase, cysteine acid phosphatase and purple acid phosphatase are types of acidic phytase (Bhavsar, Khire 2014). Alkaline phytase includes β -propeller phytase, which has a subtype, Phy D-3 phytase, produced by *Bacillus* sp. (Mullaney, Ullah 2003; Bhavsar, Khire 2014).

Microbial phytases have been studied due to their diverse action, economic advantage at different scale, high activity and high production turn-around time (Ushasree et al. 2017). Phytase is known to promote mineral absorption and increase their bioavailability, reduce non-utilized phytate, decrease mineral deficiency and improve bone health in animals (Dahiya 2016). Until now, phytase has been used commercially as poultry, swine and fish feed, thus saving the use of irreplaceable and expensive inorganic phosphate as well as preventing fungal blooms that leads to eutrophication (Lei, Porres 2003; Bajaj, Wani 2015; Nasrabadi et al. 2018). Microbial phytase is used as an innovative approach in the livestock industry as feed, and in the farming industry as fertilizer, as well as for environmental protection (Akhmetova et al. 2013; Suleimanova et al. 2015). Other applications where phytase is of great interest are the food industry, healthcare and medicine, and aquaculture (Kumar et al. 2010; Shobirin et al. 2010; Caipang et al. 2011; Rocky-Salimi et al. 2016).

An ideal phytase should have high specific activity for phytate substrate, be functionally stable at a wide range of pH and temperature, be resistant towards proteases, be thermostable, and retain stability during harsh processing conditions, during storage of feed as well as in the gut (Lei, Porres 2003; Savita et al. 2017). Methods to improve thermostability of phytase were described (Coutinho et al. 2020). These included identifying novel resistant microbial sources, engineering recombinant phytase from genetically modified microorganisms, substituting amino acids in the enzyme molecule and immobilizing the unstable and thermosensitive phytase to insoluble supports.

Several phytase-producing bacteria have been discovered and studied, such as *Bacillus* sp., *Pseudomonas* sp. and *Raoultella* sp., *Escherichia coli* (Konietzny, Greiner 2004). *E. coli* AppA phytase has optimum activity at acidic pH and shows specific activity towards phytate, deeming it to be a good source of phytase for industrial application (Golovan et al. 1999). Many studies have been conducted to discover novel phytase producing bacteria that could satisfy the industrial needs, but have not been up to the mark (Goodfellow, Fiedler 2010; Nasrabadi et al. 2018), which has led to recombinant engineering of many phytase sources (Ushasree et al. 2017). Thus, keeping in mind the considerable number of sources and phytase producers,

the present study was aimed to use the combination of plate and activity assay for preliminary screening of novel phytase producers from different soil samples from various source locations. The study resulted in potential phytase producing isolates, which can be further characterized and used for various commercial applications.

Materials and methods

Sample collection

The samples were collected as per standard protocols published on a national agricultural portal of India (http://agritech.tnau.ac.in/agriculture/agri_soil_sampling.html). The method involved dividing the field into different units and collecting five samples from each unit by making a 'V' shaped cut, mixing the soil from each unit in a sterile manner and storing the amalgam in sterile flacons at 4 °C. Eleven sites were chosen from diversified sources, which included a forest area in Ankleshwar (Gujarat), Trikoni Garden in Mumbai (Maharashtra), agricultural fields in Baruch (Gujarat), and swine, poultry as well as other agricultural fields in Jawhar district (Maharashtra) (Table 1). These included soils from forest, garden, and fields with rice, oil seed (Kursani), udad (pulse), nachini, millet (jowar), vegetation (lady finger) and cotton, as well as swine and poultry enclosures. The collected soils were transported to Sunandan Divatia School of Science, NMIMS Deemed to be University, Mumbai in ice boxes, where all further analyses were performed. The soil samples were processed by sieving (2 mm) to remove litter and lumps. After processing, the soil samples were stored in sterile flacons at 4 °C until further analysis.

Enrichment, screening and isolation of bacteria

Phytate degrading microbes were enriched by inoculating 1 g of processed soil in autoclaved phytase screening broth (PSB) (1% D-glucose, 0.4% Na-phytate, 0.2% CaCl₂, 0.5% NH₄NO₃, 0.05% KCl, 0.05% MgSO₄ 7H₂O, 0.001% FeSO₄ 7H₂O, 0.001% MnSO₄ 7H₂O). D-glucose and Na-phytate solutions were filter sterilized and not autoclaved (Kerovuo et al. 1998; Kandil 2017). The inoculated broth was incubated at 180 rpm and 37 °C on a rotary shaker and screened after 24 h (day 1) and 120 h (day 5) on phytase screening agar (PSA) plates. In addition to the ingredients of PSB, 3% agar was added to prepare the PSA plates. The enriched broths were serially diluted and spread-plated on PSA plates and incubated at room temperature. The colonies that showed zones of hydrolysis similar to those of the wild type *E. coli* phytase zone were sampled and purified. To obtain pure cultures, colonies from the dilution plates were grown overnight in Luria-Bertani (LB) broth and tryptic soy (TS) broth at 37 °C and 180 rpm on a rotary shaker. This was followed by streaking of the turbid broths on LB agar plates to obtain pure cultures.

Table 1. Soil samples collected from different regions of Maharashtra and Gujarat in the winter season and used for isolation of phytase-producing microorganisms

Soil sample source	Location	Geographical coordinates	Soil type	Agroecology	Temperature
Forest	Ankleshwar, Gujrat	21.58102° N, 73.04682° E	Forest soil	Mango, guava, drumstick trees had grown in a 24281 m ² forest land	24 °C
Garden	Mumbai, Maharashtra	19.09935° N, 72.84850° E	Clay soil	Roses, <i>Aloe vera</i> , tulsi, herbs, shrubs were grown in a 790 m ² area garden	22 °C
Rice, oil seed (kursani), udad (pulse), nachini (millets)	Jawhar, Maharashtra	19.91855° N, 73.23496° E	Hill soil (rice, oil seed, udad, nachini)	Rice, oil seed, udad and nachini were grown together in a 4000 m ² field. The crops were harvested	26 °C
Swine and poultry	Jawhar, Maharashtra	19.91855° N, 73.23496° E	Silt soil (swine, poultry)	The swine and poultry enclosures were maintained for 20 and 4 years, respectively. Pigs were present during collection while the chicken were not present in the poultry since a week	26 °C
Millets (jowar), vegetation (lady finger) and cotton	Baruch, Gujrat	21.68931° N, 72.89728° E	Loam soil	These crops were grown in 4047 m ² land each. The crops were not harvested	24 °C

Phenotypic identification of bacterial isolates

All of the isolates were Gram-stained to understand their morphology and physiology and the selected bacterial isolates were further characterized using standard biochemical tests (Table 2; Ramesh et al. 2011; Dev et al. 2016). Biochemical tests included sugar fermentation tests (glucose, sucrose, lactose, maltose and mannitol), growth on MacConkey agar and eosin methylene blue agar, the indole test, methyl red and Voges-Proskauer test, nitrate reduction test, triple sugar iron test, lysine decarboxylase test, citrate test, urease test and catalase test. Gram staining and the biochemical tests were performed in duplicate. Primary bacterial identification was performed using the ABIS online tool (https://www.tgw1916.net/bacteria_logare_desktop.html) (Dev et al. 2016), which was based on morphology and biochemical properties. Identification of all isolates matched with the sequencing result at genus level, except for FD5T. The isolate was identified as *Aneurinibacillus aneurinilyticus* by ABIS and *Paenibacillus* by sequencing.

Qualitative phytase activity

The two-step counterstaining method adapted from Bae et al. (1999) eliminated false positive results due to acid producing bacteria (Van Staden et al. 2007). The six bacterial isolates and *E. coli* BL21 (DE3) cultures were spot plated (10 µL) on PSA plates and incubated at room temperature for 96 h. After incubation, the PSA plates were flooded with 2% (w/v) cobalt chloride solution at room temperature for 5 min. The cobalt chloride solution was then replaced with freshly prepared mixture of 6.25% (w/v) aqueous ammonium molybdate and 0.42% (w/v) ammonium vanadate solutions (1:1, v/v) for 2 to 3 min

(Lee et al. 2005; Park et al. 2012). The zones that did not regain turbidity and remained colourless were considered to be phytase producing positive isolates. The qualitative analysis was done in duplicate.

Molecular identification of bacterial isolates

Genomic DNA of the 6 putative phytase producers was isolated using the PureLink® Genomic DNA kit, following the manufacturer's protocol. The DNA isolates were quantified and checked for their purity on BioTek nanoplates using BioTek Gen5 software. PCR amplification of 16s rDNA gene was performed in duplicate using forward primer 151F (5'-GTGCCAGCMGCCGCGGTAA-3') and reverse primer Y36 (5'-GAAGGAGGTGWTCCADCC-3') under the following conditions: 30 cycles of initial denaturation at 95 °C for 3 min, denaturation at 95 °C for 45 s, annealing at 54 °C for 30 s, initial extension at 72 °C for 1 min 30 s and final extension at 72 °C for 10 min. The PCR products of the samples were digested with *MspI* (*HpaII*) restriction enzymes to carry out restriction fragment length polymorphism (RFLP) and the products were run on a 3% agarose gel. This process was carried out in duplicate. The shortlisted phytase producers were identified at species level via 16s rDNA sequencing using universal primers. The sequences were analysed on the NCBI database using blastn, and the 16s rDNA sequences of the species with the highest homology were extracted for *in silico* RFLP and phylogenetic analysis. *In silico* RFLP was performed on the sequenced sample data along with the extracted 16s rDNA of the homologous species on the Serial Cloner 2.6.1.0 software using the same restriction endonucleases. A neighbour-joining dendrogram with 1000 bootstrap testing was constructed using MEGA X 10.2.2 software

Table 2. Standard protocol followed for performing the biochemical tests of isolates from different soils. As a sample, 24-h-old isolated colonies suspended in sterile saline was used in all tests

Test	Reagents	Procedure
Sugar fermentation test	<ul style="list-style-type: none"> sterile peptone water, 0.2% phenol red, 1% of glucose, lactose, sucrose, maltose and mannitol, glassware: Durham's tubes 	<ul style="list-style-type: none"> To the peptone water, 0.2% phenol red was added and Durham's tube was inverted and then the test tubes were autoclaved. 1% of each sugar was added to the autoclaved set up. Aseptically the samples were inoculated and incubated at 37 °C for 24 h
Indole test	<ul style="list-style-type: none"> sterile peptone water, Kovac's reagent 	<ul style="list-style-type: none"> To the sterilized peptone water, the samples were inoculated and incubated at 37 °C for 24 h. 4 to 5 drops of Kovac's reagent was added to the tube from the walls gently to form a red colored ring
Methyl red test	<ul style="list-style-type: none"> sterile buffered glucose phosphate broth, methyl red 	<ul style="list-style-type: none"> To the sterilized broth, the samples were inoculated and incubated at 37 °C for 24 h. 4 to 5 drops of Methyl red indicator were added and observed for colour change
Voges-Proskauer test	<ul style="list-style-type: none"> sterile buffered glucose phosphate broth, α-naphthol, 40% KOH 	<ul style="list-style-type: none"> To the sterilized broth, the samples were inoculated and incubated at 37 °C for 24 h. 5 to 6 drops of α-naphthol and 2 to 3 drops of KOH were added and observed for colour change
Citrate utilization test	<ul style="list-style-type: none"> sterile Simmon's citrate agar 	<ul style="list-style-type: none"> To the solidified agar, the samples were streaked and incubated at 37 °C for 24 h. Observed for colour change
Urease test	<ul style="list-style-type: none"> sterile Christensen's agar, urea 	<ul style="list-style-type: none"> Urea was added aseptically to the autoclaved melted agar and then solidified. To the solidified agar, the samples were streaked and incubated at 37 °C for 24 h. Observed for colour change
Nitrate reduction test	<ul style="list-style-type: none"> sterile nitrate broth, sulphanilic acid, α- naphthylamine 	<ul style="list-style-type: none"> To the sterilized broth, the samples were inoculated and incubated at 37 °C for 24 h. 4 to 5 drops of sulphanilic acid and 4 to 5 drops of α-naphthylamine were added and observed for colour change
Triple sugar iron agar test	<ul style="list-style-type: none"> sterile triple sugar agar 	<ul style="list-style-type: none"> To the solidified agar, the samples were streaked and incubated at 37 °C for 24 h. Observed for colour change
Lysine decarboxylase test	<ul style="list-style-type: none"> sterile Moeller's medium, paraffin oil 	<ul style="list-style-type: none"> To the sterilized media, the samples were inoculated, overlaid with paraffin oil and incubated at 37 °C for 24 h. Observed for colour change
Catalase test	<ul style="list-style-type: none"> hydrogen peroxide solution 	<ul style="list-style-type: none"> The culture was spread on a clean glass slide and 2 to 3 drops of H₂O₂ were added. Observed for presence of bubbles

by clustering the obtained sample sequencing results with the extracted 16s rDNA sequence exploited in the *in silico* analysis.

Quantitative effect of pH on phytase activity

The phytase activity was estimated using glycine-HCl (pH 2.5) and sodium acetate (pH 5) buffer systems. The isolates did not exhibit any detectable activity in the alkaline pH range of Tris-HCl buffer (pH 7.5). Potassium dihydrogen phosphate (KH₂PO₄) was used for inorganic phosphate

standardization at concentrations 0.1 mM to 2 mM. Sodium phytate (44 mM stock) was used as substrate at 1:50 dilution in the respective buffer systems. The colour reagent solution was freshly prepared to terminate the reaction using 5% ammonium molybdate, 100% acetone, 5N sulphuric acid (1:2:1). After incubation for 30 min at 37 °C in a water bath, the phosphate released from sodium phytate hydrolysis was measured using the ammonium molybdate method, which is a modification of the Heinonen method (Suleimanova et al. 2015). The final reaction volume was 1

mL and performed in duplicate. To remove any turbidity that might remain after incubation, the reaction solution was centrifuged at 14000 rpm, 25 °C for 5 min. One unit of phytase activity (U mL⁻¹) was defined as the amount of phytase enzyme required to liberate 1 mmol of inorganic phosphate per minute by utilizing sodium phytate as the substrate under assay conditions.

To estimate the activity of intracellular phytase, 1 mL of overnight grown soil isolate cultures, in Luria-Bertani broth, were aliquoted and the cells were harvested at 10000 rpm, 25 °C for 10 min. Cells were lysed using B-cell lysis buffer (Sigma-Aldrich) and the freeze thaw method in the presence of 2% phenylmethylsulfonyl fluoride as the protease inhibitor. The lysate was collected by centrifugation at 10000 rpm, 4 °C for 10 min, discarding the pellet. The crude extract was used for measuring intracellular activity. The blanks that were set up were a substrate blank, enzyme blank and combination blank (reaction in which substrate and supernatant were not added). Bovine serum albumin stock (10 µg µL⁻¹) was used for protein standardization at concentrations 0.2 to 10 µg µL⁻¹.

Extracellular activity was also measured by checking activity in supernatant of 24 h grown cultures, but the isolates did not show any detectable extracellular activity.

Statistical analysis

The statistical significance of differences was determined by Student's *t*-tests followed by two-way ANOVA using the GraphPad Prism 9.0.0 software. For the comparison of the mean values, a 5% level of significance was considered.

Results

Enrichment, screening and isolation of bacteria

All eleven soil samples were enriched in PSB for 24 h (day 2 broths) followed by screening on sterile PSA plates. After 120 h (day 5 broths) of enrichment udad, nachini, millets, rice and oil seed samples were excluded from screening to due fungal overgrowth, leaving only six samples (forest, garden, swine and poultry enclosures, and vegetation and cotton fields) which were screened on PSA plates. The screening plates were observed after 24, 48 and 96 h. From the plates of the day 2 broth samples, only forest, swine enclosure and vegetation soil samples showed prominent halo zones after 24 h. At 48 h, garden plates also had decreased turbidity and increased growth of colonies. A few plates had filamentous fungal growth. All of the 96-h plates had excessive fungal growth and increased number of colonies in the lower dilution plates. Due to no halo zone on plates for day 2 broths of oil seed, cotton, udad, rice, millets and nachini field samples, they were eliminated and not used for bacterial isolation. Plating of the day 5 broths was done at dilution 10⁻³ due to excessive crowding in 10⁻¹ and 10⁻² dilutions in plates of day 2 broths. Dilutions higher than 10⁻⁵ were prepared for the samples that required more scattered colonies. Oil seed, udad, rice, millets and nachini

soil samples continued to not show a zone of hydrolysis and were hence eliminated.

Higher dilutions were used for further isolations. Isolation was done by sampling colonies from the 96 h plates on the basis of presence of a zone of hydrolysis around the colony. After screening, samples used for further isolation were broth plates (day 5) for cotton and broth plates (day 2 and day 5) for forest, garden, vegetation, and swine and poultry enclosure soil samples. The colonies were grown overnight in LB broth and TS broth followed by isolation on LB agar. Translucent and large colonies as well as opaque and small colonies were observed and further isolated for the forest sample on LB agar. The swine enclosure sample plate that was streaked from TS broth showed mucoid like colonies while all the other remaining sample plates had creamy and opaque colonies regardless of the broth from which LB agar plates were streaked. The isolates were named on the basis of the soil source (C, F, G, V, PS, P), the enrichment broth isolated from (D2, D5), the opacity (O, T) and the broth from which LB agar was streaked (LB, TS).

Phenotypic identification of bacterial isolates

Gram staining was used to identify the morphology and Gram nature. A total of thirteen isolated colonies from the LB agar were analysed. The samples GD2, GD5, VD2, PSD2, PSD5LB, PD2, and PD5 showed a bean shaped structure and were suspected to be yeast cultures, and were therefore eliminated. VD5, CD5 and PSD5TSB were Gram-negative bacteria and FD2, FD5T and FD5O were Gram-positive bacteria, which were confirmed by examining the growth on MacConkey agar and eosin methylene blue agar. The remaining six samples were biochemically characterized (Table 3). From the online ABIS tool the isolates were identified as *Aneurinibacillus aneurinilyticus* (FD5T), *Bacillus carboniphilus* (FD5O and FD2), *Enterobacter amnigenus* (VD5 and CD5), and *Raoultella terrigena* (PSD5TSB).

Qualitative phytase activity

Qualitative analysis of phytate degrading bacteria was done by differential staining. After the two-step staining process, the halos around the colonies remained colourless validating that the samples were true phytase producers and not acid producers. The amount of secreted phytase that led to the halo formation was measured in comparison to wild type *E. coli* BL21 (DE3). The PSD5TSB zone was greater than for *E. coli* and FD5O had similar zone size as *E. coli*. The remaining samples had zones smaller than the control (Table 4). This step along with the screening step confirmed the produced phytase as well as identified the bacterial samples that produced phytase but without any measurable activity. This method served as initial screening and identification method for phytase producing bacteria, but an additional rapid procedure of testing for intracellular enzyme activity was performed at a small scale.

Table 3. Results of biochemical tests of isolates FD5T, FD2, FD5O, PSD5TSB, VD5 and CD5. +, positive; -, negative; Y, yellow; Pi, pink; C, colourless; Pu, purple

Tests	Indication	Isolate					
		FD5T	FD5O	FD2	VD5	CD5	PSD5TSB
Glucose	Colour	-	-	-	+	+	+
	Gas	-	-	-	+	+	+
Maltose	Colour	-	-	-	+	+	+
	Gas	-	-	-	+	+	+
Lactose	Colour	-	-	-	-	-	+
	Gas	-	-	-	-	-	+
Sucrose	Colour	-	-	-	+	+	-
	Gas	-	-	-	+	+	+
Mannitol	Colour	-	-	-	+	+	+
	Gas	-	-	-	+	+	+
MacConkey agar		-	-	-	+ Y	+ Y	+ Pi (mucoid)
Eosin methylene blue agar		-	-	-	+ C	+ C	+ Pu (mucoid)
Indole		-	-	-	-	-	-
Methyl red		+	-	-	-	-	-
Voges-Proskauer		+	+	+	+	+	+
Nitrate reduction		+	-	-	+	+	+
Triple sugar iron	Butt	Y	Y	Y	Y	Y	Y
	Slant	Y	Y	Y	Y	Y	Y
	Gas	-	-	-	+	+	+
	H ₂ S	-	-	-	-	-	-
Lysine decarboxylase		-	-	-	-	-	+
Citrate		-	-	-	Y	Y	Y
Urease		-	-	-	-	-	-
Catalase		-	+	+	+	+	+

Molecular identification of bacterial isolates

The gDNA of overnight grown FD5T, FD5O, FD2, VD5, CD5 and PSD5TSB samples was isolated using a PureLink® Genomic DNA kit. The size of isolated gDNA was greater than 10 kb. The quantity and quality of the isolated DNA of samples were measured by nanoplate UV transmission (Table 5). PSD5TSB had the highest concentration of extracted gDNA among the samples. Amplification of the 16s rDNA of the genome was done using universal primers, 151F as the forward primer and Y36 as the reverse primer. The 1 kb PCR amplicon was amplified and resolved on an agarose gel.

Table 4. The diameters of the halo zones formed by the isolates (FD5T, FD2, FD5O, PSD5TSB, VD5, CD5) and wildtype *E. coli* BL21 (DE3) on counter-stained PSA plates

Isolate	Zone of hydrolysis (mm)
FD5T	10
PSD5TSB	14
CD5	9
FD5O	12
FD2	11
VD5	8
Wild type <i>E. coli</i>	13

RFLP was performed on amplified 16s rDNA product using *MspI* (*HpaII*) enzymes. PSD5TSB, VD5 and CD5 had similar band patterns, which was also observed for FD5O and FD2. FD5T showed a different pattern from the other samples. *In silico* RFLP analysis of the sequenced data showed a similar number of bands with the same band size as those of the agarose gel, validating the performed RFLP. The *in silico* comparison of the sequenced data band pattern with those of the extracted species sequence showed similarity, which indicated that the species identification could be correct. Species level identification of the unknown samples was done by 16s rDNA sequencing. The

Table 5. Nanodrop quantification and qualification of isolated gDNA from phytase-producing bacteria (FD5T, FD2, FD5O, PSD5TSB, VD5, CD5)

Isolate	Concentration of DNA (ng μL^{-1})	A_{260}/A_{280} ratio
FD5T	19.773	1.985
PSD5TSB	140.446	1.507
CD5	85.607	1.847
FD5O	9.053	1.517
FD2	12.349	1.914
VD5	112.757	1.946

samples were submitted for sequencing and were identified as *Enterobacter cloacae* (VD5 and CD5), *Paenibacillus* sp. (FD5T), *Bacillus megaterium* (FD5O and FD2), *Klebsiella variicola* (PSD5TSB). VD5 and CD5 were 99.2 and 99.6% identical to *Enterobacter cloacae* respectively (GenBank Accession no. CP046116.1, E value 0), FD5T was 98.37% identical to *Paenibacillus* sp. (GenBank Accession no. MK681944.1, E value 0), FD5O and FD2 were 99.4 and 99.2% identical to *Bacillus megaterium*, respectively (GenBank Accession no. CP045272.1, E value 0), and PSD5TSB was 99.4% identical to *Klebsiella variicola* (GenBank Accession no. CP050958.1, E value 0). A phylogenetic tree for the six samples was built based on homology of the 16s rDNA sequence (Fig. 1). CD5 and VD5 were closely clustered with *Enterobacter cloacae* strain ATCC 13047, PSD5TSB was in the same cluster as *Klebsiella aerogenes* KCTC 2190, FD5T was in the same cluster as *Paenibacillus yonginensis* strain DCY84 and FD5O along with FD2 was clustered with *Bacillus* sp. 3401BRRJ. PSD5TSB, CD5 and VD5 were closely related, which was portrayed in the phenotypic identification as well as RFLP analysis. *Paenibacillus* sp. (FD5T) had very close similarity to *Paenibacillus yonginensis* 16s rDNA. The dendrogram showed that there was similarity between *Bacillus* sp. and *Paenibacillus* sp. strains.

Quantitative effect of pH on phytase activity

Intracellular phytase enzyme activity and protein estimation were detected using the Heinonen method with pH as the limiting factor. The R^2 values were close to 1, indicating that the values obtained for the concentration of inorganic phosphate and protein concentration were acceptable. All of the readings were taken in duplicate with standard deviation close to 0 between the duplicates. The wild type *E. coli* phytase was used as a standard to confirm phytase activity. At pH 2.5, *E. coli* phytase showed an average of 135 times more activity than FD5T, PSD5TSB, CD5 and VD5 while FD5O and FD2 showed no activity at this pH. On the other hand, there was a significant decrease in activity

of *E. coli* phytase at pH 5. PSD5TSB, CD5, FD5O, FD2 and VD5 and FD5T had higher activity at pH 5 as compared to pH 2.5. At pH 7.5, all samples including *E. coli* had a small amount of inorganic phosphate liberated, which could not be detected using the assay. *E. coli* had highest activity at pH 2.5 while the unknown samples had highest activity at pH 5 (Fig. 2). This technique aided in inferring the optimum pH for phytase samples on a small scale. Furthermore, it helped in narrowing down and identifying the samples that had intracellular activity comparable with the standard. Hence, these were further analysed and examined for large scale characterization.

To compare the intracellular and extracellular activity of the phytase enzyme, the reaction mixtures for extracellular activity were prepared in the same manner as that of intracellular testing with the use of the three pH buffers. The isolates were grown in PSB for 24 h and the supernatant was used as the phytase sample and checked for extracellular activity. KH_2PO_4 standard curves were prepared and had R^2 values close to 1. All of the unknown samples had undetectable amounts of liberated inorganic phosphate, while the *E. coli* sample had a significant amount of extracellular activity in the three pH buffers. Hence, the samples expressed only intracellular phytase with measurable activity.

Discussion

Phytase is a class of phosphatase and the respective bacteria are frequently termed as phosphate-solubilizing bacteria, which are a type of plant growth promoting bacteria (Igual et al. 2001). Phosphorus is an essential macronutrient. Phytate is the predominant form of organic P present in soil. Plants cannot utilize phytate directly and it needs to be mineralized by phytase enzyme (Singh et al. 2014; Suleimanova et al. 2015; Motamedi 2016; Alori et al. 2017; Caffaro et al. 2019). Hence, phosphate-solubilizing bacteria are found in the rhizosphere of crops that grow in P-rich soil (Singh et al. 2014; Caffaro et al. 2019). The

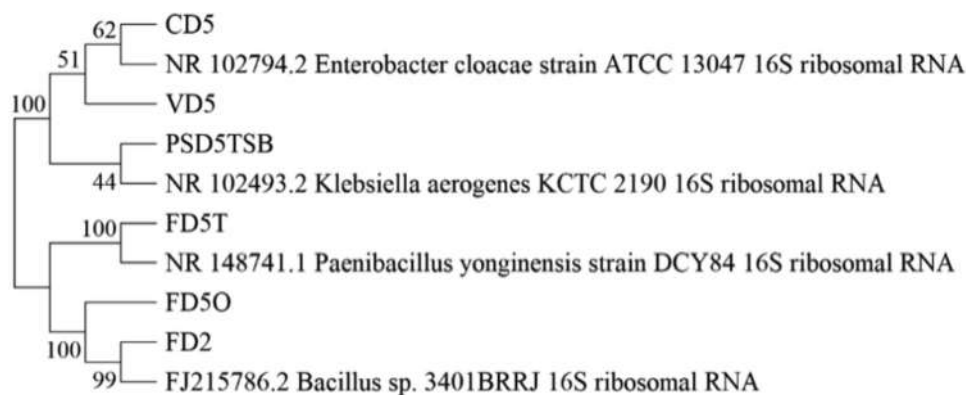


Fig. 1. Neighbor-joining dendrogram showing the relation between the samples (FD5T, FD2, FD5O, PSD5TSB, VD5, CD5) and the 16s rDNA of their closest related strains (*Enterobacter cloacae* strain ATCC 13047, *Klebsiella aerogenes* KCTC 2190, *Paenibacillus yonginensis* strain DCY84 and *Bacillus* sp. 3401BRRJ) obtained from GenBank. The tree was constructed by 1000 bootstrap tests.

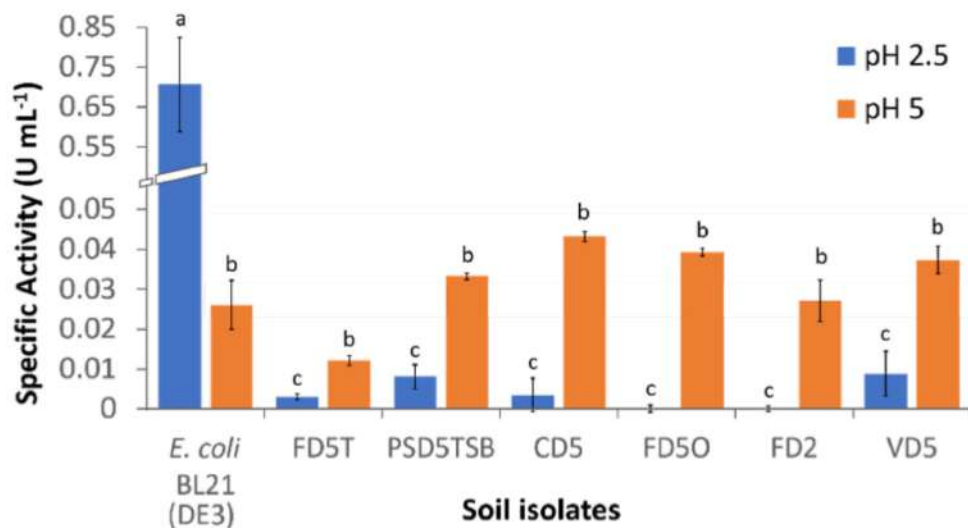


Fig. 2. Comparison of intracellular specific activity at pH 2.5, pH 5 of phytase from the samples (FD5T, FD2, FD5O, PSD5TSB, VD5, CD5) with wildtype *E. coli* BL21 (DE3) phytase as the positive control. The samples had no activity at pH 7.5. Data is shown as mean \pm standard error (U mL⁻¹). Statistical significance was estimated by Student's *t*-test and two-way ANOVA at $p \leq 0.05$. The letters depict the significant difference in mean for each soil isolate at pH 2.5 and pH 5 as well as the difference across the soil isolates in their respective pH.

intestinal microbiota in monogastric animals is devoid of species secreting phytase, leading to unutilized phytate in the gut. Phytase is thus excreted and found in enough concentration for the phytase-producing bacteria to thrive (Mittal et al. 2011). Thus, phytase-producing bacteria are not only found in the rhizospheric soil of grain, legume, nut, oil seed crops (Dahiya 2016), but also found in soil around monogastric animal shelters (Caffaro et al. 2019). A wide range of organisms (fungi, bacteria, yeast, plants and animals) are known to express phytase (Li et al. 2019).

Phenotypic identification of bacteria by standard methods of morphological and biochemical characterization has been a key step for many years in studies on phytase producing bacteria such as *Bacillus* sp., *Achromobacter* sp., *Tetrathlobacter* sp., *Klebsiella* sp. (Tye et al. 2002; Roy et al. 2009; Mittal et al. 2011; Khan, Ghosh 2012; Kumar et al. 2013). In recent years, this traditional method has been applied for primary identification of phytase producers like *Acinetobacter* sp., *Enterobacter* sp., several *Bacillus* sp. and *Lactobacillus* sp. (Ibnu Irwan et al. 2017; Alias et al. 2018; Muslim et al. 2018; Onipede et al. 2020). While these studies have biochemically identified the species using the Bergey's manual, an upcoming online tool, ABIS, has shown promising results in species level identification (Rahman et al. 2017; Stoica, Sorescu 2017; Siddique, Alif 2018; Guder, Krishna 2019). In the present study, primary identification of *Bacillus carboniphilus* (FD5O and FD2), *Enterobacter amnigenus* (VD5 and CD5), *Raoultella terrigena* (PSD5TSB) and *Aneurinibacillus aneurinilyticus* (FD5T) from the six samples was made based on their Gram-nature and biochemical character.

Phytate degrading bacteria such as *Serratia* sp.,

Enterobacter sp., *Paenibacillus* sp., *Bacillus* sp., *Lactobacillus* sp., *Geobacillus* sp. can be derived from diverse sources from rhizospheric soil, compost, poultry farms, cattle shade, fermented foods, volcanic ash and geysers (Jorquera et al. 2011; Singh et al. 2013; Sajidan et al. 2015; Kalsi et al. 2016; Savita et al. 2017; Jorquera et al. 2018). These studies, among many others, screened various samples on PSA for visualization of the zone of hydrolysis and studies were continued with sample that showed a satisfactory halo zone. Nonetheless, these zones around colonies can be false positive due to acid-producing bacteria for which the counterstaining step has to be used (Chanderman et al. 2016; Monika et al. 2017; Nasrabadi et al. 2018). Bacteria from forest (FD2, FD5O, FD5T), vegetation (VD5), cotton (CD5) and pig enclosure (PSD5TSB) soil samples were all isolated after screening on PSA for the presence of a halo zone, and were then confirmed as true phytase producers by qualitative analysis. From the eleven soil samples, oil seed, udad, rice, millets and nachini samples were eliminated due to lack of a halo zone, which aided in narrowing down the samples and saving resources for further analysis. *E. coli* is known to have high phytase activity against phytate and therefore was used as a positive control (Helian et al. 2020).

RFLP is a molecular technique for identification of the bacterial isolates, which is cost effective, fast and can be repeated again (Miao et al. 2013). This fingerprinting technique based on the 16s rDNA part of the genome has been used to study the microbial diversity present in several habitats (Kushwaha et al. 2020). While there have been few studies performing 16s rDNA RFLP procedures for initial identification of phytase expressing bacteria in recent years, this quick method helped in visually

identifying and assessing the variability of the phytase producers, as was done for *Pseudomonas* sp., *Acinetobacter* sp., *Agrobacterium* sp. and *Arthrobacter* sp., among many other species (Sanguin et al. 2016). In the current study, by comparing the pattern of resolution on the gel, RFLP results showed that there were four different species of isolated bacteria. This was confirmed by *in silico* testing and ultimately sequencing showing that FD2 and FD5O were *Bacillus megaterium*, FD5T was *Paenibacillus* sp., VD5 and CD5 were *Enterobacter cloacae*, and PSD5TSB was *Klebsiella variicola*.

For intracellular activity analysis, the BRENDA enzyme database (www.brenda-enzymes.org) provided concise kinetics of enzymes (Jeske et al. 2019) and indicated the optimal pH of phytase producing bacteria (EC 3.1.3.8). *Bacillus* sp. express β -propeller phytase (also known as alkaline phytase), which is dependent on Ca^{2+} as a cofactor for its activity at alkaline pH, while remaining inactive at acidic pH (Tran et al. 2011a; Tran et al. 2011b). FD5O and FD2 were *Bacillus megaterium*, which produces β -propeller phytase (Kumar et al. 2017). This phytase was rendered inactive due to the absence of a source of calcium in the reaction set up. *Paenibacillus* sp. are also alkaline phytase producers, which remain inactive at acidic pH (Kumar et al. 2017). A *Paenibacillus* sp. was characterized with optimal activity towards pH 5 and almost no activity at pH 7 (Acuña et al. 2011). *Enterobacter cloacae* (VD5 and CD5) and *Klebsiella variicola* (PSD5TSB) had optimal phytase activity at pH 5. The pH range for *Enterobacter* sp., which was isolated from the rhizosphere, was reported to be pH 2 to 6 (Chanderman et al. 2016), indicating no activity toward alkaline pH. In studies conducted on effect of pH, *Enterobacter cloacae* showed highest activity at pH 5 (Suliasih, Widawati 2020). Different *Klebsiella* sp. have shown optimum activity at pH 5 (Sajidan et al. 2004; Elkhalil et al. 2007), with *Klebsiella variicola* having reduced activity at pH 7 (López Ortega et al. 2013). *E. coli* enzyme, which is categorized as histidine acid phytase, had higher activity towards an acidic pH (Okamoto et al. 2017; Balaban et al. 2018). The phytase of *E. coli* BL21 (DE3) and identified isolates showed no activity at alkaline pH 7.5. At pH 2.5, *E. coli* showed an average of 135 times more activity than FD5T, PSD5TSB, CD5, and VD5 while FD5O and FD2 showed no activity at this pH. On the other hand, there was a significant decrease of activity of *E. coli* at pH 5 as compared to pH 2.5. PSD5TSB, CD5, FD5O, FD2 and VD5 had slightly higher activity than *E. coli* while FD5T did not show much improvement from pH 2.5 to pH 5.

Studies with phytases from *Bacillus megaterium* (D. Kumar et al. 2013), *Paenibacillus* sp. (Khianngam et al. 2017), *Enterobacter cloacae* (Onawola et al. 2019) and *Klebsiella variicola* (López Ortega et al. 2013) have shown that for significant detection of extracellular phytase secreted in the media, a minimum of 48 h of incubation is required. The optimal incubation time for *E. coli* phytase production was

determined to be 48 h, but it also had measurable activity after 24 h (Wang et al. 2015). The current study quantified phytase only after 24 h. During this period, apart from *E. coli*, substantial phytase was not secreted by the soil isolates and hence not detected by the assay. On the basis of these data, the isolated bacteria have great potential to excrete phytase after sufficient incubation and further condition optimization.

Preliminary screening of phytase producing bacteria by analyzing the zone of hydrolysis is a commonly used method. In this study, intracellular activity was also determined to accurately narrow down the number of samples and confirm the presence of phytase. At a mini-scale, these techniques provide higher confidence for further analysis, which occurs at a larger scale using phytase purification and characterization. The ability to solubilize phosphate is not specific to a single genera (Motamedi 2016). Due to high soil microbial diversity (Kumar et al. 2016), implementing preliminary screening reduces additional costs and resources of large scale experimentation on a large number of isolates. Therefore, this study shows an alternative at a mini scale to acquire effective bacterial phytase producers.

Acknowledgements

The authors would like to acknowledge Sunandan Divatia School of Science, NMIMS Deemed to be University, Mumbai, India for providing the opportunity and the resources for the project research. The authors also want to thank Ms. Rajvi Shah and Ms. Freyan Vakharia for their contributions in the research work. Authors would like to give special thanks to Mr. Mitesh Joshi and Ms. Shriya Sawant for their intellectual inputs and suggestions at various points in the research.

References

- Acuña J.J., Jorquera M.A., Martínez O.A., Menezes-Blackburn D., Fernández M.T., Marschner P., Greiner R., Mora M.L. 2011. Indole acetic acid and phytase activity produced by rhizosphere bacilli as affected by pH and metals. *J. Soil Sci. Plant Nutr.* 11: 1–12.
- Akhmetova A.I., Nyamsuren C., Balaban N.P., Sharipova M.R. 2013. Isolation and characterization of a new bacillary phytase. *Russ. J. Bioor. Chem.* 39: 384–389.
- Alias N., Shunmugam S., Ong P.Y. 2018. Isolation and molecular characterization of phytase producing bacteria from Malaysia hot springs. *J. Fund. Appl. Sci.* 9: 852–865.
- Alori E.T., Glick B.R., Babalola O.O. 2017. Microbial phosphorus solubilization and its potential for use in sustainable agriculture. *Front. Microbiol.* 8: 1–8.
- Bae H.D., Yanke L.J., Cheng K.J., Selinger L.B. 1999. A novel staining method for detecting phytase activity. *J. Microbiol. Meth.* 39: 17–22.
- Bajaj B.K., Wani M.A. 2015. Purification and characterization of a novel phytase from *Nocardia* sp. MB 36. *Biocatal. Biotransform.* 33: 141–149.
- Balaban N.P., Suleimanova A.D., Shakirov E. V., Sharipova M.R. 2018. Histidine acid phytases of microbial origin. *Microbiol. Russ. Fed.* 87: 745–756.

- Baruah K., Norouzitallab P., Pal A.K. 2017. Development of low cost and eco-friendly feed for various candidate species for the sustainability of commercial aquaculture and reduction of aquatic pollution. In: Bagchi D., Nair S. (eds) *Developing New Functional Food and Nutraceutical Products*. Elsevier, pp. 441–453.
- Bhavsar K., Khire J.M. 2014. Current research and future perspectives of phytase bioprocessing. *RSC Adv.* 4: 26677–26691.
- Bohn L., Meyer A.S., Rasmussen S.K. 2008. Phytate: Impact on environment and human nutrition. A challenge for molecular breeding. *J. Zhejiang Univ. Sci. B* 9: 165–191.
- Caffaro M.M., Balestrasse K.B., Rubio G. 2019. Adsorption to soils and biochemical characterization of purified phytases. *SOIL Discuss.* 6: 1–17.
- Caipang C.M.A., Dechavez R.B., Amar M.J.A. 2011. Potential application of microbial phytase in aquaculture. *ELBA Biofulux* 3: 55–66.
- Chanderman A., Puri A.K., Permaul K., Singh S. 2016. Production, characteristics and applications of phytase from a rhizosphere isolated *Enterobacter* sp. *ACSS. Bioproc. Biosyst. Eng.* 39: 1577–1587.
- Costenaro-Ferreira C., Della Flora M.A.L. 2017. Challenges for efficient use of phytase in fish nutrition. *Agrarian* 10: 95–104.
- Coutinho T.C., Tardioli P.W., Farinas C.S. 2020. Phytase Immobilization on hydroxyapatite nanoparticles improves its properties for use in animal feed. *Appl. Biochem. Biotechnol.* 190: 270–292.
- Dahiya S. 2016. Role of phytate and phytases in human nutrition. *Int. J. Food Sci. Nutr.* 1: 39–42.
- Dersjant-Li Y., Awati A., Schulze H., Partridge G. 2015. Phytase in non-ruminant animal nutrition: A critical review on phytase activities in the gastrointestinal tract and influencing factors. *J. Sci. Food Agric.* 95: 878–896.
- Dev S.S., E.A. N., Venu A. 2016. Biochemical and molecular characterization of efficient phytase producing bacterial isolates from soil samples. *Int. J. Curr. Microbiol. Appl. Sci.* 5: 218–226.
- Dvořáková J. 1998. Phytase: sources, preparation and exploitation. *Folia Microbiol.* 43: 323–338.
- Elkhalil E.A.I., Männer K., Borriss R., Simon O. 2007. In vitro and in vivo characteristics of bacterial phytases and their efficacy in broiler chickens. *British Poultry Sci.* 48: 64–70.
- Golovan S., Wang G., Zhang J., Forsberg C.W. 1999. Characterization and overproduction of the *Escherichia coli* *appA* encoded bifunctional enzyme that exhibits both phytase and acid phosphatase activities. *Can. J. Microbiol.* 46: 59–71.
- Goodfellow M., Fiedler H.P. 2010. A guide to successful bioprospecting: Informed by actinobacterial systematics. *Antonie Van Leeuwenhoek* 98: 119–142.
- Grases F., Costa-Bauza A. 2019. Key aspects of myo-inositol hexaphosphate (phytate) and pathological calcifications. *Molecules* 24: 4434.
- Greiner R., Carlsson N.G. 2006. myo-inositol phosphate isomers generated by the action of a phytate-degrading enzyme from *Klebsiella terrigena* on phytate. *Can. J. Microbiol.* 52: 759–768.
- Greiner R., Farouk A.E., Carlsson N.G., Konietzny U. 2007. myo-inositol phosphate isomers generated by the action of a phytase from a Malaysian waste-water bacterium. *Protein J.* 26: 577–584.
- Guder D.G., Krishna M.S.R. 2019. Isolation and characterization of potential cellulose degrading bacteria from sheep rumen. *J. Pure Appl. Microbiol.* 13: 1831–1839.
- Helian Y., Gai Y., Fang H., Sun Y., Zhang D. 2020. A multistrategy approach for improving the expression of *E. coli* phytase in *Pichia pastoris*. *J. Ind. Microbiol. Biotechnol.* 47: 1161–1172.
- Ibnu Irwan I., Agustina L., Natsir A., Ahmad A. 2017. Isolation and characterization of phytase-producing thermophilic bacteria from Sulili Hot Springs in South Sulawesi. *Sci. Res. J.* 5: 16.
- Igual J.M., Valverde A., Cervantes E., Velázquez E. 2001. Phosphate-solubilizing bacteria as inoculants for agriculture: Use of updated molecular techniques in their study. *Agronomie* 21: 561–568.
- Jatuwong K., Suwannarach N., Kumla J., Penkhrue W., Kakumyan P., Lumyong S. 2020. Bioprocess for production, characteristics, and biotechnological applications of fungal phytases. *Front. Microbiol.* 11: 1–18.
- Jeske L., Placzek S., Schomburg I., Chang A., Schomburg D. 2019. BRENDA in 2019: A European ELIXIR core data resource. *Nucleic Acids Res.* 47: 542–549.
- Jorquera M.A., Crowley D.E., Marschner P., Greiner R., Fernández M.T., Romero D., Menezes-Blackburn D., De La Luz Mora M. 2011. Identification of β -propeller phytase-encoding genes in culturable *Paenibacillus* and *Bacillus* spp. from the rhizosphere of pasture plants on volcanic soils. *FEMS Microbiol. Ecol.* 75: 163–172.
- Jorquera M.A., Gabler S., Inostroza N.G., Acuña J.J., Campos M.A., Menezes-Blackburn D., Greiner R. 2018. Screening and characterization of phytases from bacteria isolated from Chilean hydrothermal environments. *Microb. Ecol.* 75: 387–399.
- Kalsi H.K., Singh R., Dhaliwal H.S., Kumar V. 2016. Phytases from *Enterobacter* and *Serratia* species with desirable characteristics for food and feed applications. *3 Biotech.* 6: 1–13.
- Kandil M.M. 2017. Isolation and characterization of highly effective phytate-mineralizing *Klebsiella pneumoniae* strain MK1C adapted to arid and semiarid conditions from calcareous soil. *Alexandria Sci. Exch. J.* 38: 819–827.
- Karpagam T., Nagalakshmi P.K. 2014. Isolation and characterization of phosphate solubilizing microbes from agricultural soil. *Int. J. Curr. Microbiol. Appl. Sci.* 3: 601–614.
- Kerovuori J., Lauraeus M., Nurminen P., Kalkkinen N., Apajalahti J. 1998. Isolation, characterization, molecular gene cloning, and sequencing of a novel phytase from *Bacillus subtilis*. *Appl. Environ. Microbiol.* 64: 2079–2085.
- Khan A., Ghosh K. 2012. Characterization and identification of gut-associated phytase-producing bacteria in some fresh water fish cultured in ponds. *Acta Ichthyol. Piscat.* 42: 37–45.
- Khiannang S., Pootupaaeng-on Y., Sonloy A., Kajorn-aroonkij J., Tanasupawat S. 2017. Characterization and comparison of phytase production by *Bacillus* and *Paenibacillus* strains from Thai soils. *Malays. J. Microbiol.* 13: 318–325.
- Kim B.H., Lee J.Y., Lee P.C.W. 2015. Purification, sequencing and evaluation of a divergent phytase from *Penicillium oxalicum* KCTC6440. *J. Gen. Appl. Microbiol.* 61: 117–123.
- Konietzny U., Greiner R. 2004. Bacterial phytase: Potential application, *in vivo* function and regulation of its synthesis. *Braz. J. Microbiol.* 35: 11–18.
- Kumar D., Rajesh S., Balashanmugam P., Rebecca L.J., Kalaichelvan P.T. 2013. Screening, optimization and application of extracellular phytase from *Bacillus megaterium* isolated from poultry waste. *J. Modern Biotechnol.* 2: 46–52.
- Kumar V., Singh P., Jorquera M.A., Sangwan P., Kumar P., Verma A.K., Agrawal S. 2013. Isolation of phytase-producing

- bacteria from Himalayan soils and their effect on growth and phosphorus uptake of Indian mustard (*Brassica juncea*). *World J. Microbiol. Biotechnol.* 29: 1361–1369.
- Kumar V., Sinha A.K., Makkar H.P.S., Becker K. 2010. Dietary roles of phytate and phytase in human nutrition: A review. *Food Chem.* 120: 945–959.
- Kumar V., Yadav A.N., Saxena A., Sangwan P. 2016. Unravelling rhizospheric diversity and potential of phytase producing microbes. *SM J. Biol.* 2: 1–2.
- Kumar V., Yadav A.N., Verma P., Sangwan P., Saxena A., Kumar K., Singh B. 2017. β -Propeller phytases: Diversity, catalytic attributes, current developments and potential biotechnological applications. *Int. J. Biol. Macromol.* 98: 595–609.
- Kushwaha M., Surabhi, Marwa N., Pandey V., Singh N. 2020. Advanced tools to assess microbial diversity and their functions in restoration of degraded ecosystems. In: Singh J.S., Vimal S.R. (eds) *Microbial Services in Restoration Ecology*. Elsevier, pp. 83–97.
- Lee D.-H., Choi S.-U., Hwang Y.-I. 2005. Culture conditions and characterizations of a new phytase-producing fungal isolate, *Aspergillus* sp. L117. *Mycobiology* 33: 223–229.
- Lei X.G., Porres J.M. 2003. Phytase enzymology, applications, and biotechnology. *Biotechnol. Lett.* 25: 1787–1794.
- Li Y., Xu H., Xu J., Pang R., Xu B. 2019. Structure prediction and enzymatic properties of phytase phyS. *Adv. Enzym. Res.* 7: 57–65.
- López Ortega M.D.P., Criollo Campos P.J., Gómez Vargas R.M., Camelo Runsinque M., Estrada Bonilla G., Garrido Rubiano M.F., Bonilla Buitrago R. 2013. Characterization of diazotrophic phosphate solubilizing bacteria as growth promoters of maize plants. *Rev. Colomb. Biotecnol.* 15: 115–123.
- Miao Y.-Z., Xu H., Fei, Bao-Jin, Qiao, Dai-Rong, Cao Y. 2013. PCR – RFLP analysis of the diversity of phytate-degrading bacteria in the Tibetan Plateau. *Can. J. Microbiol.* 251: 245–251.
- Mittal A., Singh G., Goyal V., Yadav A., Aneja K.R., Gautam S.K., Aggarwal N.K. 2011. Isolation and biochemical characterization of acido-thermophilic extracellular phytase producing bacterial strain for potential application in poultry feed. *Jundishapur J. Microbiol.* 4: 273–282.
- Monika, Savitri, Kumar V., Kumari A., Angmo K., Bhalla T.C. 2017. Isolation and characterization of lactic acid bacteria from traditional pickles of Himachal Pradesh, India. *J. Food Sci. Technol.* 54: 1945–1952.
- Motamedi H. 2016. Screening cabbage rhizosphere as a habitat for isolation of phosphate-solubilizing bacteria. *Environ. Exp. Biol.* 14: 173–181.
- Mullaney E.J., Ullah A.H.J. 2003. The term phytase comprises several different classes of enzymes. *Biochem. Biophys. Res. Commun.* 312: 179–184.
- Muslim S.N., AL-Kadmy I.M.S., Khazaal S.S., Ali A.N.M., Ibrahim S.A., Al-Saryi N.A., Al-saadi L.G., Muslim S.N., Salman B.K., Aziz S.N. 2018. Screening, nutritional optimization and purification for phytase produced by *Enterobacter aerogenes* and its role in enhancement of hydrocarbons degradation and biofilm inhibition. *Microb. Pathog.* 115: 159–167.
- Nasrabadi R.G., Greiner R., Yamchi A., Nourzadeh Roshan E. 2018. A novel purple acid phytase from an earthworm cast bacterium. *J. Sci. Food Agric.* 98: 3667–3674.
- Okamoto A.K., Miranda M.A., Granjeiro J.M., Aoyama H. 2017. Enzymes with two optimum pH values. *World J. Res. Rev.* 5: 5–11.
- Onawola O.O., Akande I.S., Okunowo W.O., Osuntoki A.A. 2019. Isolation and identification of phytase-producing *Bacillus* and *Enterobacter* species from Nigerian soils. *Niger. J. Biotechnol.* 36: 127–138.
- Onipede G., Aremu B., Sanni A., Babalola O. 2020. Molecular study of the phytase gene in lactic acid bacteria isolated from ogi and kunun-zaki, african fermented cereal gruel and beverage. *Appl. Food Biotechnol.* 7: 49–60.
- Park I., Lee J., Cho J. 2012. Degradation of phytate pentamagnesium salt by *Bacillus* sp. T4 phytase as a potential eco-friendly feed additive. *Asian-Australasian J. Anim. Sci.* 25: 1466–1472.
- Rahman S.S., Siddique R., Tabassum N. 2017. Isolation and identification of halotolerant soil bacteria from coastal Patenga area. *BMC Res. Notes* 10: 4–9.
- Ramesh A., Sharma S.K., Joshi O.P., Khan I.R. 2011. Phytase, phosphatase activity and p-nutrition of soybean as influenced by inoculation of *Bacillus*. *Indian J. Microbiol.* 51: 94–99.
- Rocky-Salimi K., Hashemi M., Safari M., Mousivand M. 2016. A novel phytase characterized by thermostability and high pH tolerance from rice phyllosphere isolated *Bacillus subtilis* B.S.46. *J. Adv. Res.* 7: 381–390.
- Roy T., Mondal S., Ray A.K. 2009. Phytase-producing bacteria in the digestive tracts of some freshwater fish. *Aquacult. Res.* 40: 344–353.
- Sajidan A., Farouk A., Greiner R., Jungblut P., Müller E.C., Borriss R. 2004. Molecular and physiological characterisation of a 3-phytase from soil bacterium *Klebsiella* sp. ASR1. *Appl. Microbiol. Biotechnol.* 65: 110–118.
- Sajidan, Wulandari R., Sari E.N., Ratriyanto A., Weldekiros H., Greiner R. 2015. Phytase-producing bacteria from extreme regions in Indonesia. *Brazilian Arch. Biol. Technol.* 58: 711–717.
- Sanguin H., Wilson N.L., Kertesz M.A. 2016. Assessment of functional diversity and structure of phytate-hydrolysing bacterial community in *Lolium perenne* rhizosphere. *Plant Soil* 401: 151–167.
- Savita P., Yallappa M., Nivetha N., Suvarna V. 2017. Phytate solubilizing microorganisms and enzyme phytase to combat nutritional problems in cereal-based foods. *J. Bacteriol. Mycol.* 4: 86–89.
- Shobirin A., Hussin M., Farouk A., Salleh H.M. 2010. Phytate-degrading enzyme and its potential biotechnological application: a review. *J. Agrobiotechnology* 1: 1–16.
- Siddique R., Alif F. 2018. Isolation and identification of Orange M2R and Green GS dye decolourizing bacteria from textile sludge (soil) samples and determination of their optimum decolourization conditions. *Annu. Res. Rev. Biol.* 22: 1–12.
- Singh N.K., Joshi D.K., Gupta R.K. 2013. Isolation of phytase producing bacteria and optimization of phytase production parameters. *Jundishapur J. Microbiol.* 6: e6419.
- Singh P., Kumar V., Agrawal S. 2014. Evaluation of phytase producing bacteria for their plant growth promoting activities. *Int. J. Microbiol.* 2014: 426483.
- Van Staden J., Den Haan R., Van Zyl W.H., Botha A., Viljoen-Bloom M. 2007. Phytase activity in *Cryptococcus laurentii* ABO 510. *FEMS Yeast Res.* 7: 442–448.
- Stoica C., Sorescu I. 2017. ABIS online - Advanced Bacterial Identification Software, an original tool for phenotypic bacterial identification. *Regnum Prokaryotae* 10: 1–13.
- Suleimanova A.D., Beinhauer A., Valeeva L.R., Chastukhina I.B., Balaban N.P., Shakirov E. V., Greiner R., Sharipova M.R. 2015.

- Novel glucose-1-phosphatase with high phytase activity and unusual metal ion activation from soil bacterium *Pantoea* sp. strain 3.5.1. *Appl. Environ. Microbiol.* 81: 6790–6799.
- Suliasih S., Widawati S. 2020. Phytase production in *Enterobacter cloacae*. *Biotropia* 27: 282–291.
- Tran T.T., Hashim S.O., Gaber Y., Mamo G., Mattiasson B., Hatti-Kaul R. 2011. Thermostable alkaline phytase from *Bacillus* sp. MD2: Effect of divalent metals on activity and stability. *J. Inorg. Biochem.* 105: 1000–1007.
- Tran T.T., Mamo G., Búxo L., Le N.N., Gaber Y., Mattiasson B., Hatti-Kaul R. 2011. Site-directed mutagenesis of an alkaline phytase: Influencing specificity, activity and stability in acidic milieu. *Enzyme Microb. Technol.* 49: 177–182.
- Tye A., Siu F., Leung T., Lim B. 2002. Molecular cloning and the biochemical characterization of two novel phytases from *B. subtilis* 168 and *B. licheniformis*. *Appl. Microbiol. Biotechnol.* 59: 190–197.
- Ushasree M.V., Shyam K., Vidya J., Pandey A. 2017. Microbial phytase: Impact of advances in genetic engineering in revolutionizing its properties and applications. *Bioresour. Technol.* 245: 1790–1799.
- Wang T., Li W., Hao Y., Liu Y. 2015. Enzymatic properties of phytase from *Escherichia coli* DH5 α . *Int. Conf. Chem. Mater. Food Eng.* 22: 88–92.

Different Methods of Soil DNA Extraction

Asmita Kamble and Harinder Singh*

Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS Deemed to be University, Mumbai, India

*For correspondence: Harinder.Singh@nmims.edu

[Abstract] Soil is the major reservoir of microbial diversity. Only 1% of microbial diversity can be cultured while 99% is still not culturable. It is necessary to extract DNA from soil in order to explore the 99% microbial diversity, which will be useful to harness novel industrial enzymes and natural products. In the present study, six traditional and two kit-based methods were utilized to obtain total soil DNA from Garden soil. Quality (Absorbance ratio at A_{260}/A_{230} , A_{260}/A_{280} nm) of the extracted DNA was assessed and quantity was analyzed using the BioTek Epoch Microplate spectrophotometer. Quality of DNA is one of the important factors that should be taken in to account for downstream applications such as PCR or cloning experiments.

Keywords: Soil DNA, Non-culturable, DNA extraction, Garden soil

[Background] Soil is the largest terrestrial reservoir of microbial bio-diversity which significantly balance the critical cycle of carbon, nitrogen, and phosphorous; along with maintaining plant health, structure and fertility of the soil. 1 g of soil harbors 1,00,000-10,00,000 different bacterial and archaeal species (Satyanarayana, 2017). However, only 1% of the microbial communities can be cultured in the laboratory conditions while 99% are still unexplored because they are non-culturable. Total soil DNA extraction potentially onsets the journey of revealing hidden microbial diversity (Robe *et al.*, 2003; Fatima *et al.*, 2011 and 2014; Lamizadeh *et al.*, 2019).

Soil DNA extraction includes two major steps: microbial cell lysis followed by purification to get rid of inhibitory molecules of humic acid and fulvic acid. Cell lysis can be performed via physical, mechanical and chemical approaches; or a combination of all three methods. These methods involve the use of detergents such as Sodium Dodecyl Sulfate (SDS) along with heat treatment in buffers like Tris-HCl or sodium phosphate buffers, along with the introduction of chelating agents such Ethylenediaminetetraacetic acid (EDTA) to protect extracted DNA from DNases which are readily present in external environment. There are different components which add to the effectiveness of cell disruption such as enzymatic treatment with lysozyme, using strong chaotropic agents such as guanidium salts, physical and mechanical treatment such as using liquid nitrogen to grind soil samples, ultra-sonication, glass beads, Zirconia beads and bead-beating approach for cell lysis. Various combinations of these methods are utilized to improve the yield of isolated DNA. Purification steps are conducted after DNA extraction with the help of phenol:chloroform:isoamyl-alcohol method (PCI method), ethanol precipitation, precipitation via polyethylene glycol, isopropyl alcohol, and spin columns. Purification is a crucial step as it involves the removal of PCR and restriction digestion

inhibitors such as humic and fulvic acid which are co-extracted along with DNA. Hence, effective extraction procedure followed by stringent purification are crucial steps for isolating DNA which can be used for better understanding of microbial biodiversity (Fatima *et al.*, 2011; Bag *et al.*, 2016).

The present study was aimed to obtain a high yield of total soil DNA from garden soil by using different conventional methods of extraction as well as soil DNA kits (HiPura soil DNA kit and Dneasy power soil DNA kit). A comparative analysis was performed in terms of purity and yield using BioTek Epoch Microplate spectrophotometer for absorbance measurements.

Materials and Reagents

A. For soil collection and processing

1. Sterile 50 ml Falcon (Tarson, catalog number: 546041)
2. Sterile scapula
3. Garden soil
4. Ethanol
5. Liquid Nitrogen

B. For Extraction of total soil DNA by different methods

1. 0.1 mm Zirconia beads (BioSpec, catalog number: NC0362415)
2. 2.7 mm Glass beads (BioSpec, catalog number: 11079127)
3. 50 ml sterile Falcon (Tarson, catalog number: 546041)
4. 2 ml Eppendorf (Tarson, catalog number: 500020)
5. 1.5 ml Microcentrifuge tube (Tarson, catalog number: 5000010)
6. Filter 0.45 μm (Jsil)
7. 1,000 μl Microtips (Tarson, catalog number: 5210010)
8. 200 μl Microtips (Tarson, catalog number: 521020)
9. Disodium Hydrogen Phosphate anhydrous (Na_2HPO_4) (Molychem, catalog number: QB4Q640403)
10. Sodium Dodecyl Sulfate (SDS, $\text{CH}_3(\text{CH}_2)_{11}\text{OSO}_3\text{Na}$) (Affymetrix, catalog number: 18220)
11. Polyvinylpyrrolidone [PVPP, $(\text{C}_6\text{H}_9\text{NO})_n$] (Amresco, catalog number: 0507)
12. Sodium Chloride (NaCl , Fisher Scientific, catalog number: 27605)
13. Chloroform (CHCl_3) (Molychem, catalog number: 13620)
14. Sodium acetate trihydrate ($\text{CH}_3\text{COONa}\cdot 3\text{H}_2\text{O}$) (Fisher Scientific, catalog number: 1390M)
15. Tris (Hydroxymethyl)Aminomethane Hydrochloride (Tris-HCl, $\text{C}_4\text{H}_{11}\text{NO}_3\cdot\text{HCl}$) (SRL, catalog number: 99438)
16. Ethylenediaminetetraacetic acid (EDTA, Qualigens, catalog number: 12635)
17. Phenol:Chloroform:Iso-amylalcohol 25:24:1 (PCI) (Sigma, catalog number: P2069)
18. Sodium Dihydrogen orthophosphate Dihydrate ($\text{NaH}_2\text{PO}_4\cdot 2\text{H}_2\text{O}$) (Molychem, catalog number: 25700)

19. Cetyl trimethylammonium bromide (CTAB) (Molychem, catalog number: 22560)
20. Mannitol (HiMedia, catalog number: RM9914)
21. Lysozyme (Sigma, catalog number: 89833)
22. Proteinase K (Sigma, catalog number: P2308)
23. Sodium Lauroyl Sarcosine (Amersco, catalog number: 0719)
24. Isopropyl alcohol (Molychem, catalog number: 1750)
25. Iso-amyl alcohol (SRL, catalog number: 69931)
26. Guanidine Hydrochloride (Sigma, catalog number: G3272)
27. Glycine (MP Biomedicals, catalog number: 194825)
28. Sodium Hydroxide (Molychem, catalog number: 25800)
29. Polyethylene Glycol (6000) (LOBA, catalog number: Art.5312)
30. Hydrochloric acid (Molychem, catalog number: 23540)
31. RNase (Invitrogen PureLink Genomic DNA mini kit, catalog number: K1820-01)
32. Proteinase K (Invitrogen PureLink Genomic DNA mini kit, catalog number: K1820-01)
33. Spin Column (Invitrogen PureLink Genomic DNA mini kit, catalog number: K1820-01)
34. Wash buffer1 (Invitrogen PureLink Genomic DNA mini kit, catalog number: K1820-01)
35. Wash buffer2 (Invitrogen PureLink Genomic DNA mini kit, catalog number: K1820-01)
36. Elution buffer (Invitrogen PureLink Genomic DNA mini kit, catalog number: K1820-01)
37. Ethanol 99.9% pure (Changshu Hongsheng Fine Chemicals, catalog number: 1170)
38. Stock solution (see Recipes)
 - 1 M NaCl
 - 1 M NaH₂PO₄
 - 1 M Na₂HPO₄
 - 0.2 M Na₂HPO₄
 - 1 M Tris-HCl
 - 0.2 M EDTA
 - 1 M HCl
 - 50x TAE buffer
39. Working stock solution (see Recipes)
 - DNA extraction buffer 1
 - 50% PEG (6000)
 - 0.6 M NaCl
 - Chloroform:Isoamyl alcohol (24:1)
 - 3 M sodium acetate
 - T₁₀E₁ buffer
 - PCI
 - 0.1 M Phosphate buffer saline
 - DNA extraction buffer 2
 - 10% CTAB + 0.7 M NaCl

TEN buffer (pH 8.0)
 TEN buffers (1 ml) supplied with 0.2 mg Lysozyme
 20% SDS
 T₅₀E₁
 Lysozyme (10 mg/ml)
 Guanidine-HCl (5 M)
 10% Sodium Lauryl Sarcosine
 TENP
 Lysozyme solution
 SDS solution
 1x TAE buffer
 70% Ethanol

C. Commercial kits for soil DNA extraction

1. HiPurA soil DNA kit (HiMedia, catalog number: MB542)
2. Dneasy Power Soil (Qiagen, catalog number: 12888-50)

D. For agarose gel electrophoresis

1. Tris-Base (MP Biomedicals, catalog number: 103133)
2. EDTA (Qualigens, catalog number: 12635)
3. Glacial Acetic acid (Molychem, catalog number: 21020)
4. Ethidium bromide (Sigma-Aldrich, catalog number: E7637)
5. 1 kbp ladder (Generuler, catalog number: SM0313)
6. 6x gel Loading dye (Fremontas, catalog number: R0611)
7. Agarose (MP Biomedicals, catalog number: 218072090)

Equipment

1. Sieve (Mesh size of 0.22 mm, Sumeet Royal Interchangeable sieve)
2. Mortar and pestle (diameter of Mortar: 6 cm)
3. Pipettes 100-1,000 µl, 20-200 µl, 0.5-10 µl, 2-20 µl (Glison)
4. Water bath (Metalab, model: MSI14)
5. Centrifuge (Eppendorf, model: 5424R)
6. Micro-centrifuge (Pfact, model: 5804K)
7. Incubator Shaker (Hexatech, model: HIPL-035C)
8. Ice-machine (Wensar, model: LMIF series)
9. -20 °C refrigerator (Thermo)
10. Refrigerator (LG, model: GL-365YVQG5)
11. Sonicator (Cole-Parmer, model: 08895-22)

12. Weighing machine (Shimadzu, model: AUX220)
13. Horizontal gel electrophoresis apparatus (Techno Source, model: Sleek Gel)
14. Gel Documentation system (Bio-Rad, model: Universal Hood II)
15. Epoch Microplate spectrophotometer with Take3 plate accessories (BioTek, serial No.: 401554)
16. Microwave (LG microwave, model: MS-2347BS)

Software

1. Gen-5 Data Analysis Software (version 3.03)
2. Image Lab version 5.2.1 build11

Procedure

A. Soil sample Collection

1. Dig the soil surface up to 3 cm depth. Top soil is a precious natural resource as it contains humus (a rich source of nutrients), minerals required for growth of plants and microbial flora. Top soil is mainly rich in microbial diversity. **The guidelines for Soil sampling depth is given in Table 1.**

Table 1. The guidelines for sampling depth (Fery *et al.*, 2018) and Reference 13

Crop	Sampling depth (cm)
Grasses and grasslands	Up to 5 cm
Shallow rooted crops (e.g., rice, groundnut)	Up to 15 cm
Deep-rooted (e.g., Cotton, sugarcane)	Up to 22 cm

2. Collect three soil samples at different spots (at a distance of 3 m) from the same field.
3. Collect the soil in labeled sterile 50 ml Falcons with the help of sterile spatula.
4. Sieve the soil sample with the help of 0.22 mm mesh.
5. Mix the sieved soil sample.
6. Aliquot 10,000-15,000 mg of above-mixed soil in 50 ml sterile Falcons (maximum 15,000 mg, *i.e.*, one aliquot is enough to carry out all soil DNA extraction methods).
7. Store it in -20 °C for long term storage.

B. Soil sample processing

Two types of processing carried out for each method mentioned below: liquid nitrogen grinding and direct.

Note: Clean mortar and pestle with 100% ethanol.

1. Soil sample 1 (S1): Soil sample processed by using Liquid Nitrogen.
 - a. Weigh 10,000 mg of the soil sample.

- b. In mortar with pestle, grind the soil sample by carefully pouring liquid nitrogen. The treatment should be continued until it turns in to powder (label this soil sample as S1). The soil sample processed with liquid nitrogen will be used for all the conventional methods listed below. Since 1 g of soil harbor 10^{10} bacterial cells and 4×10^3 to 5×10^4 species diversity, each method requires 500 mg to 1 g of soil sample (Raynaud and Nunan, 2014).
2. Soil sample 2 (S2): Direct: Without liquid nitrogen processing. The soil sample processed without liquid nitrogen will be used for all the conventional and kit-based methods listed below:
 - a. Weigh 10,000 mg of the soil sample.
 - b. In mortar with pestle, grind it without liquid Nitrogen to make powder as fine as possible.

Notes:

1. The following methodology is carried out for both samples, i.e., S1 as well as S2, except for kit-based method where only S2 is used.
2. After extraction of DNA by different methods, immediately store it at $-20\text{ }^{\circ}\text{C}$.

C. Protocols for total soil DNA extraction

Figure 1 summarizes the above procedure and total soil DNA extraction methods 1 to 8.

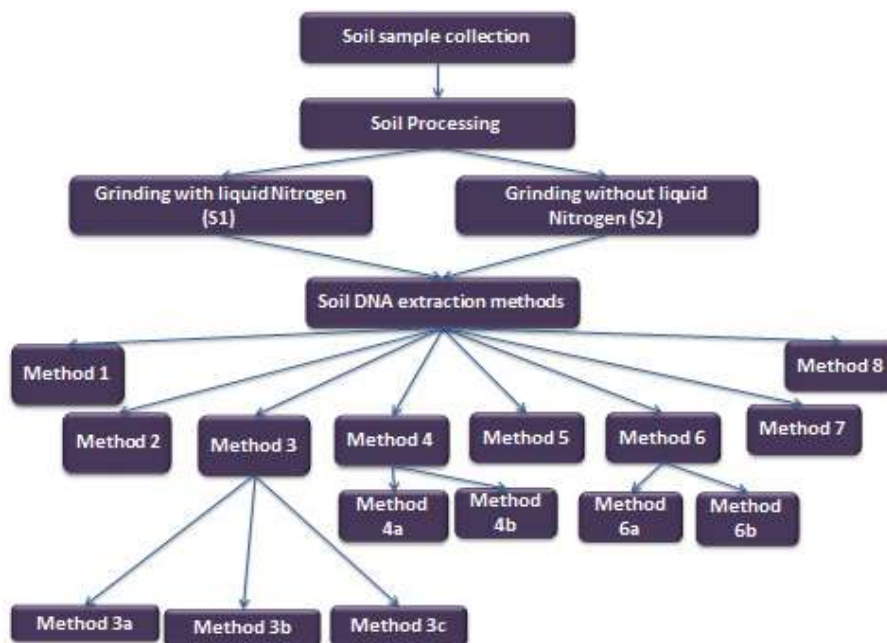


Figure 1. Workflow of soil DNA extraction protocol. The details of different steps and methods 1-8 are described under Methods section.

1. Method 1: PEG Method
 - a. Weigh 1,000 mg of the soil sample in a 50 ml sterile centrifuge tube and add 10 ml of DNA extraction buffer 1.
 - b. Incubate it at $65\text{ }^{\circ}\text{C}$ for 1 h. in a water bath.

- c. Centrifuge it at 7,000 x g 10 min at 4 °C.
 - d. Collect the supernatant in another centrifuge tube.
 - e. Add 1/2 volume 50% PEG (6000) and 1 volume 0.6 M NaCl (e.g., if the volume of supernatant collected is 5 ml, then add 2.5 ml of 50% PEG (6000) and 5 ml of 0.6 M NaCl). Mix the solution gently by inverting the tubes 3-4 times.
 - f. Add 1 volume of Chloroform:isoamylalcohol (24:1). Mix the solution gently by inverting the tubes 3-4 times.
 - g. Centrifuge at 13,000 x g 10 min at 4 °C.
 - h. Collect the supernatant.
 - i. Add 1/10th volume of 3 M sodium acetate (pH 5.2).
 - j. Add 2 volumes of ice-cold ethanol (100%). Mix the solution gently by inverting tubes 3-4 times. Incubate overnight at -20 °C, which results in high yield of DNA.
 - k. Centrifuge at 13,000 x g 10 min at 4 °C.
 - l. Remove the ethanol and air dry the pellet for 15-20 min.
 - m. Dissolve the pellet in 100 µl of T₁₀E₁ buffer.
2. Method 2: Phenol:Chloroform:Isoamylalcohol (PCI) method
- a. Follow Steps C1a-C1d of the Method 1.
 - b. Add equal volume of Phenol:Chloroform:Isoamylalcohol (25:24:1) to the supernatant. Mix the solution gently by inverting the tubes 3-4 times.
 - c. Centrifuge 13,000 x g 10 min at 4 °C.
 - d. Collect the aqueous fraction.
 - e. Add 2 volumes of ice-cold ethanol and 1/10th volume of 3 M sodium acetate (pH 5.2). Mix the solution gently by inverting tubes 3-4 times. Incubate overnight at -20 °C.
 - f. Centrifuge at 13,000 x g 10 min at 4 °C.
 - g. Remove the supernatant.
 - h. Air dry the pellet up to 15-20 min to remove ethanol.
Note: Ethanol can hinder in downstream applications. Hence it is necessary to remove it completely. Be careful not to dry pellet completely.
 - i. Dissolve the pellet in 100 µl of T₁₀E₁ buffer.
3. Method 3
- Note: In this method there are three sub-methods naming it as 3a, 3b and 3c. However, steps 'C3a to C3g' are common to all three methods.*
- a. Weigh 1,000 mg of the soil sample in a 50 ml sterile centrifuge tube and add 5 ml of 120 mM Phosphate Buffer Saline (PBS, pH 7.4).
 - b. Shake the mixture at 150 rpm 10 min at 4 °C in Hexatech shaker.
 - c. Centrifuge at 7,000 x g for 10 min at 4 °C.
 - d. Discard the supernatant and wash the pellet with PBS.
 - e. Again, centrifuge at 7,000 x g for 10 min at 4 °C.
 - f. Discard the supernatant and suspend the pellet in 10 ml DNA extraction buffer 2.

- g. Incubate at 65 °C for 1 h.
- i. Method 3a: Mannitol-PBS-PEG/NaCl
 - 1) Follow Steps C3a to C3g.
 - 2) Centrifuge the soil suspension at 8,000 x g 10 min at 4 °C.
 - 3) Collect the supernatant.
 - 4) Add 1/2 volume PEG (50%) and 1 volume of 0.6 M NaCl.
 - 5) Incubate at 4 °C overnight.
 - 6) Centrifuge at 13,000 x g 4 °C 10 min.
 - 7) Discard the supernatant and resuspend the pellet in 3 ml TE buffer.
 - 8) Proceed with Steps C2b to C2i.
 - ii. Method 3b: Mannitol-PBS-PCI
 - 1) Follow Steps C3a to C3g.
 - 2) Centrifuge the soil suspension at 7000 x g 10 min at 4 °C.
 - 3) Collect the supernatant.
 - 4) Proceed with Steps C2b to C2i.
 - iii. Method 3c: Mannitol-PBS-CTAB
 - 1) Follow Steps C3a to C3g.
 - 2) Centrifuge the soil suspension at 7,000 x g 10 min at 4 °C.
 - 3) Collect the supernatant.
 - 4) Add 50 µl 1 M NaCl and 50 µl 10% CTAB (prepared in 0.6 M NaCl).
 - 5) Incubate for 15 min at 4 °C.
 - 6) Proceed with Steps C2b to C2i.
4. Method 4
- Note: There are 2 sub-methods of method 4: 4a and 4b. Steps 'C4a-C4r' are common for both methods.*
- a. Weigh 500 mg of soil and add 1 ml of TEN buffer (Tris-Cl-EDTA-NaCl buffer).
 - b. Vortex for 1 min.
 - c. Centrifuge at 10,000 x g 10 min RT.
 - d. Decant the supernatant and wash the pellet with 1 ml TEN.
 - e. Centrifuge at 10,000 x g 10 min RT.
 - f. Discard the supernatant. Resuspend the pellet in 1 ml TEN (add 0.2 mg lysozyme) (see Recipe: TEN buffers supplied with 0.2 mg Lysozyme).
 - g. Incubate for 1 h at 37 °C.
 - h. Keep it on ice for 10 min and then 20 min at 65 °C.
 - i. Add 100 µl of 20% SDS.
 - j. Vortex for 1 min.
 - k. Incubate for 30 min **at RT**.
 - l. Centrifuge at 10,000 x g 10 min at RT.
 - m. Collect the supernatant.

- n. Add 500 μ l of 3 M sodium acetate.
- o. Incubate for 5 min **at 65 °C**.
- p. Incubate on ice for 20 min.
- q. Centrifuge at maximum '14,000 x g' for 30 min at 4 °C.
- r. Collect the supernatant
 - i. Method 4a
 - 1) Follow Steps C4a-C4r.
 - 2) Proceed with C2b to C2i.
 - ii. Method 4b
 - 1) Follow Steps C4a-C4r.
 - 2) Load the supernatant on Invitrogen PureLink Genomic DNA mini kit spin column.
 - 3) Centrifuge at 16,000 x g 2 min RT.
 - 4) Discard the flow-through.
 - 5) Add 500 μ l of wash buffer1 into the column.
 - 6) Centrifuge at 16,000 x g 1 min RT.
 - 7) Discard the flow-through.
 - 8) Add 500 μ l of wash buffer2.
 - 9) Centrifuge at 16,000 x g 1 min RT.
 - 10) Discard the flow-through.
 - 11) Centrifuge at 18,407 x g 3 min RT.
 - 12) Discard the collection tube.
 - 13) Place column in new sterile 1.5 ml Eppendorf.
 - 14) Add 50 μ l of elution buffer.
 - 15) Incubate for 2 min at RT.
 - 16) Centrifuge at 18000 x g 2 min at RT.
 - 17) Store at -20 °C.
- 5. Method 5
 - a. Weigh 1,000 mg of soil sample and add 400 μ l of 50 mM Tris-HCl and 1 mM EDTA.
 - b. To the above mixture add 4 sterile glass beads (2.5 mm).
 - c. Vortex for 1 min until it is homogenized.
 - d. Remove glass beads with sterile tweezer.
 - e. Add 50 μ l lysozyme (10 mg/ml) (see Recipe: Lysozyme (10 mg/ml))
 - f. Incubate it for 1.5 h. at 37 °C.
 - g. Add 250 μ l of Guanidine Hydrochloride (4 M).
 - h. Mix gently for 45 s.
 - i. Add 300 μ l Sodium Lauryl Sarcosine.
 - j. Vortex for 10 min at 37 °C.
 - k. Incubate **for 1 h at 70 °C**.
 - l. Add around 300 mg 0.1 mm zirconia beads.

- m. Vortex for 20 min.
 - n. Add 15 mg PVPP.
 - o. Gently vortex it.
 - p. Centrifuge at 14,000 x g 4 °C for 5 min.
 - q. Transfer the supernatant in another tube.
 - r. Wash the pellet with 200 µl TENP. Centrifuge at 14,000 x g 4 °C for 5 min. Collect the supernatant carefully.
 - s. Pool supernatant of Steps C5q-C5r.
 - t. Add 2 volumes of 100% ethanol.
 - u. Invert the tubes
 - v. Incubate at RT for 5 min.
 - w. **Centrifuge at 14,000 x g 4 °C for 5 min.**
 - x. Remove the supernatant.
 - y. Air dry for 15 min.
 - z. Resuspend the pellet with 450 µl of PB supplemented with 50 µl 3 M sodium acetate.
 - aa. Incubate at 4 °C for 1 h.
 - bb. Add 2 µl **each** of RNase and Proteinase K.
 - cc. Incubate **for 30 min** at 37 °C.
 - dd. Add 50 µl of sodium acetate (3 M) and 1 ml ethanol.
 - ee. Centrifuge at 14,000 x g 10 min 4 °C.
 - ff. Remove supernatant and wash pellet with 70% ice-cold ethanol.
 - gg. Air dry the pellet.
 - hh. Resuspend pellet in 100 µl of TE buffer.
6. Method 6
- Note: There are 2 sub-methods of method 6: 6a and 6b. Steps 'C6a-C6n' are common for both methods.*
- a. Weigh 1,000 mg of soil.
 - b. Add 100 mM sodium phosphate buffer (pH 8).
 - c. Add 1 ml of lysozyme solution (see Recipes section below: Lysozyme solution).
 - d. Incubate at 37 °C for 2 h.
 - e. Add 1 ml of SDS solution (see Recipes section below: SDS solution: 10% SDS + 1 M NaCl + Tris-HCl).
 - f. Mix the above mixture well.
 - g. Incubate it at RT for 5 min.
 - h. Incubate it on ice for 2 min.
 - i. Keep it at 65 °C for 20 min
 - j. Again on ice for 2 min.
 - k. Add 1 ml of 3 M sodium acetate.
 - l. Vortex vigorously.

- m. Centrifuge at 6,000 x g for 10min.
 - n. Collect the supernatant.
 - i. Method 6a
 - 1) Follow Steps C6a-C6n.
 - 2) Proceed with C2b to C2i.
 - ii. Method 6b
 - 1) Follow Steps C6a-C6n.
 - 2) Add equal volume of ethanol.
 - 3) Centrifuge at 14,000 x g 10 min at 4 °C.
 - 4) Remove supernatant.
 - 5) Air dry the pellet.
 - 6) Resuspend the pellet in 100 µl TE buffer.
7. Method 7: HiPura soil DNA kit
- Weigh 250 mg* of soil and follow manufacturer's instructions given in the manual. The procedure is briefly explained here. The steps and its importance of soil DNA extraction by using HiPurA soil DNA kit are also listed below.
- *Note: If it is the sediment sample, then weigh approximately 500 mg. Yield of total soil DNA depends upon the sample type and number of micro-organisms present in the sample.*
- a. Cell lysis

Bead beating and soil lysis solution (SL), along with vigorous vortexing for 10 min, can help in lysing the microbial cells. Separation of the soil particles from lysed microbial cells is done by centrifugation at 13,000 x g for 1 min. Collect the supernatant, and if there are still few soil particles in the collected supernatant, again centrifuge it at 13,000 x g for 1 min and collect the supernatant.
 - b. Removal of inhibitors

Add Inhibitor Removal Solution (IRSH) to remove inhibitors such as humic acid. Collect the supernatant; which has nucleic acid, proteins, and other lysed cellular components; by performing a round of centrifugation at 10,000 x g for 1min. Supernatant should be transparent (brownish color of supernatant indicates high humic acid content) and free of soil particles, otherwise it will hinder in purification steps by clogging the membrane.
 - c. Purification and elution

Apply the collected supernatant on to spin column to purify total genomic DNA with the aid of silica membrane (present in the spin column format) by using binding, washing and elution buffers. Binding buffer provides optimum pH environment for nucleic acids to bind on to silica membrane, washing buffers helps to remove proteins and other contaminants, and elution buffer efficiently elute total soil DNA.
8. Method 8: Dneasy Power Soil
- Weigh 250 mg of soil and follow manufacturer instructions given in the manual. Briefly the steps and its importance of soil DNA extraction by using Dneasy Power Soil is given below.

a. Cell lysis

Total soil DNA is extracted by lysis of microbial cells via cell lytic buffer and bead beating method. Hence mechanical and chemical methods are used to obtain efficient cell lysis of microbial cells. Separation of the soil particles from lysed microbial cells is done by centrifugation at 10,000 x g for 30 s. Collect the supernatant, and if there are still few soil particles in the collected supernatant, again centrifuge it at 10,000 x g for 1 min and collect the supernatant.

b. Removal of inhibitors

Add inhibitor removal solution to remove inhibitors present. Solution should be transparent. Brownish-yellow color solution is a visual indication of the presence of humic acid. Hence solution should be preferably transparent and free of soil particles otherwise it will lead to inefficient extraction of DNA.

c. Purification and Elution of genomic DNA

It is performed by using silica-coated spin column by following number of steps which involves addition of binding buffer in order to bind nucleic acid to the membrane, washing buffers to remove other contaminants such as proteins and elution buffer for eluting total genomic DNA.

D. Visualization of extracted total soil DNA

Prepare 0.8% gel

1. Weigh 800 mg of Agarose.
2. Dissolve it in 100 ml of 1x TAE.
3. Microwave it for 2 min till it is completely dissolved.
4. Let it be warm and add 4 µl of EtBr (10 mg/ml).
5. Pour it in a casting tray and allow it to solidify.
6. Take 5 µl of the extracted DNA + 2 µl 6x gel loading dye. Mix it well.
7. Load it on the gel.
8. Load 1 kbp ladder and lambda DNA HindIII digest.
9. Let it resolved in 1x TAE at 100 volts for 45 min.
10. Visualize it in Gel Documentation system and Image Lab software. The gel images of total soil DNA that was extracted by using Methods 1 to 3, 4 to 6, 7 and 8 are shown in Figures 2, 3, 4, and 5 respectively.

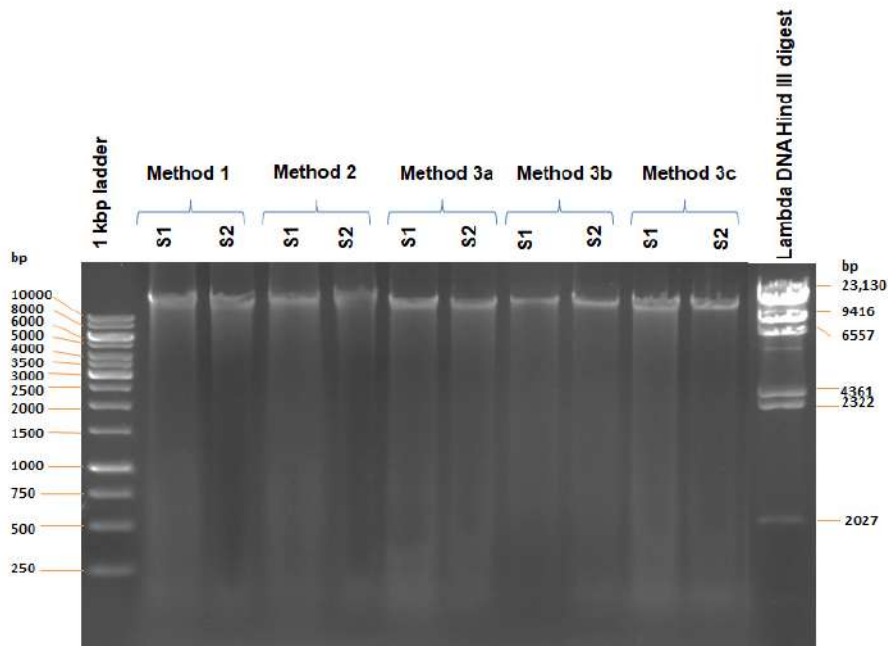


Figure 2. Gel electrophoresis of total soil DNA. The total soil DNA extracted by using methods 1 to 3 were resolved by 0.8% agarose gel electrophoresis and visualized with the help of Gel Documentation system and Image Lab software. S1: sample processed with liquid nitrogen, S2: sample processed without liquid nitrogen.

Note: High molecular weight DNA (approximately 23 kbp) was extracted using Methods 1 to 3.

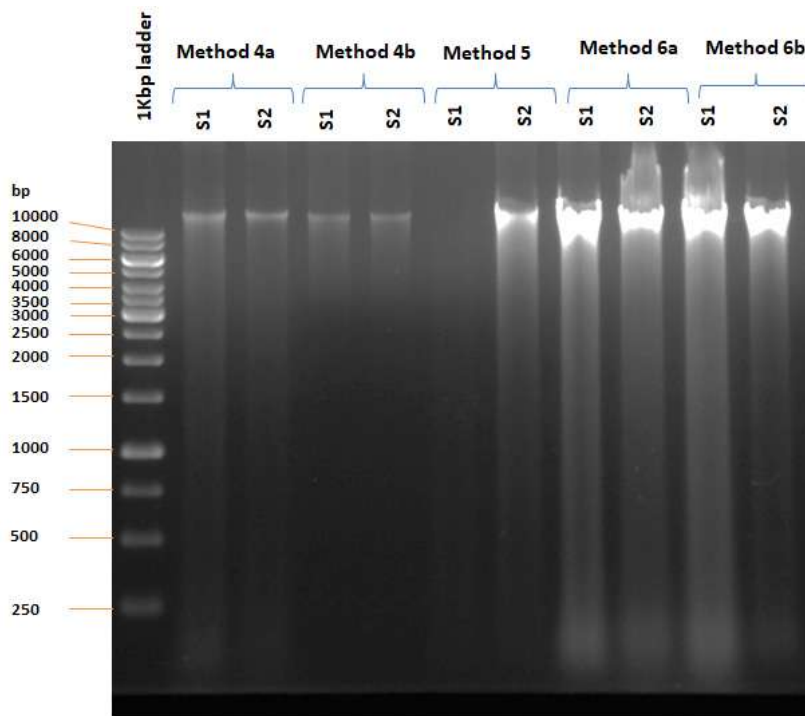


Figure 3. Gel electrophoresis of total soil DNA. The total soil DNA extracted by using methods 4 to 6 were resolved by 0.8% agarose gel electrophoresis and visualized with the help

of Gel Documentation system and Image Lab software. S1: sample processed with liquid nitrogen, S2: sample processed without liquid nitrogen.

Note: High molecular weight DNA (> 10 kbp) was extracted using Methods 4 to 6.

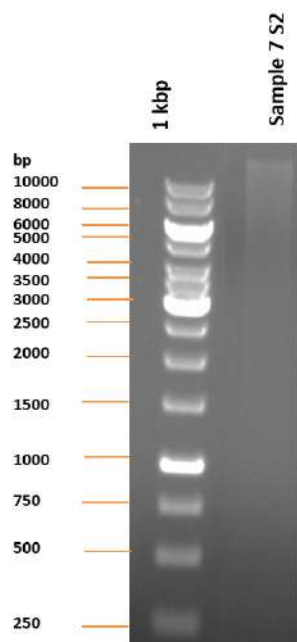


Figure 4. Gel electrophoresis of total soil DNA. The total soil DNA extracted by using method 7 were resolved by 0.8% agarose gel electrophoresis and visualized with the help of Gel Documentation system and Image Lab software. S1: sample processed with liquid nitrogen, S2: sample processed without liquid nitrogen.

Note: High molecular weight DNA (> 10 kbp) was extracted using Method 7.

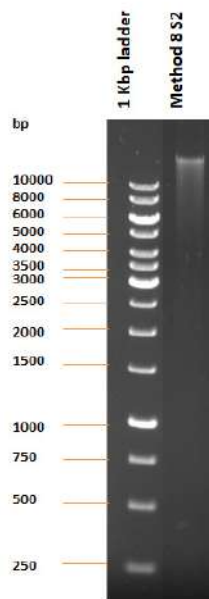


Figure 5. Gel electrophoresis of total soil DNA. The total soil DNA extracted by using method 8 were resolved by 0.8% agarose gel electrophoresis and visualized with the help of Gel Documentation system and Image Lab software. S1: sample processed with liquid nitrogen, S2: sample processed without liquid nitrogen.

Note: High molecular weight DNA (> 10 kbp) was extracted using Method 8.

E. BioTek Epoch Microplate spectrophotometer

1. Clean Take-3 plate reader with ethanol.
2. Dilute the samples if necessary (1:100 or 1:200 dilution) by using elution buffer as diluent.
3. Blank should be elution buffers used for different extraction procedures.
4. Add 2 μ l on the plate and read it with the help of BioTek ELISA plate and support by Gen 5 Software for providing user-friendly interface for analysis. The quantification of total soil DNA is collectively shown in Table 2 and Figure 6.

Table 2. Quantification of total soil DNA using BioTek Epoch Microplate spectrophotometer

Sample	ng/ μ l	A_{260}/A_{280}	A_{260}/A_{230}
Method 1 S1	9217.8	1.438	0.534
Method 1 S2	6380.6	1.487	0.482
Method 2 S1	4277.8	1.478	0.577
Method 2 S2	6003.4	1.46	0.616
Method 3a S1	3704.6	1.534	1.174
Method 3a S2	5767.7	1.458	1.416
Method 3b S1	6576.6	1.516	0.711
Method 3b S2	2814	1.515	0.627
Method 3c S1	4298	1.472	0.626
Method 3c S2	4601.7	1.507	0.677
Method 4a S1	856.3	1.19	0.689
Method 4a S2	681.2	1.442	0.745
Method 4b S1	34.1	2.33	0.048
Method 4b S2	14.8	3	0.023
Method 5a S1	104	2.33	0.184
Method 5a S2	563.8	1.468	0.649
Method 6a S1	616.9	1.732	0.691
Method 6a S2	1825.8	1.604	1.433
Method 6b S1	1567.5	1.699	1.427
Method 6b S2	453.1	1.561	1.391
Method 7 S2	11.045	1.851	1.052
Method 8 S2	19.608	1.87	1.789

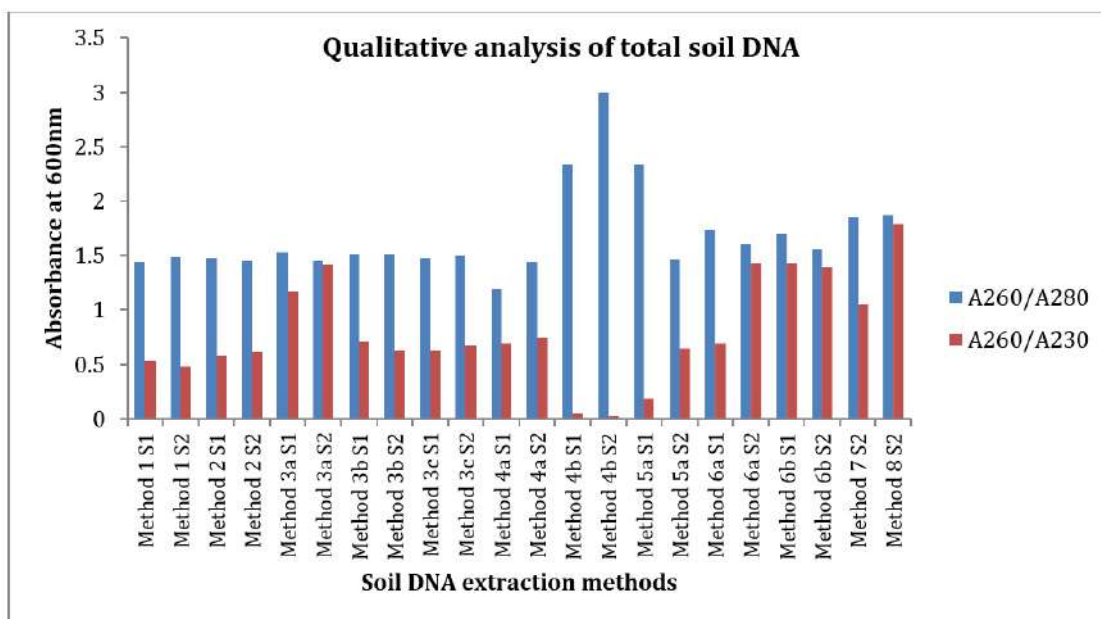


Figure 6. Qualitative analysis of total soil DNA: A ratio of A_{260}/A_{280} and A_{260}/A_{230} indicates the quality of extracted DNA. If the ratio of A_{260}/A_{280} is in between 1.8 and 2, it indicates high quality DNA. If it is less than 1.8, it indicates protein or other aromatic compounds contamination. If the ratio of A_{260}/A_{230} is less than 2 it indicates the presence of organic components such as humic acid. Humic acid present in soil samples can co-extract with the total soil DNA during the extraction procedure since both have the same charge and characteristics. Humic acid has the ability to interfere in downstream applications such as PCR (Humic acid binds to DNA polymerase and interfere in its activity) (Fatima *et al.*, 2014)

Conclusion: In the present study, conventional as well commercial kit based methods were used to extract total genomic DNA from soil. The quality and quantity of the extracted total DNA were visualized and analyzed by using the BioTek Epoch Microplate spectrophotometer (A_{260}/A_{280} and A_{260}/A_{230} ratio) and Gel electrophoresis technique. Commercial kit-based methods aid to obtain high quality of DNA as compare to conventional methods. Soil DNA extraction is the first step to explore culturable (accounts for 1% of microbial communities) as well as non-culturable microbial diversity (accounts for rest 99% of microbial communities) in a given soil sample. The DNA extracted by above-mentioned protocols can be used to perform soil metagenomics studies in order to discover novel biomolecules such as novel metabolites, enzymes, antibiotics (Kapoor *et al.*, 2015; Farias *et al.*, 2018; Castillo Villamizar *et al.*, 2019). It can also be used to study microbial diversity in different soil samples using 16S rRNA gene metagenomics. This will expand the existing microbial databases and also help to understand the effect of various environmental factors on soil microbial population and diversity (Leite *et al.*, 2014; Liu *et al.*, 2019; Matsushita *et al.*, 2019).

Recipes

A. Stock solution

1. 1 M NaCl
 - a. Add 2,922 mg of NaCl in 40 ml of MilliQ-water and then make up the volume to 50 ml with MilliQ-water
 - b. Store at RT
2. 1 M NaH₂PO₄
 - a. Weigh 7,800.5 mg in 40 ml of MilliQ-water. Mix well. Make up the volume to 50 ml with MilliQ-water
 - b. Store at RT
3. 1 M Na₂HPO₄
 - a. Add 7,080 mg of Na₂HPO₄ in 40 ml of MilliQ-water. Mix well. Make up the volume to 50 ml with MilliQ-water
 - b. Store at RT
4. 0.2 M Na₂HPO₄
 - a. Add 4 ml of 1 M Na₂HPO₄ in 16 ml of MilliQ-water
 - b. Store at RT
5. 1 M Tris-HCl
 - a. Add 7,880 mg of Tris-HCl in 40 ml of MilliQ-water
 - b. Adjust the pH to 8
 - c. Make the volume to 50 ml with MilliQ-water
 - d. Store at RT
6. 0.2 M EDTA
 - a. Add 3,725 mg of EDTA in 30 ml of MilliQ-Water. Keep this mixture on the magnetic stirrer
 - b. Add few pellets of NaOH and let it stir until EDTA completely dissolves. Check the pH of the solution. It should be pH 8 since EDTA dissolves at pH 8
 - c. Store at RT
7. 1 M HCl (1 L)

Add 83 ml of concentrated HCl and adjust the volume to 1 L with MilliQ-water
8. 50x TAE buffer

2 M Tris-Base
 1 M Glacial acetic acid
 50 mM EDTA

For total volume 1 L:

 - a. Add 242,000 mg of Tris-Base
 - b. Dissolve it in 750 ml of deionized water
 - c. Carefully add 57.1 ml of Glacial Acetic Acid
 - d. Add 100 ml of 0.5 M EDTA (pH 8)

- e. Adjust the solution to final volume of 1 L
- f. Store at RT

B. Working stock solution

Note: Filter all following preparations with 0.45 µm filter.

1. DNA extraction buffer 1

Final concentration: 120 mM Na₂HPO₄ + 5% SDS (w/v) + 0.1 g PVPP)

For 50 ml total volume follow the recipe given below:

- a. Add 6 ml of 1,000 mM Na₂HPO₄ + 0.1 mg PVPP. Mix it well
- b. Then add 2,500 mg of SDS. Mix it well and keep it at 50 °C to completely dissolve
- c. Let the froth settle down
- d. Make the volume to 50 ml with MilliQ-water

2. 50% PEG (6000)

For 10 ml total volume

- a. Add 5,000 mg of PEG 6000 in 7 ml of MilliQ-water. Dissolve it properly. Keep it at 50-60 °C water bath until dissolve
- b. Make the volume to 10 ml with MilliQ-water

3. 0.6 M NaCl

For total volume 25 ml

- a. Add 15 ml of the 1 M NaCl in 7 ml of MilliQ-H₂O
- b. Dissolve it by vortexing
- c. Make up the volume to 25 ml by using MilliQ-water

4. Chloroform:Isoamyl alcohol (24:1)

For total volume 25 ml

- a. Add 24 ml of Chloroform in 1 ml of isoamylalcohol
- b. Mix it well
- c. Store at 4 °C

5. 3 M sodium acetate

Total volume 50 ml

- a. Add 12,304.5 mg of Sodium acetate in 30 ml of MilliQ-Water
- b. Adjust the pH to 5.2
- c. Make up the volume to 50 ml with MilliQ-water

6. T₁₀E₁ buffer

Total volume 50 ml

- a. Add 0.5 ml of 1 M Tris-HCl + 0.25 ml of 0.2 M EDTA + 49.25 ml of MilliQ-water
- b. Mix it well and store at 4 °C

7. PCI

- a. Add Equilibration buffer in to the content as per manufacturer's instructions
- b. pH of the phenolic phase is between 7.8-8.2

- c. Mix it well and store at 4 °C
8. 0.1 M Phosphate buffer saline
 - 0.02 M NaH₂PO₄
 - 0.08 M Na₂HPO₄
 - 9% NaCl

For 50 ml total volume:

 - a. Add 5 ml of 1 M NaH₂PO₄ +20 ml of 0.2 M Na₂HPO₄ + 4,500 mg of NaCl
 - b. Adjust the pH 7.4
 - c. Make up the volume to 50 ml with MilliQ-water
9. DNA extraction buffer 2

For 10 ml DNA extraction buffer-2:

 - 0.1 M Tris-HCl (pH 8.0)
 - 0.2 M EDTA (pH 8.0)
 - 10% SDS
 - 0.2 M Mannitol
 - 1 M NaCl
 - 2% CTAB

Follow the steps given below:

 - a. Add 2 ml of 1 M Tris-HCl+ 1 ml of 0.2 M EDTA+ 1 mg SDS
 - b. Vortex to mix it
 - c. Keep it at 60 °C to dissolve the mixture completely
 - d. Let the foam settle down
 - e. Add 0.3643 mg mannitol. Dissolve it properly
 - f. Add 0.2 mg of CTAB. Vortex it vigorously. Again let the foam settles down
 - g. Keep it at 60 °C until dissolve
 - h. Add 0.5 g of NaCl. Vortex. Again keep it at 60 °C to dissolve completely
 - i. Filter the solution through 0.45 µm filter
10. 10% CTAB + 0.7 M NaCl

For total volume 25 ml:

 - a. Pre-warm 20 ml of MilliQ-Water
 - b. Add 2.5 mg of CTAB + 1.0227 mg NaCl in pre-warm water
 - c. Make up the volume to 25 ml with MilliQ-water
11. TEN buffer (pH 8.0)
 - 100 mM Tris-HCl
 - 50 mM EDTA
 - 500 mM NaCl

Add 5 ml of 1 M Tris-HCl + 12.5 ml of 0.2 M EDTA + 25 ml of 1 M NaCl + 42.5 ml of MilliQ-water
12. TEN buffers supplied with 0.2 mg Lysozyme

For total volume 1 ml

- a. 1 ml of TEN buffer + 0.2 mg of Lysozyme
 - b. Vortex it to dissolve
 - c. Keep it completely in water-sonicator
13. 20% SDS
- For 50 ml total volume
- a. Weigh 10,000 mg of SDS. Add it in 45 ml of MilliQ-water
 - b. Vortex it
 - c. Keep it at 40 °C in water bath until dissolve
14. T₅₀E₁
- For total volume 5 ml
- a. Add 0.25 ml of 1 M Tris-HCl + 0.025 ml of 0.2 M EDTA + 4.725 MilliQ-water
 - b. Store at RT
15. Lysozyme (10 mg/ml)
- a. Weigh 10 mg of Lysozyme in 1 ml of MilliQ-water
 - b. Vortex it until dissolve
16. Guanidine-HCl (5 M)
- For total volume 25 ml
- a. Weigh 11,941 mg of Guanidine-HCl in 20 ml of MilliQ-water. Vortex it
 - b. Make up the volume to 25 ml with MilliQ-water
17. 10% Sodium Lauryl Sarcosine
- a. Weigh 5,000 mg of Sodium Lauryl Sarcosine
 - b. Add 45 ml of MilliQ-water
 - c. Make up the volume to 50 ml with MilliQ-water
18. TENP
- 50 mM Tris-HCl
- 20 mM EDTA
- 100 mM NaCl
- 1% PVPP
- Add 2.5 ml of 1 M Tris-HCl + 5 ml of 0.2 M EDTA + 5 ml of 1 M NaCl
19. Lysozyme solution
- 150 mM Tris-HCl
- 100 mM EDTA
- Lysozyme (15 mg/ml)
- For total volume 25 ml:
- a. Add 3.75 ml of 1M Tris-HCl + 12.5 ml of 0.2 M EDTA + 375 mg of lysozyme
 - b. Make up the volume to 25 ml with MilliQ-water
20. SDS solution
- a. 1 ml of 1 M NaCl + 5 ml Tris-HCl + 1 g SDS (10%). Mix well
 - b. Let the foam settle down

- c. Make up the volume to 10 ml with MilliQ-water
21. 1x TAE buffer (1 L)
Add 20 ml of 50x TAE + 980 ml of MilliQ-water
22. 70% Ethanol
 - a. Add 70 ml of 100% ethanol + 30 ml of MiliQ-water
 - b. Store at 0 °C

Acknowledgments

The present study was supported by SVKM's NMIMS intramural seed grant. We would like to acknowledge research papers authored by Fatima *et al.*, 2011 and 2014; Bag *et al.*, 2016.

Competing interests

The authors declare no conflict of interest.

References

1. Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., Pant, A., Kumar, P., Saxena, S., Allin, K. H., Hansen, T., Arumugam, M., Vestergaard, H., Pedersen, O., Pereira, V., Abraham, P., Tripathi, R., Wadhwa, N., Bhatnagar, S., Prakash, V. G., Radha, V., Anjana, R. M., Mohan, V., Takeda, K., Kurakawa, T., Nair, G. B. and Das, B. (2016). [An improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples](#). *Sci Rep* 6: 26775.
2. Castillo Villamizar, G. A., Nacke, H., Boehning, M., Herz, K. and Daniel, R. (2019). [Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases](#). *MBio* 10(1): 10(1):1-15.
3. Fatima, F., Chaudhary, I., Ali, J., Rastogi, S., Pathak, N. (2011). [Microbial DNA extraction from soil by different methods and its PCR amplification](#). *Biochem.cell.Arch.*
4. Farias, N., Almeida, I. and Meneses, C. (2018). [New bacterial phytase through metagenomic prospecting](#). *Molecules* 23(2):1-14.
5. Fatima, F., Pathak, N. and Rastogi Verma, S. (2014). [An improved method for soil DNA extraction to study the microbial assortment within rhizospheric region](#). *Mol Biol Int* 2014: 518960.
6. Kapoor, S., Rafiq, A. and Sharma, S. (2017). [Protein engineering and its applications in food industry](#). *Crit Rev Food Sci Nutr* 57(11): 2321-2329.
7. Lamizadeh, E., Enayatizamir, N. and Motamedi, H. (2019). [Difference in some biological properties of saline and non-saline soil under sugarcane cultivation](#). *Soil Biology* 52(6): 690-695.

8. Leite, D. C., Balieiro, F. C., Pires, C.,A., Madari, B. E., Rosado, A. S., Coutinho, H. L. and Peixoto, R. S. (2014). [Comparison of DNA extraction protocols for microbial communities from soil treated with biochar](#). *Brazilian J Microbiol* 45(1):175-183.
9. Liu, M., Huang, H., Bao, S. and Tong, Y. (2019). [Microbial community structure of soils in Bamenwan mangrove wetland](#). *Sci Rep* 9(1): 8406.
10. Matsushita, Y., Egami, K. Sawada, A., Saito, M., Sano, T., Tsushima, S. and Yoshida, S. (2019) [Analyses of soil bacterial community diversity in naturally and conventionally farmed apple orchards using 16S rRNA gene sequencing](#). *Applied Soil Ecology* 141: 26-29.
11. Fery, M., Choate. J. and Murphy, E. (2018). [A guide to collecting soil samples for farms and gardens](#). EC 628.
12. Raynaud, X. and Nunan, N. (2014). [Spatial ecology of bacteria at the microscale in soil](#). *PLoS One* 9(1): e87217.
13. *Resource Management :: Soil :: Soli sampling Procedure*. Available at: http://agritech.tnau.ac.in/agriculture/agri_soil_sampling.html (Accessed: 21 November 2019).
14. Robe, P., Nalin, R., Capellano, C., Vogel, T. M. and Simonet, P. (2003). [Extraction of DNA from soil](#). *European Journal of Soil Biology* 39(4): 183-190.
15. Satyanarayana, S. D. V., Krishna, M. S. R. and Kumar, P. P. (2017). [Optimization of high-yielding protocol for DNA extraction from the forest rhizosphere microbes](#). *3 Biotech* 7(2): 91.

16S Ribosomal RNA Gene-Based Metagenomics: A Review

Asmita Kamble*, Shriya Sawant*, Harinder Singh

Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS Deemed to be University, Vile Parle (W), Mumbai, India

*The first two authors are having equal contributions

Abstract

With the advent of contemporary molecular tools, the conventional microbiological isolation, enrichment techniques, and approaches have changed considerably. Molecular techniques such as polymerase chain reaction, cloning, and sequencing have shown that the major percentage of microbial diversity in an ecosystem remain “unculturable” or “as yet uncultivable” due to the lack of information on their biology, limited selection media, and culture conditions that could support their growth. Identifying and knowing more about them have become an important objective in the microbiological research. The ecological, environmental, and functional implications of a microbial ecosystem can be deciphered by knowing its microbial composition and interactions. The areas of whole-cell and targeted gene metagenomics are playing a key role in accomplishing this objective. The present review discusses the 16S ribosomal RNA (16S rRNA) gene metagenomics approach, which has found major applications in identifying the composition of a given microbial ecosystem. Different systems, processes, and analysis tools are available to perform 16S rRNA metagenomics; however, there are few concerns that require more investigation to gain the maximum benefit of these techniques.

Keywords: 16S ribosomal RNA, metagenomics, microbiome, next-generation sequencing

INTRODUCTION

Microbial research has seen a major revolution in the past 25 years, especially after the introduction of contemporary molecular techniques in the last decade. The current microbial culturing method on standard media replicates the essential aspects such as pH, temperature, nutrients, and osmotic conditions that could only support the growth of small fraction of total microbial diversity, while the majority of it remain unculturable.^[1] Next-generation sequencing data have evidently shown that the humongous size of the uncultured microbial world remains unexplored in conventional culturing techniques.^[2] This is mainly due to the limitation of conventional microbiological isolation and enrichment techniques that are not capable of supporting the growth of all the microbes present in a sample, under laboratory conditions. Because of this drawback, the traditional microbiological approach of isolating and characterizing novel microbes from an environmental source has taken a back seat. Researchers are more interested in capturing and profiling the complete microbial diversity present in a given sample, rather than a small percentage of it which can be done by conventional methods. Along with the progress in the tools and techniques, the sample source has also expanded hugely, samples are now

being explored from a range of natural or artificial sources such as agricultural or environmentally relevant soil, aquatic habitat, flora present on or inside other organisms, like domestic animals or human body.^[3-5] The increased interest has initiated many local and global scale microbiome projects. One famous example is the Human Microbiome Project (HMP; <https://hmpdacc.org/>), which was initiated by the National Institutes of Health (NIH), launched in 2008 aims to identify the complete healthy human microbiome to appreciate the diversity and complexity of the microbial communities.^[6] Similar initiatives have been taken from national funding agencies from different countries such as Commonwealth Scientific and Industrial Research Organization (Australia), Canadian Institutes of Health Research (Canada) (<https://cihr-irsc.gc.ca/e/39939.html>), European Commission (Europe) (<https://www.gutmicrob iotaforhealth.com/met ahit/>), National

Address for correspondence: Dr. Harinder Singh, Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS University, Vile Parle (West), Mumbai, Maharashtra, India. E-mail: Harinder.Singh@nmims.edu

Submitted: 03-Mar-2020,

Revised: 09-May-2020,

Accepted: 20-May-2020,

Published: 12-Jun-2020

Access this article online

Quick Response Code:



Website:
www.brjnims.org

DOI:
10.4103/BMRJ.BMRJ_4_20

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Kamble A, Sawant S, Singh H. 16S ribosomal RNA gene-based metagenomics: A review. *Biomed Res J* 2020;7:5-11.

Agency for Research (France), European Molecular Biology Laboratory (Germany), Medical Research Council (Gambia), Japan Science and Technology Agency (JST, Japan), National Research Foundation (Korea), and NIH (United States).^[7] The International Human Microbiome Consortium (IHMC; <http://www.human-microbiome.org/>) coordinate the activities and policies, share microbiome data and protocols and promote the generation of robust data resource. Recently, interdisciplinary Unified Microbiome Initiative was started with an objective to discover and understand different Earth's microbial ecosystems.^[8] In the recent past, a huge impetus has been given to the human microbiome research in different countries.^[9] Most of these studies have generated huge data, documenting the diverse microbes present in various populations that are different on the basis of geographical location, lineage, eating or working lifestyle, etc.^[4,5,9,10] This practice has definitely increased the demand for nontraditional and contemporary techniques to understand the microbial world which resist, avoid, or escape the routine cultivation.^[3] From these focused researches, an area of metagenomics has developed, that basically involve the genomic examination of a populace of microorganisms. Metagenomics is a habitat-based investigation of mixed microbial populations at the DNA level.^[4] The process of metagenomics involves isolating DNA from an environmental sample or any sample to be tested, followed by sequencing and genome analysis. Targeted metagenomics approaches such as 16S ribosomal RNA (16S rRNA) gene metagenomics, include steps such as sample collection, DNA isolation, 16S rRNA polymerase chain reaction (PCR) amplification, and cleanup, followed by the next generation sequencing (NGS), and sequence analysis using various computational tools.^[11]

The past decade has witnessed many reports on metagenomics studies for microbial diversity analysis.^[12-16] Large scale specific projects such as US NIH-funded HMP consortium have concentrated on producing reference genome of a "healthy individual" using metagenomics approaches.^[9,17] Several methods such as whole genome shotgun, metagenomics sequencing of 16S rRNA were employed to thereby obtain reference microbial genome of the human body.^[9] In addition, research groups worldwide are also working on dynamics and interactions, and specific components of the microbiome with a variety of disease conditions, including cancers.^[12-14,18,19] Various steps involved in such metagenomics studies are discussed below.

Sample collection and isolation of DNA

Isolating the DNA from a given sample is an important step in metagenomics as this is the starting material and decides the result quality for all the downstream processes. Isolation of good quality and quantity of DNA is necessary for moving forward with the sequencing and analysis processes. There are three critical factors in the isolation process which are important: a proper sample collection, isolating the intact and high-quality DNA, and isolating DNA free of PCR contaminants.^[14] Proper sample collection mainly means it is

necessary to ensure that the collected sample is a representative of the environment location under consideration. If the location in question is of a natural environment, it is very crucial to check and avoid the man-made/artificial interference/contamination, which can modify the biodiversity analysis. For example, a deep-sea marine sample collection should be done with the precaution to avoid the microbial contaminants from casual or incorrect sample collection practices or instruments.^[20] Similarly, a natural skin microflora sample should be from an untreated skin surface, as the diversity will change if it is being treated with any chemicals such as soap or cosmetics.^[21] A root rhizosphere soil should be collected strictly from the immediate surroundings of the plant roots, as it drastically differs from the neighboring root-free soil.^[15] The DNA isolation protocol must be chosen appropriately to avoid excessive shearing or degradation of DNA. Good yield of high-molecular weight DNA is preferred, although it is not a strict requirement. The high yield is preferable because it can increase the size of the population under analysis, as opposed to the low yield of genomic DNA. The yield, in turn, is dependent on an efficient cell lysis protocol, which is another important factor playing a role in DNA isolation. Cell lysis can be done either direct or indirect lysis. In direct lysis technique, cells are lysed in the sample itself, and then, DNA is recovered, whereas in indirect lysis technique, cells are first separated from the sample, and then, DNA is extracted. It has been noted that more DNA is isolated using the direct lysis method as compared to indirect, although higher purity is obtained in an indirect lysis method.^[14] Another alternative for improving the yield of DNA is to preculture the sample under required growth conditions, or collecting the sample from an environment where the sample is exposed to favorable conditions, thus enhancing the trait of interest naturally.^[22] This is particularly important when targeting a specific population of interest. For example, aiming for a novel thermostable microbe, a sample from sources like hot springs will be favorable instead of normal sample source. Isolating novel microbes capable of utilizing a particular substrate can be improved by growing the sample in minimal media with the required substrate to select the preferred population. It can be argued that these modifications will change the diversity, but it should be noted that the objective of such studies is to find novel candidates, rather than assessing diversity. Another important issue which is faced with specific samples such as soil is the PCR inhibitors such as humic and fulvic acids which are coextracted along with total soil DNA during extraction procedures.^[23,24] To remove or avoid the PCR inhibitors, DNA isolation kits are available and different protocols have been reported.^[25-29] Another point which has been observed is the content of DNA from nontarget organisms. For example, genomic DNA isolation from a human skin or oral sample will contain only a small percentage of bacterial DNA, and largely will be of human DNA, which gets easily isolated with any DNA isolation protocol. Similarly, the soil sample will have DNA from fungi, protozoa, and viruses along with the bacteria. If the next step is to do a specific PCR to amplify the target

gene, then this contaminant DNA will not be a hindrance. However, when whole-genome sequencing is performed, the analysis part will be tedious to remove the nontarget DNA data. Hence, the DNA isolation protocol can be selected based on the type of sample and the objective of the study.

16S ribosomal RNA gene sequencing and metagenomics

The selection of genes for amplification is a major step toward metagenomics. In most of the studies, 16S rRNA gene sequencing has been widely used for diversity analysis in the polymicrobial population.^[16,17,21] The use of rRNA gene sequence and its importance to characterize and study evolution of bacteria dates back to 1970s, where Carl Woese described the use of molecular sequences to determine the evolutionary relationships. rRNA gene is a present in all self-replicating systems, can be readily isolated, and its sequence changes slowly overtime, allowing the detection of relatedness among different bacterial species.^[30] The 16S rRNA is universally present in all prokaryotes and has multiple sub-regions, namely V1–V9 which can be used for the distinct identification of various prokaryotes. Along with the hypervariable region, there are the regions conserved across all prokaryotes which allows the designing of universal primers (F01, 8F, 357F, 515F, 1237F, 519R, 1100R, 1391R) to amplify the 16S rRNA gene [Figure 1 and Table 1].^[31,32] These properties of the 16s rRNA gene make it a useful marker for taxonomical classification and separation, giving rise to various 16S identification tools and databases such as the Ribosomal Database Project (<https://rdp.cme.msu.edu/>),^[33] Greengenes (<https://greengenes.secondgenome.com/>),^[34] and SILVA (<https://www.arb-silva.de/>).^[35] However, sequencing the entire 16S rRNA gene is not necessary for microbial diversity analysis. Single or combination of different variable regions can be used for diversity analysis. The choice of region varies and comparative studies have been carried out using the different regions of the 16S rRNA compositional analysis. Regions V3–V5 is more often used for this purpose, although the region V1–V2 is also reported to be used.^[36] It was observed that for identification of maximum archaeal sequences, regions V1–V4 showed best results for species richness at genus level and V3–V5 region for family level identification. For the analysis of bacterial sequences, V1–V4 regions showed best results for sample richness at species level and has least error rates.^[37] It is important that the region used for sequencing should be chosen rightly, for example, the V3–V4 region of 16S rRNA is a preferred region giving the best results and have low error rates as compared to V8–V9 and are more appropriate for clustering analysis.^[38,39] However, other regions of the 16S rRNA have also been used for sequencing in various studies.^[40,41] Because of the universality of the gene sequence, amplification of the 16S rRNA gene can lead to potential complications, if reagents are contaminated, leading to the amplification of unwanted products. In most of the bacteria, there are multiple copies of the 16S rRNA gene per cell, which can affect the quantitative studies toward understanding the proportion of different microbes in a sample.^[42] Recently,

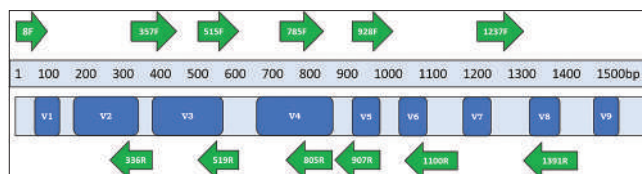


Figure 1: The figure shows the location of universal polymerase chain reaction primers used to amplify multiple sub regions (V1 to V9) of 16S ribosomal RNA gene for analyzing microbial diversity from various environmental samples

Table 1: Universal primers for 16S ribosomal RNA gene polymerase chain reaction

Name of primer	Sequence
8F	AGAGTTTGATCCTGGCTCAG
357F	CTCCTACGGGAGGCAGCAG
515F	GTGYCAGCMGCCGCGGTAA
785F	GGATTAGATACCCTGGTA
928F	TAAACTYAAAKGAATTGACGGG
1237F	GGGCTACACACGYGCWAC
336R	ACTGCTGCSYCCCGTAGGAGTCT
519R	GTATTACCGCGGCTGCTGG
805R	GACTACHVGGGTATCTAATCC
907R	CCGTCGAATTCCTTTRAGTTT
1100R	GGGTGCGCTCGTTG
1391R	GACGGCGGTGTGTRCA

The table shows the sequence details for important universal primers used for 16S ribosomal RNA gene PCR. PCR: Polymerase chain reaction

researchers have used other universal gene which is present in a single copy per cell (like *rpoB*), thus giving a correct estimation of the population numbers.^[43] A simple similarity search of universal primers showed that there is a slight possibility of not amplifying unique and novel sequences. The solution to this problem is to update the universal primers sequence periodically to include the new gene information that has been added recently to the database. The template DNA concentration used for PCR is crucial, there should be sufficient copies of the target for significant amplification. In certain cases, like DNA from the human body, the concentration of the target DNA template can be low, as most of it is human DNA, PCR amplification will not be efficient.^[14] An optimization PCR should be performed using specific target primers to decide the template concentration for an efficient amplification.

The selection of sequencing method depends on the requirement and specific objective of the research. Since metagenomics generally involves the detection of a mix population, simple sanger sequencing fails to sequence it. The next-generation sequencing such as pyrosequencing, illumina, nanopore, and PacBio sequencing are mostly used here. These techniques have been reviewed extensively, and readers are advised to refer to these review articles.^[44-47] Among the different NGS technologies, 454/Roche is one of the first sequencer to be manufactured. It clonally amplifies random DNA fragments attached to microscopic beads deposited in the picotitre plate.^[48]

The average read length of the sequence produced using 454/Roche is 600–800 bp. On the other hand, Illumina technology by Solexa amplifies DNA fragments immobilized on a surface resulting in a cluster of identical DNA fragments. The cluster density ranges from 170 to 1400 K clusters/mm², depending on the type of Illumina system used, and 150–300 bp read length can be obtained using this method.^[49] There are few more sequencing methods developed in recent times like Ion Torrent by Thermo Fischer Scientific which amplifies a read length from 200 to 600 bp with an output of 0.3–25 gigabases. This technology has smaller run time, i.e., 2.5–4 h as compared to other technologies.^[50] Third-generation sequencers have recently been started, which aim to amplify longer reads in a real-time sequencing. Single-molecule real-time (SMRT) sequencing from Pacific Biosciences (PacBio) was the first popular third-generation sequencer, SMRT sequencing introduced the capability of real-time sequence acquisition for read lengths > 1 kb using sequencing by synthesis and optical detection.^[51] In 2014, Oxford Nanopore Technologies released nanopore sequencing in the form of the MinION, a handheld sequencer that uses a grid of membrane-embedded biological nanopores.^[52] The MinION has the distinct advantage of being highly portable and capable of sequencing when plugged into a laptop.^[53]

For sequencing of 16S rRNA for taxonomic classification, reads up to 200–250 bp are satisfactory.^[54,55] The read length plays an important role in metagenomics studies since identifying up to species level can be difficult with shorter reads. Now, various kits are available, which can read up to 400 bp, making it easier for the identification of 16S rRNA sequences.^[56] Another important parameter in NGS metagenomics studies is Depth of Sequencing/Coverage, which is the number of times a genome has been sequenced. This primarily depends on the requirement and aim of the study. The sequencing depth varies from 30x-50x for whole-genome sequencing, to 100x for whole-exome sequencing and ChIP-Seq. For RNA sequencing, sequencing depth is calculated in terms of numbers of millions (M) of reads to be sampled. For profiling highly expressed genes, 5–25 million reads per sample should be enough, whereas for experiments looking for indepth transcriptome analysis, 100–200 million reads per sample might be required (<https://sapac.support.illumina.com/bulletins/2017/04/considerations-for-rna-seq-read-length-and-coverage-.html>). If the main goal of the study is to only identify the major bacterial phyla in the sample, a smaller sampling depth up to 300–400 M reads per lane is sufficient. A study by Ley *et al.* in 2008 concluded the presence of two major phyla in vertebrate gut microbiome of which 75% were firmicutes and 18% bacteroidetes by using 350 sequences/sample depth coverage.^[56,57] Shallow depth sampling can be carried for the detection of major community data/large scale pattern. On the other hand, studies that aim to identify species-level diversity in a sample, higher sequence read is required. For 16S rRNA sequencing, sequence reads of 1000sequences/sample or more are recommended to provide

species-level identification. Study carried out by Yang *et al.* in 2018 which involved diversity analysis up to the species level of oral microbiome in oral cancer patients with respect to different oral cancer stages, sequence depth of 10,000 sequences/sample was used as a threshold for analysis.^[19] Low-quality reads (Q-score <20) are generally not useful as metagenomics software such as Quantitative Insights Into Microbial Ecology (QIIME) filter out low quality reads that do not satisfy the requirement.^[58]

Data analysis

Several factors are important while assembling the data obtained from sequencing. Two major concerns while processing the data from raw format are the presence of appropriate read lengths and reduction of data processing requirements. Depending on the read length of the sequences, the pipelines used for gene prediction will differ. MG-RAST (<http://api.mg-rast.org/api.html>) and QIIME2 (QIIME: <https://qiime2.org/>) are an open source pipelines that suggests automatic phylogenetic and functional analysis of metagenomes, widely used to organize 16S rRNA sequences of length 100 bp and above,^[59] whereas IGM/M (The Integrated Microbial Genomes with Microbiome Samples: <https://img.jgi.doe.gov/m/>) prefers data input in the form of contigs. However, the longer the sequence length, the better is the potentiality of obtaining authentic information by comparing it with the genetic data by homology searching.^[60] To address the second concern while annotating the data, algorithms such as uclust^[61] or CD-HIT^[62] are used which assemble similar reads into contigs and clusters making further processing of raw data easier. Longer and more complex sequences cannot be analyzed without assembly, and therefore, might be lost from the data set if they are not arranged appropriately.

For typical metagenomics data for the identification of bacterial diversity, common pipelines such as MG-RAST and QIIME2 involve five major processes that sequences have to undergo before the actual visualization of the data [Figure 2].

First, the raw sequences are demultiplexed. Every Next Generation Sequencer can sequence the hundreds of samples in one run by multiplexing, where a unique barcode is added

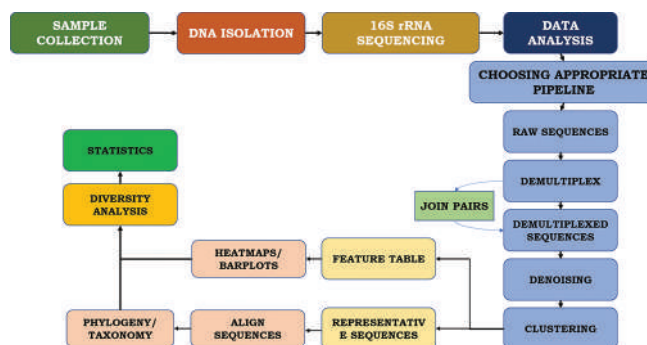


Figure 2: The steps involve are collection of various ecological samples, DNA extraction followed by 16S ribosomal RNA sequencing with the help of primers [Figure 1], and data analysis with the help of appropriate pipeline

to either one or both ends of each sequence. Once the samples are sequenced, these unique barcodes help in de-multiplexing, i.e., identifying individual sample sequences. Depending on the type of raw data, various options are available for demultiplexing. For example, q2-demux and q2-cutadapt are used in QIIME 2 for sequences in EMP (Earth Microbiome Project; <https://earthmicrobiome.org/>) format and multiplexed barcode in-sequence format respectively. The sequences obtained in the raw data can be either single-end (sequences are sequenced only in one direction) or paired-end sequences (each sequence is sequenced bi-directionally). After de-multiplexing, the pairs of paired-end sequences have to be identified and joined to each other for accurate information.^[63]

The second step in the pipelines is denoising. This step aims to filter out noisy reads, de-replicate (reduce repetition), remove singletons, remove chimeric sequences, and correct errors in marginal sequences. In QIIME2, denoising can be carried out by two methods, Divisive Amplicon Denoising Algorithm (DADA2) and Deblur. For Deblur, basic quality score-based filtering is required which is not necessary in DADA2. The outputs obtained after denoising are variants of “sequence variant (SV)” or “amplicon SV,” and these are 100% operational taxonomic units (OTUs).^[63] This step is a prerequisite for the next step, i.e., clustering. OTU clustering is implemented by either closed reference, open reference or *de novo* strategy. In closed reference clustering, all sequences are compared against a reference sequence collection, and any reads that do not match the reference sequences are discarded from downstream analysis, whereas in an open-reference clustering, reads are clustered against the reference and those that do not match the reference are clustered *de novo*. The final outcomes after denoising and clustering are feature table artefact and representative sequences artefact. The feature table gives a summary of all annotated features in the data (mRNA, genes, and sequences of 16S rRNA) along with the information such as the number of times a feature has been repeated in the data and in how many samples. On the other hand, representative sequences artefact gives information about the actual DNA sequence to every annotated feature which are later designated a taxonomy after classification using 16S rRNA databases like SILVA/RDP/GreenGenes. These are used in all downstream analysis and are the central record of all observations of a sample.^[64]

A series of programs further aid to classify the obtained sequences and feature table into taxonomy. Each sequence is allotted its taxonomy which helps in carrying out phylogenetic analysis. The taxonomic compositions can also be viewed as bar plots and heat maps. Following this, detailed analysis including alpha and beta analysis can also be carried out using this data. Alpha analysis measures the level of diversity within individual samples. Beta diversity measures the level of diversity or dissimilarity between the samples and dissimilarity index measures of microbiome composition dissimilarity along with principle coordinate analysis which are depicted in the form of Bray-Curtis matrix (quantitative) and Jaccard

matrix (qualitative). Rarefaction index can be calculated for the analysis of species richness for a given number of individual samples, based on the construction of so-called rarefaction curves.^[14,18,19,65] Diversity analyses, including Shannon Index, Simpson Index, and Faith PD are also determined using the pipelines. The whole data analysis result in a list of bacteria identified in the sample which can be further distributed in separate lists of Phylum, class, order, family, genus, and species and is dependent on the database chosen as a reference, for example, RDP, SILVA, Greengenes, etc., The diversity is given in the form of absolute count and percent composition. The identification is based on the fact that a sequence is matched with greater than or equal to 97% similarity to a sequence present in the database. This will give only the species of genus present in the database, and any new genus will not get highlighted in the results. The unclassified candidates are clustered under phylum Saccharibacteria formerly known as Candidate Division TM7.^[66] This group lacks the culture isolates, so, the practical role or implication for them is missing and further investigation is required for more information. Whole-genome sequencing can be one way to identify these novel microbes, and their genetic content which can further help in deciphering their role.

CONCLUSION

According to environmental microbiologists, <2% of bacteria can be cultured in laboratories from the different ecological environments. The percent culturability (a percentage of culturable bacteria in comparison with total cell counts) of bacteria varies for different habitats. For instance, it ranges from 0.001% to 15% for habitats such as for seawater, freshwater, mesotrophic lake, unpolluted estuarine waters, activated sludge, soil, and sediments.^[67-70] On the other hand, in habitats like the human oral cavity, 50% of microflora can be cultured.^[71] However, 16S rRNA gene sequence-based metagenomics approaches, widely used in the recent past, have enhanced our understanding of microbial diversity at such locations. These microbial diversity analysis studies have definitely led to the increased discovery of novel bacterial lineages. Some of the famous projects such as the HMP headed by NIH became possible only due to the advances in the metagenomics approaches. Not only do these studies help us in understanding the diversity of microbiome in different organs of the human body, but it also helps to understand the changing dynamics of this microbiome in the state of disease and other unhealthy conditions. It has helped in the finding and understanding the diversity in environmentally relevant microbiome population from different terrestrial and aquatic sources. Although most of the studies have been concentrated on bacterial studies, more initiatives should be taken to explore other microbes such as fungi and viruses. Efforts should also be made to enrich the available 16S rRNA gene database to remove the redundancy in entries, reducing the percentage of partial sequences, as species-level identification has more relevance from the perspective of applications. These studies

can be further enriched by whole-genome analysis, and related omics studies such as metatranscriptomics, metaproteomics and metametabolomics, to understand the functionality and dynamics of a microbial community. The network association between microbes and with the environment at different levels is important to understand the overall function and characteristics of an individual habitat and its microbial ecology. This can help in curing or modifying certain habitats such as soil for various agricultural applications. These metagenomics studies can also lead to exploration and discovery of novel genes, proteins, enzymes, metabolites, and active compounds that can be of medical, environmental, and agricultural significance. With new population diversity getting reported very frequently, it can help in revealing novel microbes that have application or potential in the medical, agricultural, and biotechnology industry. Future work in this field expects enhancement in cost-effective and streamlined protocols and workflow, technical expertise, efficient and robust computational/bioinformatics tools and pipelines. Recent metagenomics research has made significant contributions toward enhancing the microbial database; however, the large repertoire of microbes presents around us, warrants further research in order to understand and exploit them for the benefit of mankind.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- Stewart EJ. Growing unculturable bacteria. *J Bacteriol* 2012;194:4151-60.
- Pace NR, Stahl DA, Lane DJ, Olsen GJ. The analysis of natural microbial populations by ribosomal RNA sequences. In: Marshall KC, editors. *Advances in Microbial Ecology*. Advances in Microbial Ecology. Vol. 9. Boston, MA: Springer; 1986.
- Pace NR. Analyzing natural microbial populations by rRNA sequences. *ASM News* 1985;51:4-12.
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem Biol* 1998;5:R245-9.
- Kirk HJ, Kelley ST, Pace NR. New perspective on uncultured bacterial phylogenetic division OP11. *Appl Environ Microbiol* 2004;70:845-9.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007;449:804-10.
- Proctor LM, Creasy HH, Fettweis JM, Lloyd-Price J, Mahurkar A, Zhou W, *et al.* The integrative human microbiome project. *Nature* 2019;569:641-8.
- Alivisatos AP, Blaser MJ, Brodie EL, Chun M, Dangl JL, Donohue TJ, *et al.* MICROBIOME. A unified initiative to harness Earth's microbiomes. *Science* 2015;350:507-8.
- Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207-14.
- Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M, *et al.* Novel tetracycline resistance determinant from the oral metagenome downloaded from. *antimicrob. Agents Chemother* 2003;47:1430-2.
- Suenaga H. Targeted metagenomics: A high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol* 2012;14:13-22.
- Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Østerås M, *et al.* Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* 2009;79:266-71.
- Banerjee J, Mishra N, Dhas Y. Metagenomics: A new horizon in cancer research. *Meta Gene* 2015;5:84-9.
- Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 2017;66:70-8.
- Kim NH, Park JH, Chung E, So HA, Lee MH, Kim JC, *et al.* Characterization of a soil metagenome-derived gene encoding wax ester synthase. *J Microbiol Biotechnol* 2016;26:248-54.
- Xu X, He J, Xue J, Wang Y, Li K, Zhang K, *et al.* Oral cavity contains distinct niches with dynamic microbial communities. *Environ Microbiol* 2015;17:699-710.
- Dehingia M, Devi KT, Talukdar NC, Talukdar R, Reddy N, Mande SS, *et al.* Gut bacterial diversity of the tribes of India and comparison with the worldwide data. *Sci Rep* 2015;5:18563.
- Sánchez-Sanhueza G, Bello-Toledo H, González-Rocha G, Gonçalves AT, Valenzuela V, Gallardo-Escárate C. Metagenomic study of bacterial microbiota in persistent endodontic infections using Next-generation sequencing. *Int Endod J* 2018;51:1336-48.
- Yang CY, Yeh YM, Yu HY, Chin CY, Hsu CW, Liu H, *et al.* Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. *Front Microbiol* 2018;9:862.
- Turner S, Pryer KM, Miao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol Soc Protozool* 1999;46:327-38.
- Méthé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, *et al.* A framework for human microbiome research. *Nature* 2012;486:215-21.
- Elend C, Schmeisser C, Pop M, Creasy HH, Giglio MG, Huttenhower C, *et al.* A framework for human microbiome research. *Nature* 2012;486:215-21.
- Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors-occurrence, properties and removal. *J Appl Microbiol* 2012;113:1014-26.
- Watson RJ, Blackwell B. Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Can J Microbiol* 2000;46:633-42.
- Kamble A, Singh H. Different methods of soil DNA extraction. *Bio Protocol* 2020;10:1-23.
- Fatima F, Pathak N, Rastogi Verma S. An improved method for soil DNA extraction to study the microbial assortment within rhizospheric region. *Mol Biol Int* 2014;2014:518960.
- Fatima F, Chaudhary I, Ali J, Rastogi S, Pathak N. Microbial DNA extraction from soil by different methods and its PCR amplification. *Biochem Cell Arch* 2011;11:85-90.
- Bag S, Saha B, Mehta O, Anbumani D, Kumar N, Dayal M, *et al.* An improved method for high quality metagenomics DNA extraction from human and environmental samples. *Sci Rep* 2016;6:26775.
- Foong CP, Lakshmanan M, Abe H, Taylor TD, Foong SY, Sudesh K. A novel and wide substrate specific polyhydroxyalkanoate (PHA) synthase from unculturable bacteria found in mangrove soil. *J Polym Res* 2018;25:1-9.
- Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci U S A* 1977;74:5088-90.
- Lan Y, Rosen G, Hershberg R. Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. *Microbiome* 2016;4:18.
- Yang B, Wang Y, Qian PY. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 2016;17:135.
- Olsen GJ, Overbeek R, Larsen N, Woese CR. The ribosomal database project: Updated description. *Nucleic Acids Res* 1991;19:4817.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;72:5069-72.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* The

- SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590-6.
36. Abbai NS, Govender A, Shaik R, Pillay B. Pyrosequence analysis of unamplified and whole genome amplified DNA from hydrocarbon-contaminated groundwater. *Mol Biotechnol* 2012;50:39-48.
 37. Kim M, Morrison M, Yu Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Methods* 2011;84:81-7.
 38. Yong HS, Song SL, Chua KO, Lim PE. High diversity of bacterial communities in developmental stages of *Bactrocera carambolae* (Insecta: Tephritidae) revealed by illumina miseq sequencing of 16S rRNA gene. *Curr Microbiol* 2017;74:1076-82.
 39. Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L. Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS One* 2008;3:e2836.
 40. Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 2003;55:541-55.
 41. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;22:1658-9.
 42. Klappenbach JA, Dunbar JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 2000;66:1328-33.
 43. Adékambi T, Drancourt M, Raoult D. The rpoB gene as a tool for clinical microbiologists. *Trends Microbiol* 2009;17:37-45.
 44. Kumar KR, Cowley MJ, Davis RL. Next-generation sequencing and emerging technologies. *Semin Thromb Hemost* 2019;45:661-73.
 45. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Front Genet* 2015;6:348.
 46. Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, *et al.* Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics* 2014;15:96.
 47. Sandmann S, de Graaf AO, van der Reijden BA, Jansen JH, Dugas M. GLM-based optimization of NGS data analysis: A case study of Roche 454, Ion Torrent PGM and Illumina NextSeq sequencing data. *PLoS One* 2017;12:e0171983.
 48. Thomas RK, Nickerson E, Simons JF, Jänne PA, Tengs T, Yuza Y, *et al.* Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 2006;12:852-5.
 49. Wommack KE, Bhavsar J, Ravel J. Metagenomics: Read length matters † downloaded from. *Appl Environ Microbiol* 2008;74:1453-63.
 50. Merriman B, Torrent I, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* 2012;33:3397-417.
 51. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 2015;13:278-89.
 52. Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 2016;14:265-79.
 53. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;530:228-32.
 54. Zhang SW, Jin XY, Zhang T. Gene prediction in metagenomic fragments with deep learning. *Biomed Res Int* 2017;2017:4740354.
 55. Brady A, Salzberg SL. Phymm and PhymmBL: Metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* 2009;6:673-6.
 56. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: Evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* 2008;6:776-88.
 57. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335-6.
 58. Hernandez BY, Zhu X, Goodman MT, Gatewood R, Mendiola P, Quinata K, *et al.* Betel nut chewing, oral premalignant lesions, and the oral microbiome. *PLoS One* 2017;12:e0172196.
 59. Staggs C, Galloway M. Development of a local cloud-based bioinformatics architecture. Latifi S (ed.). In: *Advances in Intelligent Systems and Computing*. Cham: Springer Verlag; 2018. p. 559-65.
 60. Knietzsch A, Bowien S, Whited G, Gottschalk G, Daniel R. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* 2003;69:3048-60.
 61. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460-1.
 62. Chan CK, Hsu AL, Halgamuge SK, Tang SL. Binning sequences using very sparse labels within a metagenome. *BMC Bioinformatics* 2008;9:215.
 63. Crognale S, Tonanzi B, Valentino F, Majone M, Rossetti S. Microbiome dynamics and phaC synthase genes selected in a pilot plant producing polyhydroxyalkanoate from the organic fraction of urban waste. *Sci Total Environ* 2019;689:765-73.
 64. Kioroglou D, Mas A, Portillo MD. Evaluating the effect of QIIME balanced default parameters on metataxonomic analysis workflows with a mock community. *Front Microbiol* 2019;10:1084.
 65. Palmer RJ, Cotton SL, Kokaras AS, Gardner P, Grisius M, Pelayo E, *et al.* Analysis of oral bacterial communities: Comparison of HOMI NGS with a tree-based approach implemented in QIIME. *J Oral Microbiol* 2019;11:1586413.
 66. Ferrari B, Winsley T, Ji M, Neilan B. Insights into the distribution and abundance of the ubiquitous candidatus Saccharibacteria phylum following tag pyrosequencing. *Sci Rep* 2014;4:3957.
 67. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* 1995;59:143-69.
 68. Hugenholtz P, Pace NR. Identifying microbial diversity in the natural environment: A molecular phylogenetic approach. *Trends Biotechnol* 1996;14:190-7.
 69. Schmidt TM, DeLong EF, Pace NR. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* 1991;173:4371-8.
 70. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 1990;345:60-3.
 71. Wade W. Unculturable bacteria--the uncharacterized organisms that cause oral infections. *J R Soc Med* 2002;95:81-3.

Finding novel enzymes by *in silico* bioprospecting approach

Asmita Deepak Kamble and Harinder Singh

Department of Biological Sciences, Sunandan Divatia School of Science, NMIMS Deemed to be University, Mumbai, Maharashtra, India

1. Introduction

Enzymes are natural macromolecules and biocatalysts. They catalyze chemical reactions in organism under suitable physiological conditions. They have unique features, like high catalytic efficiency, specificity, eco-friendly conditions, which make them an attractive candidate for industries. The global enzyme market size from 2020 to 2027 is estimated to grow at a compound annual growth rate of 7.1% (Enzymes Market Size, Share & Trends Analysis Report by Application (Industrial Enzymes, Specialty Enzymes), by Product (Carbohydrase, Proteases, Lipases), by Source, by Region and Segment Forecasts, 2020–2027, 2020). Proteases, lipases, and carbohydrases are the most widely used classes of enzyme. They account for more than 70% of the total enzyme market (Li et al., 2012a).

Enzymes have wide application in various fields such biofuels (Choi et al., 2020), chemical industry (Basso and Serban, 2020), paper and pulp industry (Angural et al., 2020), medical (Vishnoi et al., 2020), food and feed industry (Handa et al., 2020), detergent and laundry sectors (Duffeck et al., 2020), textiles (Verma et al., 2021), personal care and cosmetics (Sunar et al., 2016), and nutraceutical (Vieira and Delerue Matos, 2020). For instance, Li et al. (2018) designed 18 enzymatic steps process to synthesize the anticancer compound, noscapine in *Saccharomyces cerevisiae* (Li et al., 2018). In 2018, Frances H. Arnold won Nobel Prize in Chemistry. She and her team replace harmful synthetic processes with mutated microbial enzymes (Arnold, 2018). The enzymes are also being used for synthesis and modification of optically pure drugs (Margolin, 1993). The native enzymes failed because of instability at broad range of pH, temperature, and solvent conditions (Shukla, 2020). There are continuous efforts to explore newer and better enzymes, as well as to modify and improve the existing repertoire of enzymes for industrial application.

There can be two ways to explore novel enzymes: conventional (Suleimanova et al., 2015; Sajidan et al., 2015) and *in silico* bioprospecting approach (Kamble et al., 2019; Tan et al., 2016a). *In silico* bioprospecting approach is an efficient, cost- and time-effective way to explore and identify novel enzymes from the existing databases that fulfill industrial demands (Kamble et al., 2019; Tan et al., 2016a). The identified novel enzymes through *in silico* bioprospecting approach can be subjected to advance genetic (Zhang et al., 2019) and/or chemical modifications (Naowarajna et al., 2021) or immobilization approaches (Zhang et al., 2020a,b). These modified enzymes can be applied in suboptimal or extreme condition reactions, such as temperature (cold/hot) (Li et al., 2019; Xu et al., 2015), salinity (Nasri et al., 2011), and pH (Grahame, 2016). For example, Coker (2016) generated the chimeric psychrophilic/mesophilic protease via directed evolution (Coker, 2016). This chapter discusses about different ways to perform *in silico* bioprospecting to identify novel enzymes from existing pool of sequences deposited in databases. This will be followed by various approaches that can be used to perform *in silico* characterization of novel enzyme candidates.

2. *In silico* bioprospecting

Bioprospecting of enzyme can be performed by conventional and computational approach (Kamble et al., 2019). Basic workflow of bioprospecting (conventional and *in silico*) is described in Fig. 26.1.

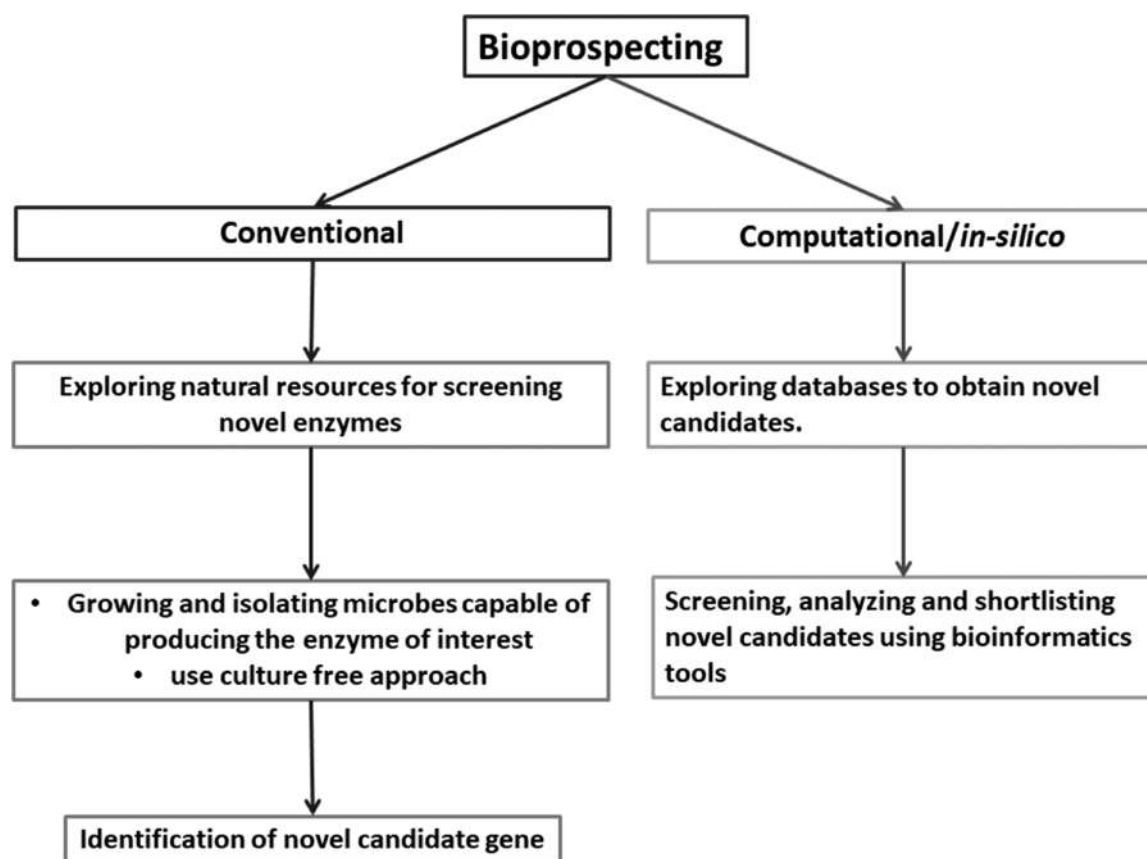


FIGURE 26.1 Conventional and in silico/computational bioprospecting workflow.

2.1 Conventional approach

The conventional approach involves exploring natural resources for screening novel enzymes (Banik et al., 2018; Boland et al., 1991; Poddar and Das, 2018; López-López et al., 2013; Suleimanova et al., 2015), growing and isolating microbes capable of producing the enzyme of interest or using culture-free approach (Sajidan et al., 2015; Kumar et al., 2013; Yao and Fan, 2007), identification of novel candidate gene in a genome sequence, and the experimental validation (Kamble et al., 2019; Arnold, 2018; Reetz et al., 2006; Kim and Lei, 2008). The conventional approach is cost- and time-ineffective (Li et al., 2012a,b). The conventional approaches depend upon the samples purity and quality (Rolf et al., 2020). The shortlisting of microbes is based on selective media (Priyodip et al., 2018). The PCR amplification with specific or degenerative primers requires well-established protocol (Yao and Fan, 2007). Recently, an extensive and thorough protocol optimization was performed for genomic DNA isolation from different soil sources (Kamble and Singh, 2020). The screening process via conventional approach may or may not yield positive candidates enzyme (Kamble et al., 2019; Bae et al., 1999).

2.2 Computational approach

The computational approach, i.e., *in silico* bioprospecting, yields novel candidates from the existing database. It is simple and effective approach (Voß et al., 2020; Li et al., 2012a; Stewart, 2012). The *in silico* bioprospecting takes advantage of these enzyme diversity data. The computational approach involves following steps:

1. Exploring databases to obtain novel candidates.
2. In silico characterization: Screening, analyzing, and shortlisting novel candidates using bioinformatics tools.

2.2.1 Exploring databases to obtain novel candidates

2.2.1.1 Why there is a need to explore databases to obtain novel candidate?

The National Center for Biotechnology Information (NCBI) consisted of 34 databases integrated Entrez system. These 34 databases hold around 3 billion records. The total number of nucleotide and the protein sequences deposited are

42,97,31,711 and 87,42,72,642, respectively, as of September 9, 2020 (Sayers et al., 2021). As mentioned above, the conventional methods are performed, either by screening of strains (Kumar et al., 2013) or via metagenomics approach (Castillo Villamizar et al., 2019), and have limitations. Alternatively, the shortlisted candidates can be subjected to OMICS approach in which the microbial proteome is extracted, analyzed via OMICS, and identified the presence of active enzymes (Kalia, 2017). Many researchers lately are taking the metagenomics approach, which is based either on activity, i.e., functional metagenomics (Castillo Villamizar et al., 2019), or sequence-based metagenomics (Daniel, 2005). However, it is also possible to explore the available metagenomic protein sequence dataset for *in silico* bioprospecting (Vaquero et al., 2015).

2.2.1.2 Ways to search databases via *in silico* bioprospecting approach to obtain novel enzymes

You can refer to number of databases, listed in Table 26.1, which can be used for *in silico* bioprospecting to obtain novel industrially important enzymes (Voß et al., 2020), antibiotics (Nirwan et al., 2021), therapeutic compounds (Deng et al.,

TABLE 26.1 Tools used for exploring and characterizing novel enzymes via *in silico* bioprospecting approaches.

S. No.	Database	Details	References	Weblink
1.	Uniprot	It is the protein database. The data are collected from Swiss-Prot, TrEMBL, and PIR-PSD database.	Magrane and Consortium (2011)	https://www.uniprot.org/
2.	Conserved Domains database	The protein functional units are annotated based on conservation pattern.	Lu et al. (2020)	https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
3.	Protein database	The protein sequences are distributed in GenBank, RefSeq, TPA, Swiss-Prot, RIP, PDB database.	Pruitt et al. (2012) Berman et al. (2007)	www.ncbi.nlm.nih.gov/home/proteins.shtml
4.	Protein cluster	It includes the RefSeq protein sequence encoded by chlorophyll genome. It also includes prokaryotic genome.	Klimke et al. (2009)	https://www.ncbi.nlm.nih.gov/proteinclusters
5.	Structure	The protein sequence data are derived from PDB database. The information of sequences is linked to sequence database, NCBI taxonomy database, and bibliographic information.	Madej et al. (2012)	https://www.ncbi.nlm.nih.gov/structure
6.	Protein Data Bank (PDB)	The database consists of X-ray diffraction or NMR structure data of protein and also consists of gene sequence. The 3D structure of protein is readily available.	Berman et al. (2007) Burley et al. (2021)	https://www.rcsb.org/
7.	InterPro	It helps to predict protein family, domain, and binding site. It uses CATH, CDD, PRINTS, Pfam database for the prediction.	Blum et al. (2021)	https://www.ebi.ac.uk/interpro/
8.	Brenda	It is a collection of information-related enzymatic reaction, structure, function, kinetic parameters, stability, and sequence data.	Schomburg et al. (2004)	https://www.brenda-ews6Y6s.org/
9.	SABIO-RK (BKMS-react), BRENDA, KEGG, MetaCyc	It includes information about the enzymatic reactions, metabolic pathways, as well as experimental conditions.	Lang et al. (2011)	http://bkms-react.tu-bs.de/index.php

Continued

TABLE 26.1 Tools used for exploring and characterizing novel enzymes via *in silico* bioprospecting approaches.—cont'd

S. No.	Database	Details	References	Weblink
10.	A Database of Enzyme Catalytic Mechanisms (EzCatDb)	This includes information about enzymatic reactions, cofactors, metabolites, and active site domain and catalysis.	Nagano (2005)	http://ezcatdb.cbrc.jp/EzCatDB/
11.	Mechanism and Catalytic site Atlas: (M-CSA)	It includes information about the catalytic residues, cofactors, and reaction mechanism.	Ribeiro et al. (2018)	https://www.ebi.ac.uk/thornton-srv/mcsa/
12.	Thermodynamic Database for Proteins and Mutants (ProTherm)	This contains information related to the change in thermal stability as mutations are introduced.	Gromiha et al. (2000)	https://www.iitm.ac.in/bioinfo/ProTherm/
13.	ProtaBank	It includes mutated protein sequence information.	Wang et al. (2018)	https://www.protabank.org/
14.	Kyoto Encyclopedia of Genes and Genomes (KEGG)	This database is a collection of genomic, chemical, and system function information.	Yi et al. (2020)	https://www.genome.jp/kegg/
15.	Protein Families (Pfam)	This is a protein family database. The multiple sequence alignment and hidden Markov model helps to cluster these families.	El-Gebali et al. (2019)	http://pfam.xfam.org/
16.	PRINTS	This is a protein motif database.	Attwood et al. (1994)	http://130.88.97.239/PRINTS/index.php
17.	CATH (Class, architecture, topology, and homologues)	This is a protein database. It consists of protein superfamily information such as Class, architecture, topology, and homologous sequences.	Knudsen and Wiuf (2010) Dawson et al. (2017)	https://www.cathdb.info/
18.	Structural Classification of Proteins (SCOP)	This helps in protein structure classification.	Andreeva et al. (2020)	http://scop.mrc-lmb.cam.ac.uk/
19.	Homology-derived Secondary Structure of Proteins (HSSP)	This includes protein secondary structure database.	Touw et al. (2015)	https://swift.cmbi.umcn.nl/gv/facilities/
20.	SWISS-3DIMAGE	The 3D structure of proteins information is gained.	Peitsch et al. (1995)	http://www.pdg.cnb.uam.es/cursos/Leon_2003/pages/visualizacion/programas_manuales/spdbv_userguide/us.expsasy.org/sw3d/index.html
21.	Biological Magnetic Resonance Data Bank (BioMagResBank)	This has the collection of NMR database of proteins, amino acids, and nucleotides.	Ulrich et al. (2008)	https://bmrdb.io/
22.	SWISS-MODEL Repository	This database consists of 3D structure.	Bienert et al. (2017)	https://swissmodel.expasy.org/repository
23.	iProClass	This is a protein classification database.	Wu et al. (2004)	https://proteininformationresource.org/pirwww/dbinfo/iproclass.shtml
24.	TIGRFAM	This is a protein family database.	Haft et al. (2013)	http://tigrfams.jcvi.org/cgi-bin/index.cgi
25.	OWL	It is a protein sequence database. The information is derived from SWISS-PROT, PIR, GenBank, and NRL-3D primary sequence database.	Bleasby et al. (1994)	http://130.88.97.239/OWL/

TABLE 26.1 Tools used for exploring and characterizing novel enzymes via in silico bioprospecting approaches.—cont'd

S. No.	Database	Details	References	Weblink
26.	The database of three-dimensional interacting domains (3DID)	This is a database which involves interaction information of protein with known 3D structure.	Mosca et al. (2014)	https://3did.irbbarcelona.org/
27.	DOMINE	This is a protein domain interaction database.	Yellaboina et al. (2011)	https://manticore.niehs.nih.gov/cgi-bin/Domine
28.	Binding MOAD (the Mother of All Databases)	It is collection of protein–ligand crystal structure information.	Smith et al. (2019)	https://bindingmoad.org/
29.	Phospho.ELM	It is a protein phosphorylation site database.	Dinkel et al. (2011)	http://phospho.elm.eu.org/
30.	SuperSite	The database is a collection of drug–protein binding site information along with its mechanism of action and conservation of binding site information.	Bauer et al. (2009)	https://www.hsls.pitt.edu/obrc/index.php?page=URL1237565601
31.	STITCH	This consists of protein–compound interaction network information.	Kuhn et al. (2008)	http://stitch.embl.de/
32.	Reactome	This includes the detail information about the protein molecules involved in various biochemical pathways. It also provides a network chart of biochemical processes. The information is collected from UniPort, KEGG, OMIM, etc.	Fabregat et al. (2017)	https://reactome.org/
33.	UniHI	This gives information about the protein–protein interaction.	Kalathur et al. (2014)	http://www.unihi.org/
34.	Bionemo	It is a collection of data related to proteins and genes involved in biodegradation metabolism. It also includes information related to structural, biochemical information about the proteins and genes involved in biodegradation pathways.	Carbajosa et al. (2009)	https://www.hsls.pitt.edu/obrc/index.php?page=URL1233778230
35.	The proteolysis map (PMAP)	Proteolysis pathway database.	Igarashi et al. (2009)	https://www.hsls.pitt.edu/obrc/index.php?page=URL1233764714
36.	CAZy (the Carbohydrate-Active enZymes Database)	The carbohydrates degradation enzyme database.	Cantarel et al. (2009)	http://www.cazy.org/
37.	The Lipase engineering database (LED)	The lipase engineering database.	Fischer and Pleiss (2003)	http://www.led.uni-stuttgart.de/
38.	MEROPS	This is a peptidase database.	Rawlings et al. (2018)	https://www.ebi.ac.uk/merops/
39.	Protease substrate specificity web server (PROSPER)	It predicts the protease catalytic substrate and cleavage sites based on the sequence comparison with the protease superfamilies.	Song et al. (2011)	https://prosper.erc.monash.edu.au/home.html
40.	SABIO-RK	This includes information about the biochemical and kinetic parameters.	Wittig et al. (2006)	http://sabio.h-its.org/

Continued

TABLE 26.1 Tools used for exploring and characterizing novel enzymes via *in silico* bioprospecting approaches.—cont'd

S. No.	Database	Details	References	Weblink
41.	EBI Metagenomics	It is an automated pipeline for storage of metagenomic data, functional and metabolic analysis.	Mitchell et al. (2018)	https://www.ebi.ac.uk/metagenomics/
42.	Human Pan Microbe Communities Database (HPMCD)	It consists of manually curated and searchable human gastrointestinal microbial data.	Forster et al. (2016)	http://www.hpmcd.org/
43.	KEGG MGENES Database	It consists of gene data obtained from large-scale environmental sequencing studies.	Yi et al. (2020)	https://www.genome.jp/mgenes/
44.	Metagenomes Online	It consists of curated protein data obtained from environmental metagenomic studies.	Nasko et al. (2018) Eric Wommack et al. (2012)	http://metagenomesonline.org/
45.	Viral Informatics Resource for Metagenome Exploration (VIROME)	The viral metagenomic sequences are analyzed.	Eric Wommack et al. (2012)	http://virome.dbi.udel.edu/
46.	Piphillin	It helps in prediction of gene composition. This is done with the help of 16S rRNA OTU from human clinical samples metagenomic dataset.	Narayan et al. (2020)	http://piphillin.secondgenome.com/
47.	Web services for metagenomic analysis (WebMGA)	These are a collection of customizable web servers. This helps in fast metagenomic analysis.	Wu et al. (2011)	http://weizhong-lab.ucsd.edu/webMGA/
48.	MG-RAST	It is a collection of: repository + annotation + analysis tool for metagenomic analysis.	Keegan et al. (2016)	https://www.mg-rast.org/
49.	Naïve Bayesian Classification (NBC) tool	This helps in taxonomic classification of metagenomic data.	Rosen et al. (2011)	http://nbc.ece.drexel.edu/
50.	CoMet	It uses protein domain signatures for comparative analysis of metagenomic data.	Lingner et al. (2011)	http://comet2.gobics.de/
51.	MetaBioME	This helps in searching commercially important enzymes in the metagenomic dataset.	Sharma et al. (2009)	http://metasystems.riken.jp/metabiome/
52.	16S Classifier	It uses 16s rRNA hypervariable regions for taxonomic classification.	Chaudhary et al. (2015)	http://metagenomics.iiserb.ac.in/16Sclassifier/application.php
53.	ENZYME	It is the collection of information related to characterized enzymes and its EC number.	Bairoch (2000)	https://enzyme.expasy.org/
54.	ExplorEnz	The IUBMB enzyme nomenclature is curated and disseminated.	McDonald et al. (2009)	https://www.enzyme-database.org/
55.	The Structure–Function Linkage Database (SFLD)	It helps in hierarchical classification of enzymes as well as predicting structure–function relationship.	Eyal et al. (2014)	http://sfld.rbvi.ucsf.edu/archive/django/index.html

TABLE 26.1 Tools used for exploring and characterizing novel enzymes via in silico bioprospecting approaches.—cont'd

S. No.	Database	Details	References	Weblink
56.	Catalytic Families (CatFam)	The enzyme sequence profile is used for understanding protein catalytic functions.	Yu et al. (2008)	http://www.bhsai.org/downloads/catfam.tar.gz
57.	MetaCyc	The experimental data of metabolic pathways and enzymes are stored in this database.	Karp et al. (2002)	https://metacyc.org/
58.	EAWAG-BBD	It gives information about the biocatalytic reactions and biodegradation pathways.	Ellis and Wackett (2012)	http://eawag-bbd.ethz.ch/
59.	Fungal Peroxidase Database (fPoxDB)	It is a database of fungus peroxidases. It helps in comparative and evolutionary genomics of peroxidases.	Choi et al. (2014)	http://peroxidase.riceblast.snu.ac.kr/index.php?a=view
60.	Enzyme Structure Database	It is a collection of known enzyme structure and these data are submitted into PDB.	Laskowski et al. (2018)	https://www.ebi.ac.uk/thornton-srv/databases/enzymes/
61.	Integrated Relational Enzyme Database (IntEnz)	It helps in enzyme nomenclature.	Fleischmann et al. (2004)	https://www.ebi.ac.uk/intenz/
62.	REBASE	It is a collection of information related to DNA restriction and modification system.	Roberts et al. (2015)	http://rebase.neb.com/rebase/rebase.html
63.	ESTerases and Alpha/beta-Hydrolase Enzymes and Relatives (ESTHER Database)	It helps in analyzing the proteins and protein domains of hydrolyses.	Lenfant et al. (2013)	http://bioweb.supagro.inra.fr/ESTHER/general?what=index
64.	A peroxidase database (PeroxiBase)	It helps in identifying putative functions and transcription regulation of peroxidase.	Passardi et al. (2007)	https://www.uniprot.org/database/DB-0072
65.	KinBase	It is a collection of information related to protein kinases.	Caenepeel et al. (2004)	https://www.hsls.pitt.edu/obrc/index.php?page=URL1054842958
66.	ArthropodaCyc	It helps in predicting and identifying enzymes and their pathways that are unique to a given organism and/or group of organism.	Baa-Puyoulet et al. (2016)	http://arthropodacyc.cycadsys.org/
67.	A database for plant carbohydrate-active enzymes (PlantCAZyme)	It helps in providing sequence and data related to CAZymes.	Ekstrom et al. (2014)	http://bcbl.unl.edu/plantcazyme/
68.	A database of ORphan ENZYme Activities (ORENZA)	It helps in predicting enzymatic activities of enzyme which does not have any representative sequence in the database.	Lespinet and Labedan (2006)	http://www.orenza.universite-paris-saclay.fr/
69.	MEME	Functional motifs.	Bailey et al. (2015)	https://meme-suite.org/meme/
70.	ProtoNet	The orthologous sequences were clustered.	Rappoport et al. (2012)	http://www.protonet.cs.huji.ac.il/
71.	TargetP/ChloroP	Subcellular localization.	Emanuelsson et al. (2000) Heinegård and Paulsson (1987)	http://www.cbs.dtu.dk/services/TargetP/ http://www.cbs.dtu.dk/services/ChloroP/

Continued

TABLE 26.1 Tools used for exploring and characterizing novel enzymes via *in silico* bioprospecting approaches.—cont'd

S. No.	Database	Details	References	Weblink
72.	TMHMM server	Predicting membrane protein.	Krogh et al. (2001)	http://www.cbs.dtu.dk/services/TMHMM/
73.	SignalP	Detecting Signal peptide.	Almagro Armenteros et al. (2019)	http://www.cbs.dtu.dk/services/SignalP/
74.	ProtParam tool	The molecular weight, theoretical pI, total number of positive and negative residues, extinction coefficient, aliphatic index, instability index, and grand average of hydropathicity (GRAVY) are predicted.	Gasteiger et al. (2005)	https://web.expasy.org/protparam/
75.	The self-optimized prediction method (SOPMA)	The percentage of α -helix, β -pleated sheet and coils were analyzed.	Geourjon and Deléage (1995)	https://vlab.amrita.edu/?sub=3&brch=275&sim=1454&cnt=1
76.	PsiPred	The identification of structural details such as disordered region of protein, disulfide bond, internal repeats, pest region, ubiquitination site, n-glycosylation site, intron–exon organization.	McGuffin et al. (2000)	http://bioinf.cs.ucl.ac.uk/psipred/
78.	Modeller	Structure modeling.	Eswar et al. (2006)	https://salilab.org/modeller/
79.	Computed Atlas of Surface Topography of proteins (CASTp)	Predicting the conserved active catalytic residues in selected protein model.	Tian et al. (2018)	http://sts.bioe.uic.edu/castp/index.html
80.	PROCHECK	Helps in understanding stereochemistry quality of the protein structure.	Laskowski et al. (1993)	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
81.	Protein Structure analysis (PROSA)	It helps in refinement of structure, structure prediction, and modeling.	Wiederstein and Sippl (2007)	https://prosa.services.came.sbg.ac.at/prosa.php
82.	VERIFY3D	It is used for evaluation of structure.	Eisenberg et al. (1997)	https://servicesn.mbi.ucla.edu/viewer/?job=9678&p=verify3d
83.	ERRAT	It is used to verify the reliability of the modeled structure.	Colovos and Yeates (1993)	https://servicesn.mbi.ucla.edu/ERRAT/
84.	Qualitative Model Energy Analysis (QMEAN)	It is used to check the quality and analyze the stereochemistry of the modeled structure.	Benkert et al. (2007)	https://swissmodel.expasy.org/qmean/

2018), and unravel unidentified biochemical pathways and functions (Fabregat et al., 2017). There are around $4-6 \times 10^{30}$ prokaryotes that inhabit the earth (Whitman et al., 1998), from which 95%–99.9% of microorganisms are uncultivable (Li et al., 2009). We can either improvise our cultivation-based approaches, or adopt cultivation-independent approach to identify them (Li et al., 2009). Pace and colleagues introduced cultivation-independent approach that targeted isolation of all the possible genomes from all the microorganisms in a given habitat (Pace et al., 1986). The first metagenome-based community characterization using 16s rDNA genes was performed by Schmidt et al. (1991), Li et al. (2009). Thus, it is preferable to consider both cultivable and uncultivable microbial community from any environmental samples (Stewart, 2012; Foong et al., 2018).

The basic workflow of metagenomics involves extraction and shearing of metagenomic DNA from any environmental samples, construct metagenomics library, sequence, assembled and classified taxonomically (Foong et al., 2018; Priyadarshini and Singh, 2019; Kamble et al., 2020; Haynes, 2013; Gao et al., 2016; Méndez-García et al., 2018). The combination of proteomics and transcriptomics paves a way to a successful enzyme prediction (Tan et al., 2019). In silico bioprospecting can also be used to find novel enzymes when the substrate is known (Zhang et al., 2020a,b). To identify a novel candidate, the sequence databases can be searched on the basis of homology (Voß et al., 2020), conserved motif (Vaquero et al., 2015), consensus-guided approach (Toyama et al., 2018), or simply by keyword search (Gupta et al., 2018). The candidates obtained can be filtered out on the basis of percentage identity (Bunterngsook et al., 2021), query coverage (Gupta et al., 2018), and e-value (González-Torres et al., 2020). The parameters such as protein coding gene sequence, or protein domain, or protein function, can be utilized to search novel enzymes from the databases (Kamble et al., 2019).

The approach to search novel candidate through homology or presence or absence of conserved domain involves (Zhang et al., 2020a,b) blast search against the databases to find homologous sequence (Zhang et al., 2020a,b; Voß et al., 2020), and analyze evolutionary relationship using servers/softwares such as MEGA-X in order to remove distant relatives (Zhang et al., 2020a,b; Vaquero et al., 2015); tertiary structure prediction is done by using softwares/servers such as Phre2, Swiss-model, or Modeller (Zhang et al., 2020a,b; Bunterngsook et al., 2021), analysis of consensus and nonconsensus residues (Zhang et al., 2020a,b; Toyama et al., 2018), and analysis of catalytic site (Zhang et al., 2020a,b). The combined *in silico* and experimental approach can also be adopted to harness novel candidate (Zhang et al., 2020a,b). For example, the combined experimental and *in silico* approach was adopted to explore camel rumen metagenomics data and discover a novel alkali-thermostable xylanase (Zhang et al., 2020a,b; Priyadarshini and Singh, 2019; Datta et al., 2020).

2.2.1.3 Studies using *in silico* bioprospecting of novel enzymes

Bunterngsook et al. (2021) performed *in silico* bioprospecting to search for putative lytic polysaccharide monoxygenases (LPMOs) from thermophilic microbial metagenomics dataset of soil present near bagasse collection site by using Blastp tool against the Carbohydrate-Active enZYmes (CAZy) database (Bunterngsook et al., 2021). Similarly, the nr database present in NCBI was used to search the putative L-Asparaginases (ASNases) (González-Torres et al., 2020). Araújo et al. (2020) retrieved 172 hypothetical proteins (HPs) from 32 *Corynebacterium pseudotuberculosis* biovar *ovis* genomes using NCBI database and online platform RAST (Rapid Annotations using Subsystems Technology) (Araújo et al., 2020). In the year 2017, around 6000 clones' active or purified enzymes were extracted from the metagenomics database (Zhang et al., 2020a). Góngora-Castillo et al. (2020) developed a strategy to extract novel proteases S8A from metagenomics database such as Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST) database (Góngora-Castillo et al., 2020). The nine novel proteases S8A were discovered that exhibited an S8 catalytic triad and belonged to phyla Proteobacteria, Actinobacteria, Planctomycetes, Bacteroidetes, and Cyanobacteria (Góngora-Castillo et al., 2020). Ariaeenejad et al. (2020) mined novel alkali-thermostable endo- β -1,4-glucanase from raw sheep rumen metagenomics data. Balik Pulau (BP) mangrove soil metagenome data deposited in MG-RAST database were explored to find polyhydroxyalkanoate (PHA) synthase gene (Foong et al., 2018).

2.2.1.4 Studies using *in silico* bioprospecting on the basis of homology

A known candidate is used to search the existing database in order to extract novel enzyme sequences. The sequence can be in the form of protein or DNA sequence. The gene sequence used to search the databases should be a full length complete sequence, and retain conserved domain(s), as the incomplete sequence yields nonfunctional proteins. The similar sequences that are >90% identical are discarded. This helps to shortlist novel candidate and not the closest homologous sequences. The sequences which are retained will be approximately 80% identical and exhibits the presence of conserved domain(s)/residues. These parameters ensure the selection of novel candidates with conserved domain and away from the closet homologous sequences. This comparison enables to find out the similarities and differences among the aligned sequences, aids in verifying protein homologies, and builds phylogenetic tree that trace evolutionary linkages (Kamble et al., 2019; Zhang et al., 2020a,b). For example, Gupta et al. (2018) searched NCBI database by using keywords such as 'Hypothetical Protein of *Triticum aestivum*,' 'Hypothetical Proteins of wheat.' The shortlisted sequences were then subjected to Blastp (Gupta et al., 2018). Rolf et al. (2020) identified novel cyclic GMP-AMP synthase (cGAS) homologues using similar approach (Rolf et al., 2020). Yata et al. (2021) compared *Cucumis sativus* Urease (CsU), *Fragaria vesca* Urease (FvU) with the reference Pigeon pea Urease (PpU) by using computational tools (Yata et al., 2021). Voß et al. (2020) mined 96 putative novel β -etherases by using peptide pattern recognition tool (Voß et al., 2020). Şahutoğlu et al. (2020) studied the structural and functional aspect of two novel endo- β -N-acetylglucosaminidase isoenzymes (Şahutoğlu

et al., 2020). Sadeghian et al. performed *in silico* analyses to identify thermostable multicopper oxidases that exhibit bilirubin oxidases (BODs) by using Blast search and NCBI database (Sadeghian et al., 2020). The novel α -amylases were mined by using sheep rumen metagenomics data (Motahar et al., 2020).

Sequence similarity network (SSN) can aid in finding the homologous sequences. The SSN can be generated in Pythoscape tool. The user-friendly tool to develop SSN is Enzyme Similarity Tool (EFI-EST network tool) and result can be evaluated via Cytoscape software (Gerlt et al., 2015).

The computational software(s) such as MEME can be used to search conserved motif. The approach sometimes does not yield any significant candidate. For example, the conserved motifs in *Pseudozyma antarctica* lipase B family were identified by using MEME software (Vaquero et al., 2015). The *ado* gene from *Synechococcus elongatus* PCC7942 was used as a query to search IMG/MER hot spring database (Shakeel et al., 2018). Toyama et al. (2018) retrieved a novel β -glucosidase by exploring microbial metagenome of lake Poraquê via sequence similarity search (Toyama et al., 2018).

2.2.1.5 *In silico* bioprospecting via analyzing conserved domain

The conserved domain can be used to identify novel candidate. For example, the NCBI protein and UniProt database can be used to retrieve HP sequences. The data from NCBI database were run on Conserved Domains Database (CDD)-search suite to obtain clusters of sequences similar to the input HP sequences. The data were arranged on the basis of conserved domain. The data obtained from CDD were cross-checked on Protein families (Pfam) and InterProScan (Guha et al., 2020). Tan et al. (2016b) retrieved novel histidine acid phytases from IMG/M database. The 18 metagenomes of fungus garden in IMG/M database were explored on the basis of conserved domain of histidine acid phosphatases. The total number of 11 putative sequences was shortlisted (Tan et al., 2016b). The microbiome dataset generated through next-generation sequencing technology along with putative domain information can be used to explore novel candidates (Thornbury et al., 2018). Maleki et al. (2020) mined two novel thermostable cellulases from camel rumen microbiome samples metagenomic sequence data. Ariaeenejad et al. (2018) mined thermostable, halotolerant, and glucose-tolerant novel β -glucosidase via *in silico* bioprospecting approach. They used metagenomics data of sheep rumen microbiota (Ariaeenejad et al., 2018). Musumeci et al. (2017) mined novel monooxygenases from cold marine sediment metagenomic dataset deposited in IMG database.

2.2.1.6 The combination of homology and conserved domain approach

The combination of approaches such as homology based or conserved domain information along with the data deposited in the databases can be used for *in silico* bioprospecting of novel enzymes. The approaches are implemented by using computational tools such as Blastp tool. The parameters such as e-value, presence or absence of conserved domain, and conserved motif are used to filter out data to get relevant novel candidate(s). For example, Sahoo et al. (2020) used genome information and computational tools such as Blastp to retrieve HPs from 26 microbial genomes which can contribute in algal lipid metabolism. Only those candidates, which possess the conserved domains, were retained (a total of 285 sequences) and characterized by using various computational tools (Sahoo et al., 2020).

2.2.2 *In silico* characterization of novel candidates

The *in silico* characterizations of putative sequences are performed using various computational tools (Table 26.1).

2.2.2.1 Primary sequence analysis and secondary structure analysis

The primary sequence analysis involves isoelectric point (pI), total amino acid composition, instability index, aliphatic index, and grand average of hydropathicity (GRAVY). The ExPasy-ProtParam and PEPSTATS tool was used to evaluate physiochemical properties of shortlisted candidates (Bhagwat et al., 2021; Sharma et al., 2017; Guha et al., 2020; Sharma, 2020; Smith and Caruso, 2013). The secondary structure analysis involves evaluation of the number of helices, sheets, turns, and coils present in the protein sequence. It is done by using tools such as Position Specific iteration (PSI)-Blast based secondary structure PREDiction (PSIPRED), ESPript, the Self-Optimized Prediction Method with Alignment (SOPMA), SIB MyHits Motif scan, and Pfam SOPMA server (Bhagwat et al., 2021; Hemmati and Mehrabadi, 2020; Şahutoğlu et al., 2020; Motahar et al., 2020).

2.2.2.2 Tertiary structure analysis and evaluation

Iterative Threading ASSEMBLY Refinement (I-TASSER) (González-Torres et al., 2020; Şahutoğlu et al., 2020; Hemmati and Mehrabadi, 2020), MODELLER (Shakeel et al., 2018), HHpred web server, Phyre2 server (Ariaeenejad et al., 2020;

Maleki et al., 2020), and Swiss-model (Bunterngsook et al., 2021) can be used for 3D structure analysis. The modeling depends on the identity, similarity, and extent of coverage between the target and the template (Bhagwat et al., 2021). The combination of SWISS-MODEL and I-TASSER ROBETTA web server was used to construct the complete structure of endo- β -N-acetylglucosaminidases (Şahutoğlu et al., 2020). The 3D modeling was performed using HHpred web server and Modeller software (Musumeci et al., 2017). The evaluation of model structure is done usually by using computational programs, such as PROCHECK, Qualitative Model Energy Analysis (QMEAN), and Verify 3D, for checking its reliability for further analysis (Hemmati and Mehrabadi, 2020).

2.2.2.3 Phylogenetic analysis

The most widely used phylogenetic software is Molecular Evolutionary Genetics Analysis (MEGA) (Bunterngsook et al., 2021), PhyloPythiaS (Tan et al., 2016b), EMBL-EBI (Yata et al., 2021), and MAFFT server (Voß et al., 2020).

2.2.2.4 Functional analysis

The functional analysis helps to gain structural insights about the model structure related to the function. It involves analysis of parameters such as protein interaction, glycosylation, disulfide bond, conserved domain, conserved motif, catalytic mechanism, molecular dynamics (MD), and molecular docking. The study of enzyme mechanism involves molecular structure, free energy of activation, binding process, cleavage of chemical bonds, and molecular interactions. The information about the catalytic mechanism is collected in databases such as BRENDA, Enzyme Function Initiative (EFI), and ExPASy-ENZYME (Zhang et al., 2020a,b). The orientation of side chain of enzyme residues and the packing of polypeptide chain of enzymes helps in recognition of enzyme, and its binding and catalysis reaction (Priyadarshini and Singh, 2019). Sharma (2020) studied functional properties of stress-responsive acid phosphatase gene TaPase via ProFun server (Sharma, 2020). The melting temperature of a novel thermo/organo tolerant BOD was studied via SCooP servers (Sadeghian et al., 2020).

2.2.2.5 Protein interaction network

The Transmembrane Helices, Hidden Markov Model (TMHMM) Server was used to identify transmembrane motifs of insect trypsin enzymes. The trypsin interaction network was revealed by using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING). The information derived from these computational tools was used to design specific inhibitors to insect trypsin enzymes (Hemmati and Mehrabadi, 2020). STRING database is used to predict functional protein association network of novel abiotic stress proteins in *T. aestivum* (Gupta et al., 2018). STRING was used for studying the protein–protein network of HPs from the core genome of *C. pseudotuberculosis* biovar *ovis* (Araújo et al., 2020).

2.2.2.6 Conserved motif, domains, peptide, epitope, glycosylation, disulfide bonds, antigenicity, and localization

Hemmati and Mehrabadi (2020) discovered highly conserved peptide motifs in the insect trypsin enzyme via using the Multiple Em for Motif Elicitation (MEME) program (Hemmati and Mehrabadi, 2020). The SIB MyHits Motif scan and Pfam was used to find possible motifs in *Beauveria bassiana* chitinase (Bhagwat et al., 2021). Conserved Domain Database was used to predict the conserved domain of the unique endo- β -1,4-glucanase (Ariaeenejad et al., 2018). Similarly, the Conserved Domain Database was used to predict conserved domains in a novel α -amylase (Motahar et al., 2020). Functional domains and families were studied by comparing list of HPs against Pfam, CATH, SVM-Prot, CDART, SMART, and ProtoNet databases (Gupta et al., 2018). The *in silico* functional characterization and domain analysis of HPs involved in lipid accumulation were performed by using CATH, Pfam, protein analysis through evolutionary relationships (PANTHER), MEME, ProtoNet, Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The information about the structural properties was obtained via Self-Optimized Prediction Method with Alignment (SOPMA), SCRATCH, NetNGlyc, and GeneWise (Sahoo et al., 2020). The functional annotations of 172 HPs from *C. pseudotuberculosis* biovar *ovis* genomes were carried out with the help of GO FEAT tool (Araújo et al., 2020).

The signal peptide is recognized by using SignalP. For example, SignalP was used for identifying signal peptide in novel mesophilic phytase and novel AA9-type LPMO (Tan et al., 2016b; Bunterngsook et al., 2021). The B-cell epitope prediction was performed by BCPREDS (Guha et al., 2020). NetNGlyc1.0 server used for predicting for N-glycosylation sites and DiANNA 1.1 web server was used for predicting disulfide bonds in novel CalB-Type Lipase (Vaquero et al., 2015). The ANTIGENpro server was used to predict antigenicity. The binding prediction of HLA class II was performed

using the MHC II Analysis Resource at the Immune Epitope Database (IEDB) server (González-Torres et al., 2020). Plant-mPLoc, MultiLoc2, SherLoc2, WoLF Psort, and TargetP were used to find out localization of HPs (Gupta et al., 2018). Guha et al. (2020) used Psortdb and CELLO for analyzing for localization of HPs (Guha et al., 2020). Araújo et al. (2020) identified subcellular localization of HPs from the core genome of *C. pseudotuberculosis* biovar *ovis* via Psortdb (Araújo et al., 2020). Sahoo et al. (2020) used TargetP, ChloroP, WoLF PSORT, TMHMM server, SignalP for signal peptide recognition in HPs (Sahoo et al., 2020). Hidden Markov Model (TMHMM) tool is used to get information regarding membrane topology of 15 chitinases, and CELLO v.2.5. was used to determine their subcellular localization (Bhagwat et al., 2021).

2.2.2.7 Molecular dynamics simulation studies

MD simulation studies capture the atomic movements (from ps to ns) of enzyme or enzyme complexes. The studies provide the insights into enzyme structural flexibility and function. The GROMINGEN MACHINE for CHEMICAL SIMULATIONS (GROMACS), ASSISTED MODEL BUILDING AND ENERGY REFINEMENT (AMBER), and CHEMISTRY AT HARVARD MACROMOLECULAR MECHANICS (CHARMM) are most widely used software to perform MD simulations. It is a multistep procedure. It involves force field selection, topology file generation, system preparation for conducting simulation, energy minimization, equilibration, and analysis. There are two approaches, i.e., molecular mechanics (MM) or quantum mechanics (QM) (Zhang et al., 2020a,b). The MD simulations of Np-ADO and its mutants, Cts-ADO were performed at different temperatures. The thermally sensitive regions were recognized by the Root Mean Square Deviation (RMSD) and the Root Mean Square Fluctuation (RMSF) analysis (Shakeel et al., 2018). The stability of the binding of novel inhibitors of *Mycobacterium tuberculosis* (Mtb)-MurB oxidoreductase enzyme was supported by MD simulation data (Nirwan et al., 2021). The Nanoscale Molecular Dynamics (NAMD) software was used for MD simulation of two novel N-acetylglucosaminidase enzymes (Şahutoğlu et al., 2020).

The conformational stability, total energy of the system, and structure compactness of *C. sativus* Urease were evaluated by using MD simulation using the Sander module of AMBER (Yata et al., 2021). The catalytic mechanism of D-lactate dehydrogenase is revealed by adopting experimental and *in silico* approach, i.e., MD simulation (Zhang et al., 2020a,b).

2.2.2.8 Molecular docking

Molecular docking approach helps to gain insight into mechanisms of catalysis and inhibitor discovery (Zhang et al., 2020a,b). The Docking analysis of *B. bassiana* chitinases was performed using AutoDock Vina program (Bhagwat et al., 2021; Yata et al., 2021; Musumeci et al., 2017). Ligand docking with two novel N-acetyl-glucosaminidase enzymes was performed using SwissDock (Şahutoğlu et al., 2020). Nirwan et al., (2021) discovered 12 compounds from ZINC database that have potential to inhibit *M. tuberculosis* MurB (Nirwan et al., 2021).

3. Conclusion

The numbers of sequences in databases are increasing at an exponential rate with many incomplete annotations. There are about 40% of the sequences that fall into the category of uncharacterized protein while there are 20% of the sequences which are not annotated correctly in the databases. To overcome this limitation, the text-based approach on the basis of enzymatic function can be carried out in a few curated databases such as the CAZy database or the Lipase Engineering Database (LED) (Zaparucha et al., 2018). The MD simulation and molecular docking can be performed with protein candidates; however, it is a difficult task to select significant program, algorithms, and parameters to perform simulation (Zhang et al., 2020a,b). The *in silico* bioprospecting approach, which involves exploration of the existing databases and *in silico* characterization, can potentially reduce down burden of traditional screening process and drastically improve the novel enzyme discovery process. The protein engineering approach can certainly help in improving properties of the novel as well as existing proteins/enzymes.

References

- Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423.
- Andreeva, A., Kulesha, E., Gough, J., Murzin, A.G., 2020. The SCOP database in 2020: Expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res.* 48, D376–D382.
- Angural, S., Rana, M., Sharma, A., Warmoota, R., Puri, N., Gupta, N., 2020. Combinatorial biobleaching of mixedwood pulp with lignolytic and hemicellulolytic enzymes for paper making. *Indian J. Microbiol.* 60, 383–387.
- Araújo, C.L., Blanco, I., Souza, L., Tiwari, S., Pereira, L.C., Ghosh, P., Azevedo, V., Artur Silva, A.F., 2020. In silico functional prediction of hypothetical proteins from the core genome of *Corynebacterium pseudotuberculosis* biovar *ovis*. *Peer J.* 8, 1–24.

- Ariaeenejad, S., Nooshi-Nedamani, S., Rahban, M., Kavousi, K., Pirbalooti, A.G., Mirghaderi, S.S., Mohammadi, M., Mirzaei, M., Salekdeh, G.H., 2018. A novel high glucose-tolerant β -glucosidase: targeted computational approach for metagenomic screening. *Front. Bioeng. Biotechnol.* 10.
- Ariaeenejad, S., Sheykh Abdollahzadeh Mamaghani, A., Maleki, M., Kavousi, K., Foroozandeh Shahraki, M., Hosseini Salekdeh, G., 2020. A novel high performance in-silico screened metagenome-derived alkali-thermostable endo- β -1,4-glucanase for lignocellulosic biomass hydrolysis in the harsh conditions. *BMC Biotechnol.* 20, 1–13.
- Arnold, F.H., 2018. Directed evolution: bringing new chemistry to life. *Biocatalysis* 57, 4143–4148.
- Attwood, T.K., Beck, M.E., Bleasby, A.J., Parry-Smith, D.J., 1994. PRINTS - A database of protein motif fingerprints. *Nucleic Acids Res.* 22, 3590–3596.
- Baa-Puyoulet, P., Parisot, N., Febvay, G., Huerta-Cepas, J., Vellozo, A.F., Gabaldón, T., Calevro, F., Charles, H., Colella, S., 2016. ArthropodaCyc: A CycADS powered collection of BioCyc databases to analyse and compare metabolism of arthropods. *Database* 2016, 1–9.
- Bae, H.D., Yanke, L.J., Cheng, K.J., Selinger, L.B., 1999. A novel staining method for detecting phytase activity. *J. Microbiol. Methods* 39, 17–22.
- Bailey, T.L., Johnson, J., Grant, C.E., Noble, W.S., 2015. The MEME suite. *Nucleic Acids Res.* 43, W39–W49.
- Bairoch, A., 2000. The ENZYME database in 2000. *Nucleic Acids Res.* 28, 304–305.
- Banik, S., Biswas, S., Karmakar, S., 2018. Extraction, purification, and activity of protease from the leaves of *Moringa oleifera*. *F1000Research* 7.
- Basso, A., Serban, S., 2020. Overview of immobilized enzymes' applications in pharmaceutical, chemical, and food industry. In: Guisan, J., Bolivar, J., López-Gallego, F., Rocha-Martin, J. (Eds.), *Immobilization of Enzymes and Cells. Methods in Molecular Biology*. Humana Press Inc., pp. 27–63
- Bauer, R.A., Günther, S., Jansen, D., Heeger, C., Thaben, P.F., Preissner, R., 2009. SuperSite: Dictionary of metabolite and drug binding sites in proteins. *Nucleic Acids Res.* 37, 195–200.
- Benkert, P., Tosatto, S.C.E., Schomburg, D., 2007. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins Struct. Funct. Bioinf.* 71, 261–277.
- Berman, H., Henrick, K., Nakamura, H., Markley, J.L., 2007. The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* 35, 2006–2008.
- Bhagwat, P., Amobonye, A., Singh, S., Pillai, S., 2021. A Comparative Analysis of GH18 Chitinases and Their Isoforms from *Beauveria Bassiana*: An In-Silico Approach, *Process Biochemistry*. Elsevier Ltd.
- Bienert, S., Waterhouse, A., De Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T., 2017. The SWISS-MODEL repository-new features and functionality. *Nucleic Acids Res.* 45, D313–D319.
- Bleasby, A.J., Akrigg, D., Attwood, T.K., 1994. OWL - A non-redundant composite protein sequence database. *Nucleic Acids Res.* 22, 3574–3577.
- Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., Finn, R.D., 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354.
- Boland, M.J., Hesselink, P.G.M., Papamichael, N., Hustedt, H., 1991. Extractive purification of enzymes from animal tissue using aqueous two phase systems: pilot scale studies. *J. Biotechnol.* 19, 19–33.
- Bunterngsook, B., Mhuanong, W., Kanokratana, P., Iseki, Y., Watanabe, T., Champreda, V., 2021. Identification and characterization of a novel AA9-type lytic polysaccharide monooxygenase from a bagasse metagenome. *Appl. Microbiol. Biotechnol.* 105, 197–210.
- Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G.V., Christie, C.H., Dalenberg, K., Di Costanzo, L., Duarte, J.M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D.S., Ghosh, S., Green, R.K., Guranović, V., Guzenko, D., Hudson, B.P., Lawson, C.L., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Persikova, I., Randle, C., Rose, A., Rose, Y., Sali, A., Segura, J., Sekharan, M., Shao, C., Tao, Y.P., Voigt, M., Westbrook, J.D., Young, J.Y., Zardecki, C., Zhuravleva, M., 2021. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* 49, D437–D451.
- Caenepeel, S., Charydzak, G., Sudarsanam, S., Hunter, T., Manning, G., 2004. The mouse kinome: Discovery and comparative genomics of all mouse protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11707–11712.
- Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* 37, 233–238.
- Carbajosa, G., Trigo, A., Valencia, A., Cases, I., 2009. Bionemo: Molecular information on biodegradation metabolism. *Nucleic Acids Res.* 37, 598–602.
- Castillo Villamizar, G.A., Nacke, H., Boehning, M., Herz, K., Daniel, R., 2019. Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. *mBio* 10, 1–15.
- Chaudhary, N., Sharma, A.K., Agarwal, P., Gupta, A., Sharma, V.K., 2015. 16S classifier: A tool for fast and accurate taxonomic classification of 16S rRNA hypervariable regions in metagenomic datasets. *PLoS One* 10, 1–13.
- Choi, J., Détry, N., Kim, K.T., Asiegbu, F.O., Valkonen, J.P., Lee, Y.H., 2014. FPoxDB: Fungal peroxidase database for comparative genomics. *BMC Microbiol.* 14, 1–8.
- Choi, K.R., Jiao, S., Lee, S.Y., 2020. Metabolic engineering strategies toward production of biofuels. *Curr. Opin. Chem. Biol.* 59, 1–14.
- Coker, J.A., 2016. Extremophiles and biotechnology: current uses and prospects. *F1000Research* 5, 396.
- Colovos, C., Yeates, T.O., 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* 2, 1511–1519.
- Daniel, R., 2005. The metagenomics of soil. *Nat. Rev. Microbiol.* 3, 470–478.
- Datta, S., Rajnish, K.N., Samuel, M.S., Pugazlndhi, A., Selvarajan, E., 2020. Metagenomic applications in microbial diversity, bioremediation, pollution monitoring, enzyme and drug discovery. A review. *Environ. Chem. Lett.* 18, 1229–1241.

- Dawson, N.L., Lewis, T.E., Das, S., Lees, J.G., Lee, D., Ashford, P., Orengo, C.A., Sillitoe, I., 2017. CATH: An expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* 45, D289–D295.
- Deng, Y., Zhu, L., Cai, H., Wang, G., Liu, B., 2018. Autophagic compound database: a resource connecting autophagy-modulating compounds, their potential targets and relevant diseases. *Cell Prolif.* 51.
- Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J., Diella, F., 2011. Phospho. ELM: A database of phosphorylation sites-update 2011. *Nucleic Acids Res.* 39, 261–267.
- Duffeck, C.E., de Menezes, C.L.A., Boscolo, M., da Silva, R., Gomes, E., da Silva, R.R., 2020. *Citrobacter diversus*-derived keratinases and their potential application as detergent-compatible cloth-cleaning agents. *Braz. J. Microbiol.* 51, 969–977.
- Eisenberg, D., Lüthy, R., Bowie, J.U., 1997. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277, 396–404.
- Ekstrom, A., Taujale, R., McGinn, N., Yin, Y., 2014. PlantCAZyme: a database for plant carbohydrate-active enzymes. *Database (Oxford)* 2014, 1–8.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432.
- Ellis, L.B., Wackett, L.P., 2012. Use of the University of Minnesota Biocatalysis/Biodegradation Database for study of microbial degradation. *Microb. Inf. Exp.* 2, 1–10.
- Emanuelsson, O., Nielsen, H., Brunak, S., Von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016.
- Enzymes Market Size, 2020. Share & Trends Analysis Report by Application (Industrial Enzymes, Specialty Enzymes), by Product (Carbohydrase, Proteases, Lipases), by Source, by Region, and Segment Forecasts, 2020 - 2027. Grand view research.
- Eric Wommack, K., Bhavsar, J., Polson, S.W., Chen, J., Dumas, M., Srinivasiah, S., Furman, M., Jamindar, S., Nasko, D.J., 2012. VIROME: A standard operating procedure for analysis of viral metagenome sequences. *Stand. Genom. Sci.* 6, 427–439.
- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., Sali, A., 2006. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinf.* 1–47.
- Eyal, A., Brown, S., Almonacid, D.E., Barber, A.E., Huang, C.C., Lauck, F., Mashiyama, S.T., Meng, E.C., Mischel, D., Morris, J.H., Ojha, S., Schnoes, A.M., Stryke, D., Yunes, J.M., Ferrin, T.E., Holliday, G.L., Babbitt, P.C., 2014. The structure-function linkage database. *Nucleic Acids Res.* 42, 521–530.
- Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnao, V., D'Eustachio, P., Stein, L., Hermjakob, H., 2017. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinf.* 18, 1–9.
- Fischer, M., Pleiss, J., 2003. The lipase engineering database: A navigation and analysis tool for protein families. *Nucleic Acids Res.* 31, 319–321.
- Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K.B., Bairoch, A., Schomburg, D., Tipton, K.F., Apweiler, R., 2004. IntEnz, the integrated relational enzyme database. *Nucleic Acids Res.* 32, 434–437.
- Foong, C.P., Lakshmanan, M., Abe, H., Taylor, T.D., Foong, S.Y., Sudesh, K., 2018. A novel and wide substrate specific polyhydroxyalkanoate (PHA) synthase from unculturable bacteria found in mangrove soil. *J. Polym. Res.* 25, 23.
- Forster, S.C., Browne, H.P., Kumar, N., Hunt, M., Denise, H., Mitchell, A., Finn, R.D., Lawley, T.D., 2016. HPMCD: The database of human microbial communities from metagenomic datasets and microbial reference genomes. *Nucleic Acids Res.* 44, D604–D609.
- Gao, W., Wu, K., Chen, L., Fan, H., Zhao, Z., Gao, B., Wang, H., Wei, D., 2016. A novel esterase from a marine mud metagenomic library for biocatalytic synthesis of short - chain flavor esters. *Microb. Cell Factories* 15, 1–12.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. *Proteomics Protoc. Handb.* 571–607.
- Geourjon, C., Deléage, G., 1995. Sopma: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11, 681–684.
- Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., Whalen, K.L., 2015. Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta Protein Proteonom.* 1854, 1019–1037.
- Góngora-Castillo, E., López-Ochoa, L.A., Apolinar-Hernández, M.M., Caamal-Pech, A.M., Contreras-de la Rosa, P.A., Quiroz-Moreno, A., Ramírez-Prado, J.H., O'Connor-Sánchez, A., 2020. Data mining of metagenomes to find novel enzymes: a non-computationally intensive method. *3 Biotech* 10, 1–8.
- González-Torres, I., Perez-Rueda, E., Evangelista-Martínez, Z., Zárate-Romero, A., Moreno-Enríquez, A., Huerta-Saquero, A., 2020. Identification of L-asparaginases from *Streptomyces* strains with competitive activity and immunogenic profiles: a bioinformatic approach. *Peer J.* 8, 1–23.
- Grahame, D.S.A., 2016. Exploring Pepsin's Alkaline Instability via Bioinformatic Analysis and a Rationale Protein Design Approach.
- Gromiha, M.M., An, J., Kono, H., Oobatake, M., Uedaira, H., Prabakaran, P., Sarai, A., 2000. ProTherm, version 2.0: Thermodynamic database for proteins and mutants. *Nucleic Acids Res.* 28, 283–285.
- Guha, S., Das, S., Ganguli, S., 2020. A comparative genomics pipeline for *in silico* characterization and functional annotation of short hypothetical proteins. *J. Trop. Life Sci.* 10, 141–148.
- Gupta, S., Singh, Y., Kumar, H., Raj, U., Rao, A.R., Varadwaj, P.K., 2018. Identification of novel abiotic stress proteins in *Triticum aestivum* through functional annotation of hypothetical proteins. *Interdiscipl. Sci. Comput. Life Sci.* 10, 205–220.
- Haft, D.H., Selengut, J.D., Richter, R.A., Harkins, D., Basu, M.K., Beck, E., 2013. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, 387–395.



- Handa, V., Sharma, D., Kaur, A., Arya, S.K., 2020. Biotechnological applications of microbial phytase and phytic acid in food and feed industries. *Biocatal. Agric. Biotechnol.* 25.
- Haynes, M., 2013. Function-based metagenomic library screening and heterologous expression strategy for genes encoding phosphatase activity. In: *Brenner's Encyclopedia of Genetics*, second ed., pp. 378–381.
- Heinegård, D., Paulsson, M., 1987. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Methods Enzymol.* 145, 336–363.
- Hemmati, S.A., Mehrabadi, M., 2020. Structural ensemble-based computational analysis of trypsin enzyme genes discovered highly conserved peptide motifs in insects. *Arch. Phytopathol. Plant Protect.* 53, 335–354.
- Igarashi, Y., Heureux, E., Doctor, K.S., Talwar, P., Gramatikova, S., Gramatikoff, K., Zhang, Y., Blinov, M., Ibragimova, S.S., Boyd, S., Ratnikov, B., Cieplak, P., Godzik, A., Smith, J.W., Osterman, A.L., Eroshkin, A.M., 2009. PMAP: Databases for analyzing proteolytic events and pathways. *Nucleic Acids Res.* 37, 611–618.
- Kalathur, R.K.R., Pinto, J.P., Hernández-Prieto, M.A., MacHado, R.S.R., Almeida, D., Chaurasia, G., Futschik, M.E., 2014. UniHI 7: An enhanced database for retrieval and interactive analysis of human molecular interaction networks. *Nucleic Acids Res.* 42, 408–414.
- Kalia, V.C., 2017. Mining metagenomes for novel bioactive molecules. In: *Mining of Microbial Wealth and MetaGenomics*, pp. 1–9.
- Kamble, A., Singh, H., 2020. Different Methods of Soil DNA Extraction. *Bio-101 e3521*, pp. 1–23.
- Kamble, A., Srinivasan, S., Singh, H., 2019. In-silico bioprospecting: finding better enzymes. *Mol. Biotechnol.* 61, 53–59.
- Kamble, A., Sawant, S., Singh, H., 2020. 16S ribosomal RNA gene - based metagenomics: a review. *Biomed. Res. J.* 5–11.
- Karp, P.D., Riley, M., Paley, S.M., Pellegrini-Toole, A., 2002. The MetaCyc database. *Nucleic Acids Res.* 30, 59–61.
- Keegan, K.P., Glass, E.M., Meyer, F., 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function. In: *Methods in Molecular Biology*. Humana Press Inc., pp. 207–233
- Kim, M.S., Lei, X.G., 2008. Enhancing thermostability of *Escherichia coli* phytase AppA2 by error-prone PCR. *Appl. Microbiol. Biotechnol.* 79, 69–75.
- Klimke, W., Agarwala, R., Badretin, A., Chetvermin, S., Ciufu, S., Fedorov, B., Kiryutin, B., O'Neill, K., Resch, W., Resenchuk, S., Schafer, S., Tolstoy, I., Tatusova, T., 2009. The National Center for biotechnology information's protein clusters database. *Nucleic Acids Res.* 37, 216–223.
- Knudsen, M., Wiuf, C., 2010. The CATH database. *Hum. Genom.* 4, 207–212.
- Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* 305, 567–580.
- Kuhn, M., von Mering, C., Campillos, M., Jensen, L.J., Bork, P., 2008. STITCH: Interaction networks of chemicals and proteins. *Nucleic Acids Res.* 36, 684–688.
- Kumar, D., Rajesh, S., Balashanmugam, P., Rebecca, L.J., Kalaichelvan, P.T., 2013. Screening, optimization and application of extracellular phytase from *Bacillus megaterium* isolated from poultry waste. *J. Mod. Biotechnol.* 2, 46–52.
- Li, X., Zhang, Z., Song, J., 2012b. Computational enzyme design approaches with significant biological outcomes: progress and challenges. *Comput. Struct. Biotechnol. J.* 2, e201209007.
- Li, Y., Li, S., Thodey, K., Trenchard, I., Cravens, A., Smolke, C.D., 2018. Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. USA* 115, E3922–E3931.
- Li, S., Yang, X., Yang, S., Zhu, M., Wang, X., 2012a. Technology prospecting on enzymes: application, marketing and engineering. *Comput. Struct. Biotechnol.* 2, 1–11.
- Lang, M., Stelzer, M., Schomburg, D., 2011. BKM-react, an integrated biochemical reaction database. *BMC Biochem.* 12.
- Laskowski, R.A., Jablńska, J., Pravda, L., Vařeková, R.S., Thornton, J.M., 2018. PDBsum: Structural summaries of PDB entries. *Protein Sci.* 27, 129–134.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., Chatonnet, A., 2013. ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: Tools to explore diversity of functions. *Nucleic Acids Res.* 41, 423–429.
- Lespinet, O., Labedan, B., 2006. ORENZA: A web resource for studying ORphan ENZyme activities. *BMC Bioinf.* 7, 1–11.
- Li, L.L., McCorkle, S.R., Monchy, S., Taghavi, S., van der Lelie, D., 2009. Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol. Biofuels* 2, 1–11.
- Li, J., Li, X., Gai, Y., Sun, Y., Zhang, D., 2019. Evolution of *E. coli* phytase for increased thermostability guided by rational parameters. *J. Microbiol. Biotechnol.* 29, 419–428.
- Lingner, T., Abhauer, K.P., Schreiber, F., Meinicke, P., 2011. CoMet - A web server for comparative functional profiling of metagenomes. *Nucleic Acids Res.* 39, 518–523.
- López-López, O., Cerdán, M., González-Siso, M., 2013. Hot spring metagenomics. *Life* 3, 308–320.
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268.
- Madej, T., Addess, K.J., Fong, J.H., Geer, L.Y., Geer, R.C., Lanczycki, C.J., Liu, C., Lu, S., Marchler-Bauer, A., Panchenko, A.R., Chen, J., Thiessen, P.A., Wang, Y., Zhang, D., Bryant, S.H., 2012. MMDB: 3D structures and macromolecular interactions. *Nucleic Acids Res.* 40, 461–464.
- Magrane, M., Consortium, U.P., 2011. UniProt Knowledge base: A hub of integrated protein data. *Database* 2011, 1–13.

- Maleki, M., Shahraki, M.F., Kavousi, K., Ariaeenejad, S., Hosseini Salekdeh, G., 2020. A novel thermostable cellulase cocktail enhances lignocellulosic bioconversion and biorefining in a broad range of pH. *Int. J. Biol. Macromol.* 154, 349–360.
- Margolin, A.L., 1993. Enzymes in the synthesis of chiral drugs. *Enzym. Microb. Technol.* 15, 266–280.
- McDonald, A.G., Boyce, S., Tipton, K.F., 2009. ExplorEnz: The primary source of the IUBMB enzyme list. *Nucleic Acids Res.* 37, 593–597.
- McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405.
- Méndez-García, C., Bargiela, R., Martínez-Martínez, M., Ferrer, M., 2018. Metagenomic protocols and strategies. In: Nagarajan, M. (Ed.), *Metagenomics: Perspectives, Methods, and Applications*. Elsevier Inc., pp. 15–54
- Mitchell, A.L., Scheremetjew, M., Denise, H., Potter, S., Tarkowska, A., Qureshi, M., Salazar, G.A., Pesseat, S., Boland, M.A., Hunter, F.M.I., Ten Hoopen, P., Alako, B., Amid, C., Wilkinson, D.J., Curtis, T.P., Cochrane, G., Finn, R.D., 2018. EBI Metagenomics in 2017: Enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res.* 46, D726–D735.
- Mosca, R., Céol, A., Stein, A., Olivella, R., Aloy, P., 2014. 3did: A catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 42, 374–379.
- Motahar, S.F.S., Khatibi, A., Salami, M., Ariaeenejad, S., Emam-Djomeh, Z., Nedaei, H., Kavousi, K., Sheykhabdolazadeh Mamaghani, A., Salekdeh, G.H., 2020. A novel metagenome-derived thermostable and poultry feed compatible α -amylase with enhanced biodegradation properties. *Int. J. Biol. Macromol.* 164, 2124–2133.
- Musumeci, M.A., Lozada, M., Rial, D.V., Cormack, W.P.M., Jansson, J.K., Sjöling, S., Carroll, J.L., Dionisi, H.M., 2017. Prospecting biotechnologically-relevant monooxygenases from cold sediment metagenomes: an *in silico* approach. *Mar. Drugs* 15, 2–19.
- Nagano, N., 2005. EzCatDB: The enzyme catalytic-mechanism database. *Nucleic Acids Res.* 33, 407–412.
- Naowarojna, N., Cheng, R., Lopez, J., Wong, C., Qiao, L., Liu, P., 2021. Chemical modifications of proteins and their applications in metalloenzyme studies. *Synth. Syst. Biotechnol.* 6, 32–49.
- Nasko, D.J., Wommack, K.E., Ferrell, B.D., Polson, S.W., 2018. Fast and sensitive protein sequence homology searches using hierarchical cluster BLAST. *bioRxiv*. 19711, 1–23.
- Nasri, N., Kaddour, R., Rabhi, M., Plassard, C., Lachaal, M., 2011. Effect of salinity on germination, phytase activity and phytate content in lettuce seedling. *Acta Physiol. Plant.* 33, 935–942.
- Narayan, N.R., Weinmaier, T., Laserna-Mendieta, E.J., Claesson, M.J., Shanahan, F., Dabbagh, K., Iwai, S., Desantis, T.Z., 2020. Piphillin predicts metagenomic composition and dynamics from DADA2-corrected 16S rDNA sequences (*BMC Genom.* (2020) 21 (56), 1-2). <https://doi.org/10.1186/s12864-019-6427-1>.
- Nirwan, S., Chahal, V., Kakkar, R., 2021. Structure-based virtual screening, free energy of binding and molecular dynamics simulations to propose novel inhibitors of Mtb-MurB oxidoreductase enzyme. *J. Biomol. Struct. Dyn.* 39 (2), 656–671. <https://doi.org/10.1080/07391102.2020.1712258>.
- Pace, N.R., Stahl, D.A., Lane, D.J., Olsen, G.J., 1986. The analysis of natural microbial populations by ribosomal RNA sequences. In: Marshall, K.C. (Ed.), *Advances in Microbial Ecology*. Springer, Boston, MA, pp. 1–55.
- Passardi, F., Theiler, G., Zamocky, M., Cosio, C., Rouhier, N., Teixeira, F., Margis-Pinheiro, M., Ioannidis, V., Penel, C., Falquet, L., Dunand, C., 2007. PeroxiBase: The peroxidase database. *Phytochemistry* 68, 1605–1611.
- Peitsch, M.C., Timothy, N.C.W., Dave, R.S., Joel, L.S., 1995. The Swiss-3D image collection and PDB-Browser on the World-Wide Web. *Trends Biochem. Sci.* 20 (2), 82–84.
- Poddar, A., Das, S.K., 2018. Microbiological studies of hot springs in India: a review. *Arch. Microbiol.* 200, 1–18. <https://doi.org/10.1007/s00203-017-1429-3>.
- Priyadarshini, P., Singh, B., 2019. *Computational Resources and Techniques in Enzyme Research, Biomass, Biofuels, Biochemicals: Advances in Enzyme Technology*. Elsevier B.V.
- Priyodip, P., Balaji, S., 2018. Microbial degradation of myo-inositol hexakisphosphate (IP6): specificity, kinetics, and simulation. *3 Biotech* 8.
- Pruitt, K.D., Tatusova, T., Brown, G.R., Maglott, D.R., 2012. NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Res.* 40, 130–135.
- Rappoport, N., Karsenty, S., Stern, A., Linial, N., Linial, M., 2012. ProtoNet 6.0: Organizing 10 million protein sequences in a compact hierarchical family tree. *Nucleic Acids Res.* 40, 313–320.
- Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., Finn, R.D., 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632.
- Reetz, M.T., Carballeira, J.D., Vogel, A., 2006. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem.* 118, 7909–7915.
- Ribeiro, A.J.M., Holliday, G.L., Furnham, N., Tyzack, J.D., Ferris, K., Thornton, J.M., 2018. Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* 46, D618–D623.
- Roberts, R.J., Vincze, T., Posfai, J., Macelis, D., 2015. REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* 43, D298–D299.
- Rolf, J., Siedentop, R., Lütz, S., Rosenthal, K., 2020. Screening and identification of novel cGAS homologues using a combination of in vitro and in vivo protein synthesis. *Int. J. Mol. Sci.* 21.
- Rosen, G.L., Reichenberger, E.R., Rosenfeld, A.M., 2011. NBC: The naïve Bayes classification tool webserver for taxonomic classification of metagenomic reads. *Bioinformatics* 27, 127–129.
- Sadeghian, I., Rezaie, Z., Rahmatbadi, S.S., Hemmati, S., 2020. Biochemical insights into a novel thermo/organo tolerant bilirubin oxidase from *Thermosediminibacter oceani* and its application in dye decolorization. *Process Biochem.* 88, 38–50.

- Sahoo, S., Mahapatra, S.R., Das, N., Parida, B.K., Rath, S., Misra, N., Suar, M., 2020. Functional elucidation of hypothetical proteins associated with lipid accumulation: prioritizing genetic engineering targets for improved algal biofuel production. *Algal. Res.* 47, 101887.
- Şahutoğlu, A.S., Duman, H., Frese, S.A., Karav, S., 2020. Structural insights of two novel N-acetyl-glucosaminidase enzymes through *in silico* methods. *Turk. J. Chem.* 44, 1703–1712.
- Sajidan, Wulandari, R., Sari, E.N., Ratriyanto, A., Weldekiros, H., Greiner, R., 2015. Phytase-producing bacteria from extreme regions in Indonesia. *Braz. Arch. Biol. Technol.* 58, 711–717.
- Sayers, E.W., Beck, J., Bolton, E.E., Bourexis, D., Brister, J.R., Canese, K., Comeau, D.C., Funk, K., Kim, S., Klimke, W., Marchler-Bauer, A., Landrum, M., Lathrop, S., Lu, Z., Madden, T.L., O’Leary, N., Phan, L., Rangwala, S.H., Schneider, V.A., Skripchenko, Y., Wang, J., Ye, J., Trawick, B.W., Pruitt, K.D., Sherry, S.T., 2021. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 49, D10–D17.
- 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.
- Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., Schomburg, D., 2004. BRENDA, the enzyme database: Updates and major new developments. *Nucleic Acids Res.* 32, D431–D433.
- Shakeel, T., Gupta, M., Fatma, Z., Kumar, R., Kumar, R., Singh, R., Sharma, M., Jade, D., Gupta, D., Fatma, T., Yazdani, S.S., 2018. A consensus-guided approach yields a heat-stable alkane-producing enzyme and identifies residues promoting thermostability. *J. Biol. Chem.* 1–30.
- Sharma, A.D.E.V., 2020. Molecular modeling and 3D analysis of water stress responsive tapase phosphatase encoding gene in wheat (*Triticum aestivum*). *Agriculture* 4.
- Sharma, V.K., Kumar, N., Prakash, T., Taylor, T.D., 2009. MetaBioME: A database to explore commercially useful enzymes in metagenomic datasets. *Nucleic Acids Res.* 38, 468–472.
- Sharma, N., Thakur, N., Raj, T., SavitriBhalla, T.C., 2017. Mining of microbial genomes for the novel sources of nitrilases. *Biomed. Res. Int.* 1–14.
- Shukla, P., 2020. *Microbial Enzymes and Biotechniques - Interdisciplinary Perspectives*. Springer Singapore.
- Smith, A.A., Caruso, A., 2013. Characterization and Homology Modeling of a Cyanobacterial Phosphoenolpyruvate Carboxykinase Enzyme. *Hindawi*.
- Smith, R.D., Clark, J.J., Ahmed, A., Orban, Z.J., Dunbar, J.B., Carlson, H.A., 2019. Updates to binding MOAD (Mother of All Databases): Poly-pharmacology tools and their utility in drug repurposing. *J. Mol. Biol.* 431, 2423–2433.
- Song, J., Tan, H., Perry, A.J., Akutsu, T., Webb, G.I., Whisstock, J.C., Pike, R.N., 2012. PROSPER: An integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One* 7.
- Stewart, E.J., 2012. Growing unculturable bacteria. *J. Bacteriol.* 194, 4151–4160.
- Suleimanova, A.D., Beinhauer, A., Valeeva, L.R., Chastukhina, I.B., Balaban, N.P., Shakirov, E.V., Greiner, R., Sharipova, M.R., 2015. Novel glucose-1-phosphatase with high phytase activity and unusual metal ion activation from soil bacterium *Pantoea* sp. strain 3.5.1. *Appl. Environ. Microbiol.* 81, 6790–6799.
- Sunar, K., Kumar, U., Deshmukh, S.K., 2016. Recent applications of enzymes in personal care products. In: Singh Dhillon, G., Kaur, S. (Eds.), *Agro-Industrial Wastes as Feedstock for Enzyme Production: Apply and Exploit the Emerging and Valuable Use Options of Waste Biomass*. Elsevier Inc., pp. 279–298.
- Tan, H., Miao, R., Liu, T., Cao, X., Wu, X., Xie, L., Huang, Z., Peng, W., Gan, B., 2016a. Enhancing the thermal resistance of a novel acidobacteria-derived phytase by engineering of disulfide bridges. *J. Microbiol. Biotechnol.* 26, 1717–1722.
- Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., Gan, B., 2016b. Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Appl. Microbiol. Biotechnol.* 100, 2225–2241.
- Tan, H., Chen, X., Liang, N., Chen, R., Chen, J., Hu, C., Li, Q., Li, Q., Pei, W., Xiao, W., Yuan, Y., Chen, W., Zhang, L., 2019. Transcriptome analysis reveals novel enzymes for apo-carotenoid biosynthesis in saffron and allows construction of a pathway for crocetin synthesis in yeast. *J. Exp. Bot.* 70, 4819–4833.
- Thornbury, M., Sicheri, J., Guinard, C., Mahoney, D., Routledge, F., Curry, M., Elaghil, M., Boudreau, N., Tsai, A., Slaine, P., Getz, L., Cook, J., Rohde, J., McCormick, C., Sicheri, C.J., Thornbury, M., Guinard, C., Slaine, S.P., Getz, L., Finlayson-trick, E., Assay, E., Sicheri, J., Thornbury, M., Cook, S.J., Getz, L., 2018. Discovery and Characterization of Novel Lignocellulose-Degrading Enzymes from the Porcupine Microbiome. *bioRxiv*.
- Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J., 2018. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* 46, W363–W367.
- Touw, W.G., Baakman, C., Black, J., Te Beek, T.A.H., Krieger, E., Joosten, R.P., Vriend, G., 2015. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* 43, D364–D368.
- Toyama, D., de Moraes, M.A.B., Ramos, F.C., Zanphorlin, L.M., Tonoli, C.C.C., Balula, A.F., de Miranda, F.P., Almeida, V.M., Marana, S.R., Ruller, R., Murakami, M.T., Henrique-Silva, F., 2018. A novel β -glucosidase isolated from the microbial metagenome of Lake Poraquê (Amazon, Brazil). *Biochim. Biophys. Acta Protein Proteonom.* 1866, 569–579.
- Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C.F., Tolmie, D.E., Kent Wenger, R., Yao, H., Markley, J.L., 2008. BioMagResBank. *Nucleic Acids Res.* 36, 402–408.
- Vaquero, M.E., De Eugenio, L.I., Martínez, M.J., Barriuso, J., 2015. A novel CalB-type lipase discovered by fungal genomes mining. *PLoS One* 10, 1–11.
- Verma, V., Joshi, S., Choudhary, M., Srivastava, N., 2021. Enzymes in textile industries. In: Thatoi, H., M.S., D.S.K. (Eds.), *Bioprospecting of Enzymes in Industry, Healthcare and Sustainable Environment*. Springer, Singapore, pp. 383–394.

- Vieira, E.F., Delerue-Matos, C., 2020. Exploitation of *Saccharomyces cerevisiae* enzymes in food processing and preparation of nutraceuticals and pharmaceuticals. In: Arora, N., Mishra, J., M., V. (Eds.), *Microbial Enzymes: Roles and Applications in Industries*. Springer, Singapore, pp. 41–62.
- Vishnoi, N., Dixit, S., Mishra, J., 2020. Microbial lipases and their versatile applications. In: Arora, N.K., Mishra, J., Mishra, V. (Eds.), *Microbial Enzymes: Roles and Applications in Industries, Microorganisms for Sustainability*. Springer Singapore, Singapore, pp. 207–230.
- Voß, H., Heck, A., Schallmey, M., 2020. Database mining for novel bacterial beta-etherases, glutathione-dependent lignin-degrading enzymes. *Appl. Environ. Microbiol.* 86, 1–15.
- Wang, C.Y., Chang, P.M., Ary, M.L., Allen, B.D., Chica, R.A., Mayo, S.L., Olafson, B.D., 2018. ProtaBank: A repository for protein design and engineering data. *Protein Sci.* 27, 1113–1124.
- Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA* 95, 6578–6583.
- Wiederstein, M., Sippl, M.J., 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, 407–410.
- Wittig, U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J., Rojas, I., 2006. SABIO-RK: Integration and curation of reaction kinetics data. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 4075, 94–103.
- Wu, C.H., Huang, H., Nikolskaya, A., Hu, Z., Barker, W.C., 2004. The iProClass integrated database for protein functional analysis. *Comput. Biol. Chem.* 28, 87–96.
- Wu, S., Zhu, Z., Fu, L., Niu, B., Li, W., 2011. WebMGA: A customizable web server for fast metagenomic sequence analysis. *BMC Genom.* 12.
- Xu, W., Shao, R., Wang, Z., Yan, X., 2015. Improving the neutral phytase activity from *Bacillus amyloliquefaciens* DSM 1061 by site-directed mutagenesis. *Appl. Biochem. Biotechnol.* 175 (6), 3184–3194.
- Yao, B., Fan, A.E.Y., 2007. A novel phytase appA from *Citrobacter amalonaticus* CGMCC 1696: gene cloning and overexpression in *Pichia pastoris*. *Curr. Microbiol.* 55, 185–192.
- Yata, V.K., Biswas, A.D., Deb, A., Sanjeev, A., Mattaparthi, V.S.K., 2021. Identification of cucumis sativus urease as a potential urea binding enzyme by computational methods. *Biointerface Res. Appl. Chem.* 11, 9184–9200.
- Yellaboina, S., Tasneem, A., Zaykin, D.V., Raghavachari, B., Jothi, R., 2011. DOMINE: A comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.* 39, 730–735.
- Yi, Y., Fang, Y., Wu, K., Liu, Y., Zhang, W., 2020. KEGG: Kyoto encyclopedia of genes and genomes. *Oncol. Lett.* 19, 3316–3332.
- Yu, C., Zavaljevski, N., Desai, V., Reifman, J., 2008. Genome-wide enzyme annotation with precision control: Catalytic families (CatFam) databases. *Proteins Struct. Funct. Bioinf.* 74, 449–460.
- Zaparucha, A., De Berardinis, V., Vaxelaire-Vergne, C., 2018. Chapter 1: genome mining for enzyme discovery. In: *RSC Catalysis Series*. Royal Society of Chemistry, pp. 3–27.
- Zhang, Y., Geary, T., Simpson, B.K., 2019. Genetically modified food enzymes: a review. *Curr. Opin. Food Sci.* 25, 14–18.
- Zhang, J.-Z., Fu, L.-H., Tang, T., Zhang, S.-Y., Zhu, J., Li, T., Wang Zi-ning, S.T., 2020. Scalable enzyme mining via synthetic biology. *Synth. Biol. J.* 1–18.
- Zhang, Y., Aryee, A.N., Simpson, B.K., 2020b. Current role of *in silico* approaches for food enzymes. *Curr. Opin. Food Sci.* 31, 63–70.

**SYNOPSIS OF THE THESIS TO BE SUBMITTED
TO THE UNIVERSITY OF NMIMS (DEEMED-TO-BE-UNIVERSITY)
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (BIOLOGICAL SCIENCES)**

NAME OF CANDIDATE:	Asmita Deepak Kamble
DOCTORAL COMMITTEE:	I. Dr. Kayzad Nilgiriwala, Research Officer, Foundation of Medical Research, Mumbai II. Dr. Binoj Kutty Assistant Professor, St. Xavier's College, Mumbai
NAME AND DESIGNATION OF MENTOR:	Dr. Harinder Singh Assistant Professor, Sunandan Divatia School of Science, SVKM's NMIMS University, Vile Parle (W), Mumbai - 400 056
TITLE OF THESIS:	Understanding thermostability of phytase: Bioprospecting and Modulation
PLACE OF RESEARCH	Sunandan Divatia School of Science, SVKM's NMIMS University, Vile Parle (W), Mumbai- 400 056
NUMBER AND DATE OF REGISTRATION:	75102150002/26 th September 2017
DATE OF SUBMISSION OF SYNOPSIS:	18-July-2023
SIGNATURE OF THE CANDIDATE:	
SIGNATURE OF THE MENTOR:	

INDEX

Sr. No.	Table of Contents	Page Number
1	Chapter 1: Introduction	1-2
2	Chapter 2: Review of Literature	3-6
3	Chapter 3: Lacunae, Rationale, Aim and Objectives	7-9
4	Chapter 4: Isolation and screening of phytase producing microbes from natural source or <i>In-silico</i> bioprospecting	10-11
5	Chapter 5: <i>In-silico</i> characterization and experimental validation of shortlisted phytase	12-13
6	Chapter 6: <i>In-silico</i> identification of Hotspot and experimental validation	14-15
7	Chapter 7: Discussion	16-20
8	Chapter 8: Summary and Conclusion	21
9	Chapter 9: Significance of the study	22
10	Publications and Poster Presentations	23
11	References	24-31

Chapter 1: Introduction

Phytate, also known as myoinositol hexakisphosphate, cannot be digested by monogastric animals because they lack the phytase enzyme. Phytate accounts for 80% of the total organic phosphate pool in soil (Menezes-Blackburn et al. 2013). In the soil, phytate forms complexes and binds to metal ions that are crucial for soil fertility and plant growth. It binds to various metal ions such as calcium, iron, zinc, and magnesium, as well as trace minerals like manganese, copper, and molybdenum, thereby reducing their availability. As a result, the nutritional value of the diet decreases. To compensate for this, feed and food preparations for animal growth require supplementation of inorganic phosphate along with essential ions like calcium and iron (Goutami Banerjee, Khin Oo, XIyun Zhang, Jie Yang 2017). When phytate binds to minerals, it forms mineral phytate complexes that remain unabsorbed in the gut and are excreted without being utilized. This indigestible phytate, along with excessive phosphate supplementation in feed, can contribute to environmental pollution (Vashishth, Ram, and Beniwal 2017). Phytate also interacts with proteins, forming complexes through electrostatic interactions. This interaction affects the structure, activity, solubility, and digestibility of proteins, making them less accessible to proteases and resulting in inefficient protein digestion (Vashishth, Ram, and Beniwal 2017). Additionally, phytate binds to enzymes such as trypsin, pepsin, α -amylase, and β -galactosidase, reducing their activity and availability. It also forms complexes with starch and lipids, resulting in starch complexes and lipo-phytic acids, respectively (Goutami Banerjee, Khin Oo, XIyun Zhang, Jie Yang 2017). The degradation of phytate is important for several reasons. Firstly, it can reduce feed costs, which account for a significant portion (60-70%) of livestock farming expenses, by minimizing the need for inorganic phosphorous supplementation (Becker 2008). Phytate serves as a rich source of inorganic phosphate. Secondly, phytate degradation can increase the bioavailability of micronutrients in animal feed, thereby enhancing its nutritional value and promoting better growth performance in the poultry industry. Apart from its significance in the feed industry, phytate degradation is also crucial for improving soil fertility, plant growth, and bioethanol production (Bhavsar and Khire 2014).

Phytase enzymes, known as myoinositol hexakisphosphate phosphohydrolases, act on phytate complexes as their substrate. They catalyze the breakdown of O-P (oxygen-phosphorous) bonds in phytate, leading to the release of various products such as inositol, intermediate derivatives of

inositol, phosphate, and other micronutrients (Vashishth, Ram, and Beniwal 2017). Phytases are classified based on stereospecificity of phytate hydrolysis (3/1 or 4/6), sources of isolation, optimum pH (acidic/alkaline), and catalytic mechanisms (histidine acid phytases, β -propeller, cysteine phosphatases, and purple acid phosphatases) (Menezes-Blackburn et al. 2013). Phytases are mainly isolated from plant and microbial sources (Vikas Kumar and Sinha 2018). Phytases find extensive use in various sectors such as the feed and food industry, aquaculture, bioethanol production, the reduction of myoinositol phosphates, and even in soil improvement (Bhavsar and Khire 2014) (Vikas Kumar and Sinha 2018).

The utilization of phytase in the feed and food industry, bioethanol, and myoinositol phosphates production process are restricted due to its reduced effectiveness during high-temperature processing stages (Rebello et al. 2017) (Vikas Kumar and Sinha 2018) (Makolomakwa et al. 2017). The application of phytases is also limited in the soil amendment process as it depends upon various factors such as soil environment (pH, temperature, presence of metal ions), bioavailability of phytate, or presence of inhibitors in soil (Menezes-Blackburn et al. 2013). Hence there is a constant need to explore novel stable phytases by conventional and computational approaches (Tan, Wu, et al. 2016a) (Castillo Villamizar et al. 2019).

In general, the conventional method involves searching for new enzymes in natural sources (Castillo Villamizar et al. 2019), while the computational method utilizes existing databases to identify enzymes (Tan, Wu, et al. 2016a). These novel enzymes are further subjected to protein engineering approaches (directed evolution/ rational/ semi-rational approach) to enhance their properties (Zhang et al. 2019) (K. Kumar et al. 2015) (Sadeghian et al. 2020) (Soh et al. 2017) (Chen et al. 2015).

In this context, the present study focused on bioprospecting novel phytases through conventional, metagenomics, and *in-silico* approaches. The identified novel phytases will then be further examined and modified through rational engineering techniques to gain a better understanding as well as improve their stability.

Chapter 2: Review of literature

Looking at the worldwide demand for phytases in various industries and their limitations due to their biochemical properties, there is a need to search for novel phytases as well as improve their features (Mrudula Vasudevan et al. 2019) (Abdollahi, Ravindran, and Svihus 2013).

2.1 Bioprospecting of novel phytases

Researchers have adopted different methods to explore novel phytases. These methods included the use of natural sources (conventional), existing databases (*in-silico* bioprospecting), and metagenomics methods.

2.1.1 Conventional bioprospecting of phytases

A few studies that adopted conventional methods to explore phytases are highlighted here: Kumar et.al. isolated phytase producers from an environmental sample i.e., Himalayan soils, and identified them as *Acromobacter* sp. PB-01, *Tetra- thiobacter* sp. PB-03 and *Bacillus* sp. PB-13. They helped in the growth and phosphorous uptake of *Brassica juncea* (Vinod Kumar et al. 2013). Puppala et.al. characterized a novel acidic and thermostable phytase isolated from *Streptomyces* sp. (NCIM 5533) from Sanjivani islands Maharashtra, India. It also acts as a plant growth-promoting bacteria which was evident as it supports the growth of *Solanum lycopersicum* (Puppala et al. 2019). Amritha and Venkateswaran reported a phytase from *Lactobacillus plantarum* MTCC 1325 that can degrade phytate from sorghum. It also tolerates acid, bile and retains activity after exposure to a simulated gastrointestinal fluid environment (Amritha and Venkateswaran 2017). Mandviwala and Khire isolated thermostable phytase from *A.niger* NICM 564 (optimum activity at 50°C, optimum pH 5) (Mandviwala and Khire 2000). Sajidan et.al. explored soil samples of volcanic areas in Central Java for phytase producers. These phytases were found to belong to the *Bacillus* genus (*B. cereus*, *B. aryabhatai*, *B. psychrotolerans*) with varied ranges of optimum temperature (50- 60°C) (Sajidan et al. 2015).

In general, the conventional bioprospecting approach favors the cultivation of microbes that can be grown in the laboratory. These microbial techniques are tailored to accommodate only a small portion of the microbial community found in natural resources. Merely 1% of the soil microbial community can be cultured, leaving the remaining 99% unexplored (A. Kamble, Srinivasan, and

Singh 2019a) (A. D. Kamble and Singh 2022). To overcome the limitations of conventional methods, a metagenomic approach has been adopted by a few researchers to find novel phytases.

2.1.2 Bioprospecting of phytase through functional metagenomic approach:

Farias et.al. constructed a metagenomic library using DNA extracted from red rice crop residues mixed with castor bean cake. They employed a functional approach to screen for new phytase enzymes. One of the clones obtained from this library, called PhyRC001, displayed phytase activity. Through biochemical characterization, it was found that PhyRC001 exhibited optimal phytase activity at a pH of 7 and a temperature of 35°C (Farias, Almeida, and Meneses 2018). Functional metagenomics played a crucial role in investigating the unexplored range of genes associated with the Phosphorus cycle, as well as identifying phytase activity in environmental samples (Castillo Villamizar et al. 2019). Tan et al. employed a functional metagenomics technique to discover and utilize two new types of phytases that demonstrated the ability to break down phytate effectively at neutral pH levels (Tan, Mooij, and , Matthieu Barret , Pdraig M. Hegarty , Catriona Harrington , Alan D.W. Dobson 2014).

To successfully employ the metagenomics approach, it is crucial to have well-optimized protocols at every step, starting from the isolation of total DNA to library construction. These optimized protocols enable the screening of positive clones that possess the desired property (A. Kamble, Srinivasan, and Singh 2019a) (A. D. Kamble and Singh 2022). *In-silico* bioprospecting method helped to overcome the limitations of conventional and metagenomics approaches.

2.1.3 *In-silico* bioprospecting of novel phytases:

Database mining can also help to discover novel phytases e.g., Tan et.al. explored IMG/M database and discovered 11 putative HAP phytase genes from 18 publicly available metagenomes of fungus gardens. Among these rPhyXT52 phytase exhibited the highest activity at 52.5 °C (Tan, Wu, et al. 2016b). In another study, Tan et.al. used metagenomic datasets of an acidic peat-soil microbiome in northeastern Bavaria, Germany by using Pfam00328 (Pfam identifier) to explore novel phytases (Tan, Wu, et al. 2016a).

2.2 Protein engineering approach:

Researchers all over the world have been focusing on different aspects to increase thermostability via protein engineering approach (directed or rational engineering) and computational analysis via

software and tools. The directed evolution approach is time-consuming and requires an extensive screening process. Shivange et al. focus on the reduction of flexibility to enhance thermostability via the adopted KeySIDE approach to improve the thermostability of *Yersinia mollaretii* phytase (Ymphytase) which involved a combination of investigating iterative key residues of the wild type and identifying substitutions using directed evolution method. M6 mutant (T77K, Q154H, G187S, and K289Q) improved the residual activity at 58 °C by more than double as compared to the wild type. The mutations T77K, G187S, and K289E/Q reduced the flexibility of the loops near helices which overall increased thermostability (Shivange, Roccatano, and Schwaneberg 2016).

Researchers also adopt a rational engineering approach to improve thermostability e.g., Fei et.al. focused on protein flexibility as well as protein surface analysis and enhancement of salt bridges as a strategy for rational engineering of *Escherichia coli* AppA phytase. They analyzed protein flexibility with the help of a root mean square deviation (RMSF) graph obtained via Molecular Dynamic (MD) simulation. Surface thermal unstable residues were targeted based on their RMSF values which were above 2 Å° and above 4 Å° away from the AppA functional sites. They demonstrated that the salt bridges and the α/β - domain of *E. coli* phytase are of utmost importance (Fei et al. 2013) (Fei, Xu, Cao, et al. 2013). Han et.al. adopted a consensus sequence rational protein engineering approach to engineer *A.niger* phytase (Anp) by structural comparison with *A.fumigatus* phytase (Afp) structure with the help of molecular dynamics simulation. In AnP, the segments A35-P42, R163-Q168, and T248-K254 exhibited the highest main chain deviation. Hence these were targets for engineering. By hydrogen bond analysis of mutant candidate, it was found that S39 and S42 in segment-1, T165, N166, and R167 in segment-2, R248, D251, A252, and Q254 in segment-3 have the greatest contribution to the enhanced thermostability (Han et al. 2018).

To get an insight into the structure of phytases, some researchers utilize computational tools such as Pramanik et.al. uses computational tools to characterize *Enterobacter* phytases which were predicted to be thermostable based on its aliphatic indices, acidic as the isoelectric point was below 7, interaction with water based on lower values of GRAVY, the highest content of alpha-helical compared to other forms, function as histidine acid phosphatases and has conserved residues ‘DG–DP–LG’ (Pramanik et al. 2018). Similarly, V.Kumar et.al. characterized Histidine acid phytase sequences using computational tools and revealed the motifs that are found in obtained clusters

i.e., clusters 1 (PhyA) and 2 (PhyB) has Motif 1 “SPFCDLFTHEEWIQYDYDLQSLGKYYGYGAGNPLGPAQGIGF” and Cluster 3 (AppA) contains motif 9 “KKGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITV-HTQADTSSPDP”. This information can help to study the evolution of HAP phytase and the engineering of phytases for improving their features based on clustering (Vinod Kumar et al. 2012). Kumar et.al. studied the conformational dynamics of *A. niger* PhyA PDB ID: 3K4Q to detect thermo-sensitive regions as well as to study the effect of salt bridges on protein stability via MD simulation. As the temperature increases in a simulation study from 310K to 500K, there are changes at the structural level such as decreased main chain to main chain hydrogen bonds, increased solvent-accessible surface area, increased fluctuations of loops, increased rigidity of the catalytic site (K. Kumar et al. 2013).

Chapter 3: Lacunae, Rationale, Aim and Objectives

3.1 Lacunae/ Rationale:

Phytase hydrolyzes its substrate phytate. Phytate is a major source of phosphate in cereals and legumes and is indigestible in monogastric animals due to very limited phytase activity in the digestive tract. Hence exhibits potential to cause environmental pollution. Phytate also exhibits anti-nutritional properties by chelating micronutrients such as Ca^{2+} , Fe^{2+} , and Zn^{2+} , by forming complexes such as protein-phytate, starch-phytate, and lipid-phytate and reduces bioavailability which overall decreases the nutritional value of feed (Luo et al. 2009) (Kalsi et al. 2016) (Jaiwal et al. 2019) (Brinch-Pedersen et al. 2014). Hence, phytase is used as a feed additive in the feed industry to hydrolyze phytate to lower myoinositol derivatives and release phosphate. Thereby reducing the supplementation of inorganic phosphate in feed (which is obtained from non-renewable resources and accounts for 9% of the average feed cost), reducing the cost of feed which accounts for 60-70% of the total cost in livestock farming as well as aid in natural resource management. It also increases the bioavailability of micronutrients which help to maintain the high nutritional standard of feed and reduce environmental pollution by reducing eutrophication and deposition of low phytate content manure in soil (Jaiwal et al. 2019) (Brinch-Pedersen et al. 2014) (Cang et al. 2004) (Industrial Enzymes Market - Global Forecast by 2022 2017).

In addition to its wide application in the feed industry, phytases are also used as a bread-making additive (Rao et al. 2009). In the bread-making industry, it helped to reduce the fermentation time and improve bread texture (Rao et al. 2009). Phytase has been employed in aquaculture to counter the negative effects of phytate present in fish feed. By using phytase, the detrimental impact of phytate can be reduced. The application of phytase in aquaculture has yielded several benefits, including enhanced fish growth and improved bone mineralization due to increased bioavailability of minerals. Additionally, phytase has been shown to enhance protein digestibility and decrease the overall phosphorus burden in the aquatic environment (Rao et al. 2009). Microbial phytase genes introduced into transgenic plants enable the hydrolysis of phytate, leading to improved utilization of phosphate. These plants can be directly used as animal feed thereby reducing the cost of commercial phytase production (Rao et al. 2009). Similarly, the expression of microbial phytase genes in transgenic animals can eliminate the phosphorous pollution caused by animal agriculture (Rao et al. 2009). Microbial phytases are also utilized in soil amendment, leading to improvements

in soil quality and enhancing the growth and quality of plants (Rao et al. 2009). Besides that, phytases are used to synthesize lower myoinositol phosphates by enzymatic method. Enzymatic methods are environmentally friendly as compared to that of chemical synthesis methods. These inositol phosphate derivatives can be used as enzyme stabilizers, enzyme substrates, or inhibitors of enzymes (Rao et al. 2009) (Kaur, Vohra, and Satyanarayana 2021). It is also proposed to apply it as a painkiller and act as an inhibitor against retroviral infections including HIV. Hence, inositol phosphates that can be synthesized by using phytases, have potential therapeutic applications (Kaur, Vohra, and Satyanarayana 2021). Hence, in recent years, the use of phytase has extended beyond the feed-food industry and has found applications in various other fields. These include nutraceuticals, soil amendments, and the production of myoinositol phosphates. Commercial phytases like Natuphos® or Allzyme find extensive use in the feed industry, but their production is expensive, which ultimately restricts their application as feed additives (Vohra, and Satyanarayana 2021). Furthermore, apart from the aforementioned constraints, the production industry of inositol phosphates also requires new phytases that possess the capability to achieve stereospecificity, ensuring precise configuration. This is crucial to offer a sustainable alternative to the chemical synthesis process. Phytase applications are also limited in soil amendment due to several factors, including the soil environment (such as pH, temperature, and presence of cations), the presence of inhibitors, and the decreased availability of phytate because it binds tightly to soil components.

To address these challenges, we aim to obtain novel phytases through different approaches such as conventional, metagenomics, and *in-silico* bioprospecting and further improve its features via a protein engineering approach.

3.2 Aim and Objectives

3.2.1 Aim:

Aim: To obtain a novel phytase enzyme using a bioprospecting and protein engineering approach.

3.2.2 Objectives

- Isolation and screening of phytase-producing microbes from natural sources or *In-silico* Bioprospecting.
- *In-silico* characterization and experimental validation of shortlisted phytase.
- *In-silico* hotspot identification and experimental validation of shortlisted phytase.

Chapter 4: Isolation and screening of phytase-producing microbes from natural sources or *in-silico* bioprospecting.

This chapter delves into the exploration of new phytases through three distinct approaches: conventional, metagenomics, and *in-silico*. The conventional approach involved the isolation and screening of phytase-producing microorganisms from different soil samples. The metagenomics approach employed 16S rRNA gene sequencing of total soil DNA to investigate the overall microbial diversity and identify phytase-producing microorganisms. In the *in-silico* bioprospecting approach, researchers utilized various databases to screen microorganisms capable of producing phytase.

4.1 Conventional approach:

This approach aims to identify and isolate phytase-producing microbes from different soil samples (Nagar, Kamble, and Singh 2021). Soil samples were collected from different regions of Maharashtra and Gujarat. Soil samples were enriched in phytase screening broth (PSB) and plated on phytase screening agar plates (PSA). The potential isolates exhibited the zone of hydrolysis on PSA which indicated the release of phytase and the subsequent degradation of phytate present in the agar. The potential soil isolates named FD2, FD5T, and FD5O were screened from forest soil whereas VD5, CD5, and PSD were obtained from vegetation, cotton, and pig stag soil sample respectively. The above-maintained isolates and *E. coli* culture were tested for phytase using qualitative assays. The isolates were further identified by 16S rRNA gene amplification and sequencing. The analysis revealed that isolates VD5 and CD5 were closely related to *Enterobacter cloacae*, and FD5T was closely related to *Paenibacillus* sp. Whereas FD5O and FD2 were closely related to *Bacillus megaterium*, and PSD was closely related to *Klebsiella*. The phytase activity was estimated using the modified Heinonen method. In brief, the reaction mixture consisted of crude extract and substrate sodium phytate. The phytase activity of isolates was estimated at pH 2.5, and 5 at 37°C for 30min. The phytase activity was measured using KH_2PO_4 (0.1 mM to 2 mM) as a standard. The phytase activity assay revealed that PSD, CD5, FD5O, FD2 VD5, and FD5T had higher activity at pH 5 as compared to pH 2.5.

PSD was shortlisted based on preliminary screening of phytase-producing microbes from the soil via qualitative (phytase plate assay) and quantitative (phytase activity assay) methods. The full-length 16S rRNA gene of PSD was amplified and sequenced which revealed its closest

evolutionary relation with *Klebsiella variicola*. The phytase gene from PSD was amplified and sequenced. The presence of an N-terminal conserved region (RHGIRPP) in the sequence indicated that it belongs to the Histidine acid phytase group.

4.2 Metagenomics approach

First, the soil bacterial genomic DNA extraction method was optimized from different methods reported in the literature. The 16S rRNA gene metagenomic analysis of pig stag soil DNA sequencing was performed, and the study identified the distribution of bacterial families and genera, highlighting the *Xanthomonadaceae* family as a potential target for further investigation of phytase activity.

4.3 *In-silico* bioprospecting approach

In-silico bioprospecting approach involved three major steps: exploring the database, removing redundancy, and phylogenetic analysis. Databases such as NCBI, Scan-Prosite, UniProtKB, Conserved Domain database, HMMER, and JGI-metagenome were explored to identify potential phytase producers. The redundancy was removed using CD-Hit and MEGA7 software was used to construct a neighbor-joining phylogenetic tree using 1000 bootstrap, and Poison method. The reference of each tool mentioned above is listed below in the “Databases/tool/software Reference” section.

In-silico bioprospecting methods helped to obtain a total of 17 different potential phytase-producing genera from different databases. Out of these 17, two candidates i.e., *Pantoea vagans* (PV) and *Edwardsiella tarda* (ET) phytases were further studied computationally as well as characterized experimentally.

Chapter 5: *In-silico* characterization and experimental validation of shortlisted phytase

5.1 *In-silico* characterization of PV and ET

This consisted of primary, secondary, and tertiary structure analysis and validation and functional analysis of PV and ET. This also includes molecular docking and molecular dynamics simulation studies to understand the dynamics of PV and ET at different temperatures such as 300,350,400, and 450K for 30ns. The analysis helped to understand these two novel phytases i.e., ET and PV, and shortlist one for rational engineering to improve its stability.

Primary sequence analysis, using the ProtParam tool and InterProScan database revealed that PV and ET phytase belong to the Histidine phosphatase superfamily and possess Histidine acid phosphatase phosphohistidine signature. They were mildly acidic based on pI calculation. PV was more thermostable as compared to ET and references i.e., *Klebsiella pneumoniae* phytase (KP) and *Hafnia alvei* phytase (HA) based on aliphatic index. PV and ET both had better interaction with the surrounding aqueous environment based on the GRAVY index. The secondary structure of PV and ET, predicted by using Psipred, YASARA, PDBSum, Porter, SOPMA, and Proteus2 showed that PV and ET consisted of a higher percentage of helices compared to its other secondary structural elements. Similar observations were found in the case of their respective reference sequences i.e., KP and HA phytases. Tertiary structure modeling of PV and ET, using FunFold, Hhpred, Intfold, Orion, Phyre, Swiss model, tr-Rosetta, Topmodel, and ITASSER and validated by the EXPASY web portal. The PV and ET structures modeled using various tools exhibited a range of 87.28 to 99.1% and 74.7 to 94.4% of amino acid residues within the allowed regions in the Ramachandran plot respectively. According to the Ramachandran plot, the % of residues in the allowed regions in the case of model structures PV and ET via SWISS-MODEL were 96.41 and 92.5% respectively and that of KP and HA were 98.74% and 96.83% respectively. This indicated that the structure was reliable as the % of residues in the allowed regions of PV and ET were similar to their references i.e., KP and HA phytases respectively. It also indicated that the PV structure was more reliable as compared to the ET structure. The reference of each tool mentioned above is listed below in the “Databases/tool/software Reference” section.

Molecular Docking and Molecular Dynamics Simulation:

Molecular docking was conducted using Autodock 4 (Trott and Olson 2009) to investigate the binding of phytate to the catalytic pocket of ET and PV. The calculated binding affinity of PV and ET was -9.1 and -7.7 kcal/mol respectively. The Functional analysis and Molecular docking studies revealed that conserved C-terminal and N-terminal residues, along with other residues interacted with phytate, indicating stable and favorable interactions. The docked structures were subjected to molecular dynamics simulation studies at different temperatures (300 to 450 K) using GROMACS software to understand the stability of PV and ET. The analysis includes Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Solvent Accessible Surface Area (SASA), and Radius of Gyration (Rg). RMSD analysis revealed that the overall conformational changes in PV were less than that of ET. This indicated that PV was more stable than ET. The average backbone RMSF of PV was less compared to ET which suggested that PV was more stable than ET at higher temperatures, especially at 400 and 450K. The average backbone Rg of PV was less than that of ET which indicated that the structure of PV is stable at all temperatures as compared to ET. The average SASA of PV structure was less than that of ET which indicated less conformational changes in PV compared to ET. This suggested that PV was more stable than ET.

5.2 Experimental validation:

This study focused on cloning and over-expression of phytase genes from *P.vagans* and *E. tarda*. The full-length phytase gene from both *P.vagans* and *E. tarda* was then cloned in pET29-b and over-expressed in *E. coli* BL21. PV phytase (approximately 46kDa) was found to be overexpressed in the *E. coli* BL21(DE3) host system. However, ET phytase failed to over-express. The purified phytase of PV was then subjected to biochemical characterization i.e., pH and temperature optima. The optimum temperature of PV phytase was found to be 40°C. The optimum pH of PV phytase was found to be 4. The secondary structure analysis using Circular Dichroism (CD) suggested a folded structure of PV phytase.

Chapter 6: *In-silico* identification of Hotspot to design thermostable PV phytase and experimental validation.

6.1 Identification of hotspot in PV phytase:

This study aimed to investigate and identify regions or specific amino acid residues in PV phytase that contribute to its thermo-sensitivity. Identification of thermo-labile regions/residues was performed using various computational tools such as Disulfide by Design (Craig and Dombkowski 2013), Consensus sequence finder (Hua et al. 2020), Consurf (Celniker et al. 2013), consensus analysis with known phytases, literature (C. C. Chen and Tw 2018) (Wang et al. 2015), Hotspot wizard (Bendl et al. 2016), Fireprot (Musil et al. 2017), Cabsflex (Jamroz, Kolinski, and Kmiecik 2013) and FoldX (Schymkowitz et al. 2005). Subsequently, computational methods were employed to suggest amino acid substitutions that could enhance stability. Each method utilized specific criteria and filters to generate a list of shortlisted mutants based on their calculated ΔG values.

6.2 Experimental validation:

Experimental validation focused on understanding the stability of wildtype (WT) and mutant (M1 and M2) phytases via over-expression in the *E.coli* BL21(DE3) host system, followed by biochemical and biophysical characterization.

6.2.1 Over-expression and purification of wildtype (WT) and mutant (M1 and M2) phytases:

In this study, wild-type and mutant genes were synthesized. These genes were codon optimized and cloned into to pET29-b vector. The recombinant constructs were then transformed separately into to *E.coli* BL21(DE3) host system and induced with IPTG (1mM) at 30°C overnight. Both WT and M2 over-expressed successfully, whereas M1 didn't express and was found to be unstable. The WT and M2 phytase proteins were purified using His60Ni SuperFlow resin. The purified phytases were also confirmed by Western blot technique via anti-His antibody.

6.2.2 WT and M2 phytase characterization:

Biochemical characterization of purified WT and M2 was performed which included the effect of substrate concentration, pH, and temperature. The optimum pH of WT and M2 was found to be pH 4. The optimum temperature of WT was 45°C, whereas that of M2 was 50°C. Hence, a 5°C shift in temperature optima was observed. The substrate affinity of WT was slightly more than M2.

Biophysical characterization: It involved the use of Circular Dichroism, Isothermal Titration Calorimetry (ITC), and differential scanning fluorimetry (DSF) methods to gain insights into folding (secondary structure analysis), affinity (K_d) and stability (T_m) of PV phytase. ITC data revealed that the binding affinity of WT was similar to M2. CD data revealed that both WT and M2 were properly folded. DSF data revealed that T_m of WT and M2 was found to be 58 and 60°C respectively. This change in T_m indicated that M2 was more stable than WT.

Chapter 7: Discussion

Exploring the immense microbial diversity in natural environments like soil offers the potential to discover novel enzymes, including phytase (Suleimanova et al. 2015) (Sajidan et al. 2004). Each gram of soil contains an extensive variety of bacterial and archaeal species, ranging from 100,000 to 1,000,000 (Satyanarayana, Krishna, and Kumar 2017). In this study, various soil samples from different regions of Maharashtra and Gujarat were explored to obtain novel phytase producers (Nagar, Kamble, and Singh 2021). Both qualitative plate assay and quantitative assay to isolate and screen six potential phytase producers i.e., FD5T, FD5O, FD2 from forest soil, VD5 from vegetation soil, CD5 from cotton field, and PSD from pigstagg soil. These shortlisted soil isolates were identified via 16S rRNA gene sequencing. This analysis identified *Enterobacter cloacae* (VD5 and CD5), *Paenibacillus* sp. (FD5T), *Bacillus megaterium* (FD5O and FD2), *Klebsiella variicola* (PSD). Previous reports have indicated that *Enterobacter* (Kalsi et al. 2016) and *Klebsiella* sp. *ASRI* (Sajidan et al. 2004) isolated from soil exhibit maximum activity at pH 7 and 5 respectively which supports our results. The sequencing analysis confirmed that PSD belongs to the HAP group, as indicated by the presence of the N-terminal conserved region RHGXRX in its sequence. Moreover, the isolated PSD demonstrated promising potential for further exploration in industrial applications. Therefore, the conventional approach adopted in this presents a mini-scale alternative for identifying bacterial phytase producers (Nagar, Kamble, and Singh 2021).

The conventional approach can only be utilized to investigate the cultivable microbial population. Only 1% of the microbial communities can be cultured in laboratory conditions while 99% are still unexplored because they are non-culturable (Robe et al. 2003) (Faria Fatima, Chaudhary Ira, Ali Jasarat, Rastogi Smita 2011) (Kamble Asmita and Singh Harinder 2020). To overcome these limitations, the 16S ribosomal RNA (16S rRNA) gene metagenomics approach can be used to explore the non-cultural microbial community present in soil samples (Muwawa et al. 2021). In this study, a metagenomic approach was used to study microbial diversity in the pig stagg soil sample. Due to its high phytate content, the soil sample obtained from a pig enclosure had the potential to support the growth of phytase-producing microorganisms. (Menezes-Blackburn et al. 2013). The 16S rRNA gene Metagenomic approach was taken that involved the isolation of total soil genomic DNA from Pig stagg soil (Kamble Asmita and Singh Harinder 2020), followed by 16S rRNA gene metagenomic sequencing. The analysis was performed using the QIIME2 platform (A.

Kamble, Sawant, and Singh 2020). The result indicated *Xanthomonadaceae* family holds potential for further investigation (DiFonzo and Bordia 2003). *Xanthomonas* phytase can be further explored by designing full-length degenerative phytase gene primers to amplify the phytase gene, cloning and over-expressing it in an appropriate host system, and ultimately characterizing the purified phytase.

In-silico bioprospecting, a computational approach, is a straightforward and efficient method to discover new candidates from existing databases (Voß, Heck, and Schallmeyer 2020). It allows us to overcome the limitations of traditional approaches. The advent of sequencing technology has dramatically increased the number of sequences available in databases (Stewart 2012), thereby expanding the pool of enzyme diversity (Stewart 2012). *In-silico* bioprospecting takes advantage of this diversity by employing the following steps: exploration of databases to find novel candidates, followed by screening, analysis, and shortlisting of these candidates using bioinformatics tools (A. D. Kamble and Singh 2022). The exploration of the database can be based on homology, conserved domain, keyword search, and consensus-guided (A. Kamble, Srinivasan, and Singh 2019b). In the present study, different databases were explored such as NCBI, Scan-Prosit, UniProtKB, Conserved Domain database, HMMER, and JGI-metagenome based on homology, keyword search, conserved motif/domain, and consensus guided. The focus was to explore novel phytases that were 50-80% identical to experimentally well-characterized phytases. The steps involved were exploring databases, followed by removing redundancy and studying its evolutionary relationship by constructing a phylogenetic neighbor-joining tree. Finally, we obtained 17 different genera as potential phytase producers.

To summarize, the utilization of three different approaches - conventional, metagenomics, and *in-silico* bioprospecting - enabled the identification and screening of new phytases. Each method yielded distinct phytase producers, and the combination of all three approaches expanded the potential for discovering additional novel phytases. In this study, *Pantoea vagans* (PV) and *Edwardsiella tarda* (ET) phytases were shortlisted for *in-silico* characterization and experimental validation.

In-silico characterization of novel enzymes helped to gain information about their primary sequence, secondary structure, and tertiary structure along with the functional, evolutionary relationship with the experimentally proven homologs, and dynamics studies via molecular

dynamic simulations. In the present study, *in-silico* characterization of PV and ET phytases was conducted to gain insights into the overall structural, functional, and dynamic behavior. The primary sequence analysis suggested that PV and ET both belonged to the histidine acid phosphatase superfamily. The closest homolog of PV and ET was KP and HA respectively. The known experimentally validated phytases such as *Klebsiella* (Böhm et al. 2010), *Hafnia* (Ariza et al. 2013), *E.coli* (Lim et al. 2000) and many others also belong to the same superfamily. The overall primary sequence analysis also suggested that PV was thermostable and exhibited better interaction with the surrounding solvent compared to ET. The secondary and tertiary structure analysis revealed that the structure and topology of ET and PV matched with known structures. Examination of the catalytic sites in PV and ET phytases revealed the presence of conserved regions, namely RHGXRXR at the N-terminal and HD at the C-terminal, which resemble the corresponding regions in *Klebsiella* (Böhm et al. 2010) and *Hafnia* (Ariza et al. 2013). These findings strongly suggest that PV and ET can degrade phytate, similar to the experimentally characterized phytases of *Klebsiella* (Böhm et al. 2010) and *Hafnia* (Ariza et al. 2013). The validation results revealed that the PV structure was more stable and reliable as compared to that of ET.

Molecular dynamic simulation of protein helps to understand the conformational dynamics of protein. In the present study, molecular dynamic simulation was conducted to understand the conformational dynamics of PV and ET phytases at different temperatures (300 to 400K for 30ns). The MDS analysis, especially the average RMSF, Rg, and SASA, revealed that PV was stable as compared to ET. Overall, the computational analysis provides valuable insights into the structural characteristics and dynamics of PV and ET phytases, highlighting PV's stability compared to ET.

Experimental validation along with *in-silico* characterization enables researchers to acquire a more comprehensive understanding of enzyme characteristics. In this study, experimental validation of both PV and ET phytases was conducted to shortlist one candidate (either PV or ET) based on the results. The full-length phytase genes of PV and ET were approximately 1.3 and 1.4kbp size, comparable to known phytase genes (Ariza et al. 2013) (Abeldenov et al. 2017) (Sajidan et al. 2004). The full-length PV and ET phytase genes were cloned in the pET29-b vector and over-expressed in *E. coli* BL21(DE3). PV phytase was found to be produced in large quantities, whereas Et was not stable and didn't express. The size of PV phytase was found to be approximately 46kDa

which was comparable to its closet homolog *Klebsiella* phytase (Sajidan et al. 2004) (Böhm et al. 2010). The size of the PV phytase was also confirmed via HR-LC-MS. The purified phytase was biochemically characterized. The optimum pH of PV phytase was found to be 4 and optimum temperature 40°C which was similar to *Klebsiella* phytase (optimum pH 5 and optimum temperature 45°C). In summary, *in-silico* characterization and experimental validation resulted in the selection of PV phytase for rational engineering.

The protein engineering approach (rational/semi-rational/directed evolution) can be employed to improve characteristics of enzymes such as thermostability, resistance to proteolytic degradation, or attaining broad-range pH stability. *In-silico* identification of hotspots by using different tools/servers/ software was adopted to improve a number of enzymes such as lipase (G. Li et al. 2018), phospholipase D (L. Li et al. 2022), and phytases (Tan, Miao, et al. 2016) (Xi Wang & Jun Du & Zhi-yun Zhang & Yue-jun Fu & Wen-ming Wang & Ai-Hua Liang 2018). In the present study, Hotspots or thermo-labile regions/ residues were identified by using computational tools/servers/ software such as Disulfide by Design2, Consensus Finder, ConSurf, Hotspot Wizard, Fireprot, CABSFLEX, and based on literature. We also employed a combination of these methods to generate a comprehensive list of potential hotspots. To further refine the selection, the hotspots identified were evaluated based on the ΔG (change in free energy) values calculated using FoldX software. The ΔG values of the mutant variants were expected to be lower than those of the wild-type (WT) enzyme, indicating improved stability. A thorough analysis yielded a list of potential mutants. Two of these mutants i.e., M1 (E85W/T97C/L178Y/N212C/R350N) and M2 (A163C, Y271C) were further experimentally verified. WT and M2 were over-expressed in the *E.coli* BL21(D3) host system, whereas M1 was unstable and did not show expression. The biochemical characterization showed the optimum pH of M2 and WT as 4, which was like its closest homolog *Klebsiella* (Sajidan et al. 2004) (Böhm et al. 2010). The optimum temperatures of WT and M2 were 45°C and 50°C respectively. Hence, there was a 5°C temperature shift which signifies that M2 was stable compared to that of WT and *Klebsiella* phytase (Sajidan et al. 2004) (Böhm et al. 2010). The effect of substrate concentration analysis revealed that K_m of WT and M2 were 0.24 and 0.33mM respectively which were comparable to *Klebsiella* phytase (Greiner et al. 1997). The Biophysical characterization was performed by using CD, ITC, and DSF. The results revealed that WT and M2 were properly folded and binding affinity was not affected due to the introduction of mutation. The

T_m of WT was found to be 58°C while the T_m of M2 was 60°C. Hence, M2 was more stable than WT.

Chapter 8: Summary and Conclusion:

In this present study, three different approaches were used to identify potential phytase producers: conventional, metagenomics, and *In-silico* bioprospecting. A total number of 17 different genera were identified, including *Pantoea vagans*(PV) and *Edwardsiella tarda* (ET). The phytases from PV and ET were further studied computationally and experimentally. Hence, the combination of these three approaches resulted in identifying a wide range of potential phytase producers.

Both *in-silico* characterization and experimental validation indicated that PV phytase exhibited greater stability than ET, making it the preferred candidate for further optimization using rational engineering techniques. Various computational techniques were used to identify potential mutations in PV phytase that could improve its stability. This study is notable for its unique methodology, which involved employing a wide array of computational tools, servers, and software to produce a comprehensive analysis and generate an extensive list of potential mutant enzymes. The two mutations were further checked experimentally. Based on the experimental validation results, it was determined that M2 was more stable than WT in terms of T_m values.

Chapter 9: Significance of the study:

The significance of this study lies in its exploration of different approaches to identify potential phytase producers. By employing conventional, metagenomic, and *in-silico* bioprospecting methods, we were able to identify a wide range of potential phytase producers. This approach allowed for the identification of microbes in soil that produce effective phytase, providing an alternative, smaller-scale method. In addition to the above points, this study stands out due to its distinctive approach, utilizing a diverse range of computational tools, servers, and software to conduct a thorough analysis and create a vast list of potential mutant enzymes.

Overall, this study contributes to the understanding of potential phytase producers, explores different approaches for their identification, and provides insights into the stability and potential improvements of these enzymes through computational and experimental approaches. The study has obtained novel active and stable phytase enzyme candidates that have wide applications in the soil amendment process, feed and food industry, aquaculture, myo-inositol phosphate production, and bioethanol production.

Publications and Presentations:

Research/Review paper:

1. **Kamble, A.**, Srinivasan, S. and Singh, H. (2019) ‘In-Silico Bioprospecting: Finding Better Enzymes’, *Molecular Biotechnology*, 61(1), pp. 53–59. (Impact factor: 2.860)
2. **Kamble, A.**, Sawant, S. and Singh, H. (2020) ‘16S Ribosomal RNA Gene-Based Metagenomics : A Review’, *BMRJ*, pp. 5–11.
3. **Kamble A** and Singh H (2020) ‘Different Methods of Soil DNA Extraction’, *Bio-protocol*, e3521, pp. 1–23. (*Web of Science*: Emerging Sources Citation Index (ESCI) indexes).
4. Nagar, A., **Kamble, A.** and Singh, H. (2021) ‘Preliminary screening, isolation and identification of microbial phytase producers from soil’, *Environmental and Experimental Biology*, 19, pp. 11–22. (Additional *Web of Science* Indexes).

Book Chapter:

- **Kamble, A.D.** and Singh, H. Ch26 - Finding novel enzymes by in silico bioprospecting approach. Ed(s): M Kuddus, CN Aguilar. In: *Value-Addition in Food Products and Processing Through Enzyme Technology*, Academic Press, 2022, Pages 347-364, ISBN 9780323899291,

Conferences and poster presentations

- International conference: *Advances in Materials Science & Applied Biology (AMSAB)* held on 8th – 10th January, 2019 at NMIMS University. Title of the Poster presented in AMSAB : Computational analysis of putative phytase enzymes.

Scholarship:

- ‘Late Dr. Suresh Mahajan - IWSA Scholarship’ for the year 2022-2023.

Grant

- NMIMS seed grant: 2018 and 2020

Bibliography:

- Abdollahi, M. R., V. Ravindran, and B. Svihus. 2013. "Pelleting of Broiler Diets: An Overview with Emphasis on Pellet Quality and Nutritional Value." *Animal Feed Science and Technology* 179(1–4): 1–23. <http://dx.doi.org/10.1016/j.anifeedsci.2012.10.011>.
- Abeldenov, S et al. 2017. "Expression, Purification and Biochemical Characterization of Recombinant Phosphohydrolase Appa in Escherichia Coli." *Eurasian journal of applied biotechnology*: 1–10.
- Amritha, Girish K, and G Venkateswaran. 2017. "Use of Lactobacilli in Cereal-Legume Fermentation and as Potential Probiotics towards Phytate Hydrolysis." *Probiotics & Antimicro. Prot.*
- Ariza, Antonio et al. 2013. "Degradation of Phytate by the 6-Phytase from Hafnia Alvei: A Combined Structural and Solution Study." *PLoS ONE* 8(5): e65062.
- Becker, Geoffrey S. 2008. *Livestock Feed Costs: Concerns and Options*. http://www.econ.iastate.edu/faculty/lawrence/Lawrence_Website/estreturns.htm. (April 16, 2021).
- Bendl, Jaroslav et al. 2016. "HotSpot Wizard 2.0: Automated Design of Site-Specific Mutations and Smart Libraries in Protein Engineering." *Nucleic Acids Research* 44(1): W479–87.
- Bhavsar, K., and J. M. Khire. 2014. "Current Research and Future Perspectives of Phytase Bioprocessing." *RSC Advances* 4(51): 26677. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84903589969&partnerID=tZOtx3y1>.
- Böhm, Kerstin et al. 2010. "Crystal Structure of Klebsiella Sp. ASR1 Phytase Suggests Substrate Binding to a Preformed Active Site That Meets the Requirements of a Plant Rhizosphere Enzyme." *FEBS Journal* 277(5): 1284–96.
- Brinch-Pedersen, Henrik, Claus Krogh Madsen, Inger Bæksted Holme, and Giuseppe Dionisio. 2014. "Increased Understanding of the Cereal Phytase Complement for Better Mineral Bio-Availability and Resource Management." *Journal of Cereal Science* 59(3): 373–81.
- Cang, Long, Yu-jun Wang, Dong-mei Zhou, and Yuan-hua Dong. 2004. "Heavy Metals Pollution in Poultry and Livestock Feeds and Manures under Intensive Farming in Jiangsu Province, China." *Journal of environmental sciences (China)* 16(3): 371–74.
- Castillo Villamizar, Genis Andrés et al. 2019. "Functional Metagenomics Reveals an Overlooked Diversity and Novel Features of Soil-Derived Bacterial Phosphatases and Phytases." *mBio* 10(1): 1–15.
- Celniker, Gershon et al. 2013. "ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function." *Israel Journal of Chemistry* 53(3–4): 199–206.
- Chen, Chun Chi, and Taipei Tw. 2018. "PHYTASE HAVING IMPROVED THERMOSTABILITY." 2.

- Chen, Weiwei et al. 2015. "Enhanced Activity of an Alkaline Phytase from *Bacillus Subtilis* 168 in Acidic and Neutral Environments by Directed Evolution." *Elsevier B.V.*
<http://dx.doi.org/10.1016/j.bej.2015.02.021>.
- Craig, Douglas B., and Alan A. Dombkowski. 2013. "Disulfide by Design 2.0: A Web-Based Tool for Disulfide Engineering in Proteins." *BMC Bioinformatics* 14(1): 0–6.
- DiFonzo, Nicholas, and Prashant Bordia. 2003. "PhyA, a Secreted Protein of *Xanthomonas Oryzae* Pv. *Oryzae*, Is Required for Optimum Virulence and Growth on Phytic Acid as a Sole Phosphate Source." *Journal of Allergy and Clinical Immunology* 16(11): 973–82.
<http://dx.doi.org/10.1016/j.jaci.2012.05.050>.
- Faria Fatima, Chaudhary Ira, Ali Jasarat, Rastogi Smita, Pathak Neelam. 2011. "Microbial DNA Extraction from Soil by Different Methods and Its PCR Amplification." *Biochem.cell.Arch* 11.
- Farias, Nathálya, Isabela Almeida, and Carlos Meneses. 2018. "New Bacterial Phytase through Metagenomic Prospection." *Molecules* 23(2): 1–14.
- Fei, Baojin, Hui Xu, Yu Cao, et al. 2013. "A Multi-Factors Rational Design Strategy for Enhancing the Thermostability of *Escherichia Coli* AppA Phytase." *Journal of Industrial Microbiology and Biotechnology* 40(5): 457–64.
- Fei, Baojin, Hui Xu, Feiwei Zhang, et al. 2013. "Relationship between *Escherichia Coli* AppA Phytase's Thermostability and Salt Bridges." *Journal of Bioscience and Bioengineering* 115(6): 623–27. <http://dx.doi.org/10.1016/j.jbiosc.2012.12.010>.
- Finn, Robert D., Jody Clements, and Sean R. Eddy. 2011. "HMMER Web Server: Interactive Sequence Similarity Searching." *Nucleic Acids Research* 39: 29–37.
- Goutami Banerjee, Khin Oo, Xlyun Zhang, Jie Yang, Yingxin Zhang. 2017. "PHYTASES AND USES THEREOF."
- Greiner, Ralf, Edith Haller, Ursula Konietzny, and Klaus-dieter Jany. 1997. "Purification and Characterization of a Phytase from *Klebsiella Terrigena*." *ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS* 341(2): 201–6.
- Han, Nanyu et al. 2018. "Enhancing Thermal Tolerance of *Aspergillus Niger* PhyA Phytase Directed by Structural Comparison and Computational Simulation." *BMC Biotechnology* 18(1): 1–8.
- Hua, Yujiao et al. 2020. "Improving the Thermostability of Glutamate Decarboxylase from *Lactobacillus Brevis* by Consensus Mutagenesis." *Applied Biochemistry and Biotechnology* 191(4): 1456–69.
- "Industrial Enzymes Market - Global Forecast by 2022." 2017.
<https://www.marketsandmarkets.com/PressReleases/industrial-enzymes.asp>.
- Jaiwal, Pawan Kumar, Anil K Chhillar, Darshna Chaudhary, and Ranjana Jaiwal. 2019. *Nutritional Quality Improvement in Plants*.

- Jamroz, Michal, Andrzej Kolinski, and Sebastian Kmiecik. 2013. "CABS-Flex: Server for Fast Simulation of Protein Structure Fluctuations." *Nucleic acids research* 41(Web Server issue): 427–31.
- Kalsi, Harpreet Kaur, Rajveer Singh, Harcharan Singh Dhaliwal, and Vinod Kumar. 2016. "Phytases from Enterobacter and Serratia Species with Desirable Characteristics for Food and Feed Applications." *3 Biotech* 6(1): 1–13.
- Kamble, Asmita Deepak, and Harinder Singh. 2022. "Finding Novel Enzymes by in Silico Bioprospecting Approach." In *Value-Addition in Food Products and Processing Through Enzyme Technology*, Academic Press, 347–64.
- Kamble, Asmita, Shriya Sawant, and Harinder Singh. 2020. "16S Ribosomal RNA Gene - Based Metagenomics : A Review." *Biomedical Research Journal* 7(5): 5–11.
- Kamble Asmita, and Singh Harinder. 2020. "Different Methods of Soil DNA Extraction." *Bio-101* e3521: 1–23.
- Kamble, Asmita, Sumana Srinivasan, and Harinder Singh. 2019a. "In-Silico Bioprospecting: Finding Better Enzymes." *Molecular Biotechnology* 61(1): 53–59. <http://dx.doi.org/10.1007/s12033-018-0132-1>.
- . 2019b. "In-Silico Bioprospecting: Finding Better Enzymes." *Molecular Biotechnology* 61(1): 53–59.
- Kaur, Parvinder, Ashima Vohra, and Tulasi Satyanarayana. 2021. *Encyclopedia of Mycology Multifarious Applications of Fungal Phytases*. Elsevier Ltd.
- Kumar, Kapil, Mudit Dixit, Jm Khire, and Sourav Pal. 2013. "Atomistic Details of Effect of Disulfide Bond Reduction on Active Site of Phytase B from Aspergillus Niger: A MD Study." *Bioinformation* 9(19): 963–67. <http://www.ncbi.nlm.nih.gov/pubmed/24391358><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3867648>.
- Kumar, Kapil, Krunal Patel, D. C. Agrawal, and J. M. Khire. 2015. "Insights into the Unfolding Pathway and Identification of Thermally Sensitive Regions of Phytase from Aspergillus Niger by Molecular Dynamics Simulations." *Journal of Molecular Modeling* 21(6).
- Kumar, Vikas, and Amit K. Sinha. 2018. *Enzymes in Human and Animal Nutrition: Principles and Perspectives General Aspects of Phytases*. Elsevier Inc.
- Kumar, Vinod et al. 2013. "Isolation of Phytase-Producing Bacteria from Himalayan Soils and Their Effect on Growth and Phosphorus Uptake of Indian Mustard (Brassica Juncea)." *World Journal of Microbiology and Biotechnology* 29(8): 1361–69.
- Kumar, Vinod, Gopal Singh, A. K. Verma, and Sanjeev Agrawal. 2012. "In Silico Characterization of Histidine Acid Phytase Sequences." *Enzyme Research*.
- Li, Guanlin et al. 2018. "Identification of a Hot-Spot to Enhance: Candida Rugosa Lipase Thermostability by Rational Design Methods." *RSC Advances* 8(4): 1948–57.

- Li, Lilang et al. 2022. "Improving Both the Thermostability and Catalytic Efficiency of Phospholipase D from *Moritella* Sp. JT01 through Disulfide Bond Engineering Strategy." *International Journal of Molecular Sciences* 23(19): 1–13.
- Lim, D, S Golovan, C W Forsberg, and Z Jia. 2000. "Crystal Structures of *Escherichia Coli* Phytase and Its Complex with Phytate." *Nature structural biology* 7(2): 108–13.
- Luo, Huiying et al. 2009. "Novel Low-Temperature-Active Phytase from *Erwinia Carotovora* Var . Carotovota ACCC 10276." *J. Microbiol. Biotechnol.* 19(10): 1085–91.
- Makolomakwa, Melvin, Adarsh Kumar Puri, Kugen Permaul, and Suren Singh. 2017. "Thermo-Acid-Stable Phytase-Mediated Enhancement of Bioethanol Production Using *Colocasia Esculenta*." *Bioresource Technology* 235: 396–404. <http://dx.doi.org/10.1016/j.biortech.2017.03.157>.
- Mandviwala, T. N., and J. M. Khire. 2000. "Production of High Activity Thermostable Phytase from Thermotolerant *Aspergillus Niger* in Solid State Fermentation." *Journal of Industrial Microbiology and Biotechnology* 24(4): 237–43.
- Menezes-Blackburn, Daniel et al. 2013. "Phytases and Phytase-Labile Organic Phosphorus in Manures and Soils." *Critical Reviews in Environmental Science and Technology* 43(9): 916–54.
- Mrudula Vasudevan, Ushasree, Amit K. Jaiswal, Shyam Krishna, and Ashok Pandey. 2019. "Thermostable Phytase in Feed and Fuel Industries." *Bioresource Technology* 278(January): 400–407. <https://doi.org/10.1016/j.biortech.2019.01.065>.
- Musil, Milos et al. 2017. "FireProt: Web Server for Automated Design of Thermostable Proteins." *Nucleic Acids Research* 45(W1): W393–99.
- Muwawa, Edith M. et al. 2021. "16S rRNA Gene Amplicon-Based Metagenomic Analysis of Bacterial Communities in the Rhizospheres of Selected Mangrove Species from Mida Creek and Gazi Bay, Kenya." *PLoS ONE* 16(3): 1–22. <http://dx.doi.org/10.1371/journal.pone.0248485>.
- Nagar, Akshita, Asmita Kamble, and Harinder Singh. 2021. "Preliminary Screening, Isolation and Identification of Microbial Phytase Producers from Soil." *Environmental and Experimental Biology* 19(1): 11–22.
- Pramanik, Krishnendu et al. 2018. "Computational-Based Structural, Functional and Phylogenetic Analysis of *Enterobacter* Phytases." *3 Biotech* 8(6): 1–12. <http://dx.doi.org/10.1007/s13205-018-1287-y>.
- Puppala, Kumar Raja et al. 2019. "Characterization of Novel Acidic and Thermostable Phytase Secreting *Streptomyces* Sp. (NCIM 5533) for Plant Growth Promoting Characteristics." *Biocatalysis and Agricultural Biotechnology* 18(Ncim 5533): 101020. <https://doi.org/10.1016/j.bcab.2019.101020>.
- Rao, D. E.C.S., K. V. Rao, T. P. Reddy, and V. D. Reddy. 2009. "Molecular Characterization, Physicochemical Properties, Known and Potential Applications of Phytases: An Overview." *Critical Reviews in Biotechnology* 29(2): 182–98.

- Rebello, Sharrel, Leny Jose, Raveendran Sindhu, and Embalil Mathachan Aneesh. 2017. "Molecular Advancements in the Development of Thermostable Phytases." *Applied Microbiology and Biotechnology* 101(7): 2677–89.
- Robe, Patrick et al. 2003. "Extraction of DNA from Soil." *European Journal of Soil Biology* 39(4): 183–90.
- Sadeghian, Issa, Zahra Rezaie, Seyyed Soheil Rahmatabadi, and Shiva Hemmati. 2020. "Biochemical Insights into a Novel Thermo/Organo Tolerant Bilirubin Oxidase from *Thermosediminibacter Oceani* and Its Application in Dye Decolorization." *Process Biochemistry* 88: 38–50.
- Sajidan et al. 2015. "Phytase-Producing Bacteria from Extreme Regions in Indonesia." *Brazilian Archives of Biology and Technology* 58(5): 711–17.
- Sajidan, A. et al. 2004. "Molecular and Physiological Characterisation of a 3-Phytase from Soil Bacterium *Klebsiella* Sp. ASR1." *Applied Microbiology and Biotechnology* 65(1): 110–18.
- Satyanarayana, Sadam D.V. Optimization of high-yielding protocol for DNA extraction from the forest rhizosphere microbes, M. S.R. Krishna, and Pindi Pavan Kumar. 2017. "Optimization of High-Yielding Protocol for DNA Extraction from the Forest Rhizosphere Microbes." *3 Biotech* 7(2): 1–9.
- Schymkowitz, Joost et al. 2005. "The FoldX Web Server: An Online Force Field." *Nucleic Acids Research* 33(SUPPL. 2): 382–88.
- Shivange, Amol V., Danilo Roccatano, and Ulrich Schwaneberg. 2016. "Iterative Key-Residues Interrogation of a Phytase with Thermostability Increasing Substitutions Identified in Directed Evolution." *Applied Microbiology and Biotechnology* 100(1): 227–42.
- Soh, Lemuel M.J. et al. 2017. "Engineering a Thermostable Keto Acid Decarboxylase Using Directed Evolution and Computationally Directed Protein Design." *ACS Synthetic Biology* 6(4): 610–18.
- Stewart, Eric J. 2012. "Growing Unculturable Bacteria." *Journal of Bacteriology* 194(16): 4151–60.
- Suleimanova, Aliya D. et al. 2015. "Novel Glucose-1-Phosphatase with High Phytase Activity and Unusual Metal Ion Activation from Soil Bacterium *Pantoea* Sp. Strain 3.5.1." *Applied and Environmental Microbiology* 81(19): 6790–99.
- Tan, Hao, Xiang Wu, et al. 2016a. "A Novel Phytase Derived from an Acidic Peat-Soil Microbiome Showing High Stability under Acidic Plus Pepsin Conditions." *Journal of Molecular Microbiology and Biotechnology* 26(4): 291–301.
- Tan, Hao, Renyun Miao, et al. 2016. "Enhancing the Thermal Resistance of a Novel Acidobacteria-Derived Phytase by Engineering of Disulfide Bridges." *Journal of Microbiology and Biotechnology* 26(10): 1717–22.
- Tan, Hao, Xiang Wu, et al. 2016b. "Identification and Characterization of a Mesophilic Phytase Highly Resilient to High-Temperatures from a Fungus-Garden Associated Metagenome." *Applied Microbiology and Biotechnology* 100(5): 2225–41.

- Tan, Hao, Marlies J. Mooij, and and Fergal O’Gara , Matthieu Barret , Pdraig M. Hegarty , Catriona Harrington , Alan D.W. Dobson. 2014. “Identification of Novel Phytase Genes from an Agricultural Soil-Derived Metagenome.” *J.Microbiol.Biotechnol* 24(1): 113–18.
- Trott, Oleg, and Arthur J. Olson. 2009. “AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading.” *Journal of Computational Chemistry* 31(2): NA-NA.
- Vashishth, Amit, Sewa Ram, and Vikas Beniwal. 2017. “Cereal Phytases and Their Importance in Improvement of Micronutrients Bioavailability.” *3 Biotech* 7(1): 1–7.
- Voß, Hauke, Amata Heck, and Marcus Schallmey. 2020. “Database Mining for Novel Bacterial Beta-Etherases, Glutathione- Dependent Lignin-Degrading Enzymes.” *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* 86(2): 1–15.
- Wang, Xi et al. 2015. “Enzymology and Thermal Stability of Phytase AppA Mutants.” *RSC Advances* 5(54): 43863–72.
- Xi Wang & Jun Du & Zhi-yun Zhang & Yue-jun Fu & Wen-ming Wang & Ai-Hua Liang. 2018. “A Rational Design to Enhance the Resistance of Escherichia Coli Phytase AppA to Trypsin.” *Applied Microbiology and Biotechnology* 102(22): 9647–56.
- Zhang, Huitu et al. 2019. “Rational Design of a Yarrowia Lipolytica Derived Lipase for Improved Thermostability.” *International Journal of Biological Macromolecules* 137: 1190–98.
<https://doi.org/10.1016/j.ijbiomac.2019.07.070>.

Databases/tool/software references:

- NCBI: Sayers, Eric W. et al. 2021. “Database Resources of the National Center for Biotechnology Information.” *Nucleic Acids Research* 49(D1): D10–17.
- ScanProsite: Sigrist, Christian J.A. et al. 2013. “New and Continuing Developments at PROSITE.” *Nucleic Acids Research* 41(D1): 344–47.
- UniprotKB: Magrane, Michele, and Uni Prot Consortium. 2011. “UniProt Knowledgebase: A Hub of Integrated Protein Data.” *Database* 2011: 1–13.
- CDD: Lu, Shennan et al. 2020. “CDD/SPARCLE: The Conserved Domain Database in 2020.” *Nucleic Acids Research* 48(D1): D265–68.
- HHMER: Finn, Robert D., Jody Clements, and Sean R. Eddy. 2011. “HMMER Web Server: Interactive Sequence Similarity Searching.” *Nucleic Acids Research* 39: 29–37.

- JGI-Metagenome: I-Min A Chen et al. 2023. “The IMG/M data management and analysis system v.7: content updates and new features” *Nucleic Acids Research* 51(D1): D723–D732.
- ProtParam: Gasteiger, Elisabeth et al. 2005. “Protein Identification and Analysis Tools on the ExPASy Server.” *The Proteomics Protocols Handbook*: 571–607.
- InterProScan: Mulder, Nicola, and Rolf Apweiler. 2007. “InterPro and InterProScan.” *Comparative Genomics* 2: 59–70.
- Psipred: McGuffin, L. J., K. Bryson, and D. T. Jones. 2000. “The PSIPRED Protein Structure Prediction Server.” *Bioinformatics* 16(4): 404–5.
- YASARA: Land, Henrik, and Maria Svedendahl Humble. 2018. “YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations Henrik.” In *Protein Engineering: Methods and Protocols*, , 43–67.
- PDBSum: Laskowski, Roman A. et al. 2018. “PDBsum: Structural Summaries of PDB Entries.” *Protein Science* 27(1): 129–34.
- Porter: Mirabello, Claudio, and Gianluca Pollastri. 2013. “Porter, PaleAle 4.0: High-Accuracy Prediction of Protein Secondary Structure and Relative Solvent Accessibility.” *Bioinformatics* 29(16): 2056–58.
- SOPMA: Geourjon, C., and G. Deléage. 1995. “Sopma: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments.” *Bioinformatics* 11(6): 681–84. <https://pubmed.ncbi.nlm.nih.gov/8808585/> (January 31, 2021).
- PROTEUS: Montgomerie, Scott et al. 2008. “PROTEUS2: A Web Server for Comprehensive Protein Structure Prediction and Structure-Based Annotation.” *Nucleic acids research* 36(Web Server issue): 202–9.
- FunFOLD: Roche, Daniel B., Stuart J. Tetchner, and Liam J. McGuffin. 2011. “FunFOLD: An Improved Automated Method for the Prediction of Ligand Binding Residues Using 3D Models of Proteins.” *BMC Bioinformatics* 12(1): 160. <http://www.biomedcentral.com/1471-2105/12/160>.

- HHPred: Zimmermann, Lukas et al. 2018. “A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at Its Core.” *Journal of Molecular Biology* 430(15): 2237–43.
- IntFold: McGuffin, Liam J. et al. 2019. “IntFOLD: An Integrated Web Resource for High Performance Protein Structure and Function Prediction.” *Nucleic Acids Research* 47(W1): W408–13.
- ORION: Ghouzam, Yassine et al. 2016. “ORION: A Web Server for Protein Fold Recognition and Structure Prediction Using Evolutionary Hybrid Profiles.” *Scientific Reports* 6(March): 1–10.
- Phyre2: Kelley, Lawrence A et al. 2016. “The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis.” *Nature Protocols* 10(6): 845–58. <http://dx.doi.org/10.1038/nprot.2015-053>.
- SWISS-MODEL: Waterhouse, A., Bertoni, M. et al. 2018 “SWISS-MODEL: homology modelling of protein structures and complexes”. *Nucleic Acids Res.* 46: W296-W303.
- Tr-Rosetta: Du, Zongyang et al. 2021. “The TrRosetta Server for Fast and Accurate Protein Structure Prediction.” *Nature Protocols* 16(12): 5634–51.
- TopModel: Mulnaes, Daniel et al. 2020. “TopModel: Template-Based Protein Structure Prediction at Low Sequence Identity Using Top-Down Consensus and Deep Neural Networks.” *Journal of Chemical Theory and Computation* 16(3): 1953–67.
- I-TASSER: Yang, Jianyi et al. 2014. “The I-TASSER Suite: Protein Structure and Function Prediction.” *Nature Methods* 12(1): 7–8.
- CD-Hit: Li, Weizhong, and Adam Godzik. 2006. “Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences.” *Bioinformatics* 22(13): 1658–59.
- MEGA7: Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. 2016. “MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.” *Molecular Biology and Evolution* 33(7): 1870–74.