Constructing Gene Co-Expression Networks to Capture Interaction Modules Contributing to Biosynthesis and Accumulation of Iridoid Glycosides in a Medicinal Herb *Picrorhiza kurroa*

Thesis submitted in partial fulfilment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

By

ASHISH SHARMA (E19SOE815)



Department of Biotechnology, School of Engineering and Applied Sciences

Bennett University (Established under UP Act No 24, 2016) Plot Nos 8-11, Tech Zone II Greater Noida-201310, Uttar Pradesh, India Aug 2022 Dedicated to family, friends, and the research community

@ Copyright Bennett University, Greater Noida, India Aug 2022 ALL Rights Reserved

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the Ph.D. thesis entitled "Constructing Gene Co-Expression Networks to Capture Interaction Modules Contributing to Biosynthesis and Accumulation of Iridoid Glycosides in a Medicinal Herb *Picrorhiza kurroa*" submitted at Bennett University, Greater Noida, India, is an authentic record of my work carried out under the supervision of Dr. Rajinder Singh Chauhan. I have not submitted this work elsewhere for any other degree or diploma.

I am fully responsible for the contents of my Ph.D. Thesis.

Ashish Sharma

Department of Biotechnology

Bennett University, Greater Noida, India

25.08.2022

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled "**Constructing Gene Co-Expression Networks to Capture Interaction Modules Contributing to Biosynthesis and Accumulation of Iridoid Glycosides in a Medicinal Herb** *Picrorhiza kurroa*", submitted by **Ashish Sharma** at **Bennett University, Greater Noida, India**, is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Dr. Rajinder Singh ChauhanDean (Research & Consultancy) &Prof. & Head, Department of BiotechnologyDate: 25.08.2022

ACKNOWLEDGEMENT

First and foremost, I'd want to thank my research supervisor, Dr. Rajinder Singh Chauhan. He is someone who showed me how to conduct effective scientific study. Throughout my PhD, he has been quite conscientious about managing his time. I appreciate all his efforts to produce ideas, smart recommendations, and funds to help my PhD production. His research dedication, energy, and enthusiasm were incredible and a source of inspiration for me, especially during difficult moments in my PhD quest. I am also grateful for the opportunity he has given me to work on such a magnificent project, and my affiliation with him will be remembered for the rest of my life.

I would also like to thank the Department of Biotechnology, Ministry of Science and Technology for providing financial assistance under Centres of Excellence in the form of a Senior Research Fellow (SRF) throughout my tenure.

I am particularly grateful to my friends and colleagues Anjali Kharb, Roma Pandey, Aditi Singh, and Sumit Joshi, as well as Research Associates Dr. Poornima Sharma and Dr. Dipto Bhattacharya, for their assistance and insight into the study. I'd like to thank my friend and fellow bioinformatician Shilpa Sharma in particular for her ongoing assistance and wonderful support during my study. I am grateful to the BU administration, Mr. Vineet Jain (Chancellor, BU), Dr. Prabhu Kumar Aggarwal (Vice Chancellor, BU), Mr. Sudhanshu Varma (Chief Operating Officer, BU), Col. (Retd) Guljit Singh Chadha (Registrar) and BU Information Technology and administration staff for pointing me in this direction and motivating me to go on this adventure.

A special note of thanks to my PhD committee members Dr. Shivani Goel, Dr. Deepa Khare, Dr. Sridhar Swaminathan, Dr. Deepika Vatsa for their fruitful suggestions during perusal of my doctoral work. I'd want to express my heartfelt appreciation to all of the academic members in the Biotechnology department for their ongoing assistance and recommendations. I am appreciative to the department's technical team.

Finally, I want to thank my mother, father, and brothers for their unwavering love, care, and support, that have encouraged me to achieve a PhD degree.

TABLE OF CONTENTS

Subject			
	No.		
ABSTRACT	i		
LIST OF ABBREVIATIONS	ii		
LIST OF FIGURES			
LIST OF TABLES			
CHAPTER 1			
INTRODUCTION			
CHAPTER 2 DEVIEW OF LITED ATUDE	8		
REVIEW OF LITERATURE	8		
	9		
2.1.1. Iridoid glycosides of <i>Picrorhiza kurroa</i>	10		
2.1.2. Biomarker studies in Picrorhiza kurroa	11		
2.2. Developments in Next Generation Sequencing analysis			
2.2.1. Next generation sequencing	12		
2.2.2. Transcriptome sequencing	13		
2.2.3. Computational analytics of NGS data	13		
2.2.4. Quality control and trimming	14		
2.2.5. De novo assembly	14		
2.2.6. Functional annotation	15		
2.2.7. In silico transcript abundance	15		
2.3. Biological networks	16		
2.3.1. Protein-Protein Interaction networks			
2.3.2. Metabolic networks	17		
2.3.3. Gene interaction networks	17		
2.3.4. Gene/ transcriptional regulatory networks	18		
2.3.5. Cell signalling networks	18		
CHAPTER 3			
MATERIAL AND METHODS			
3.1. Dataset information	20		
3.1.1. Transcriptome dataset	20		
3.1.2. Library preparation of transcriptome dataset			

3.1.3. De novo transcriptome assembly				
3.1.4. Functional annotation of assembled transcriptomes				
3.1.5. Extraction of expressed transcripts				
3.1.6. Mapping transcripts to other corresponding transcripts				
3.1.7. Generation of gene expression matrix				
3.1.8. Network generation and visualizations				
3.1.9. Generation of subnetworks of gene co-expression.				
3.2. Visualization of acyltransferases related networks	25			
3.2.1. Capturing acyltransferases based on involvement in co-expression	25			
networks				
3.2.2. Molecular modelling and docking	27			
3.3. De novo GBS assembly and library preparation	29			
3.3.1. Identification of Single Nucleotide Polymorphisms (SNPs)				
3.3.2. Mapping SNPs to gene co-expression networks				
3.3.3. Analysis of functional module from gene co-expression network				
3.3.4. Categorization of the GBS sample based on the PI and PII	31			
concentration in Shoot, roots and Stolons.				
CHAPTER 4				
RESULT	33			
4.1. Gene co-expression networks using NGS-transcriptome datasets of	34			
different organs/tissues and developmental stages of Picrorhiza kurroa				
4.1.1. Reads generation and De novo sequence assembly	34			
4.1.2. Functional annotation of P. kurroa transcriptomes	35			
4.1.3. Comparative co-expression network analysis	38			
4.1.4. Comparative co-expression network analysis between the shoot-only	40			
systems				
4.1.5. Comparative co-expression network analysis between PKS15 vs				
PKSS				
4.1.6. Comparative co-expression network analysis between PKSS >	48			
PKS15 > PKS25				
4.1.7. Comparative co-expression network analysis between PKS15 vs				
PKSTS				

4.1.8. Comparative co-expression network analysis between PKSS vs			
PKSTS			
4.1.9. Comparative co-expression network analysis between PKSTS vs	57		
PKSR			
4.1.10. Comparative co-expression network analysis between PKSS,	60		
PKSTS, and PKSR.			
4.2. Co-expression modules contributing to secondary metabolite	62		
biosynthesis in Picrorhiza kurroa			
4.2.1. Structural analysis of individual co-expression networks highlighting	62		
acyltransferase function in different samples	02		
4.2.2 Comparison of acyltransferase networks of experimentally cultured	70		
shoot samples (PKS-15 and PKS25 versus field-grown shoots	70		
(PKSS)			
423 Comparison of acyltransferase networks among different tissues of	70		
field grown plant samples (PKSS_PKSTS and PKSR)	19		
4.2.4 PAHD acultransformed subnetwork conturing as expressing modules	0.1		
4.2.4. BATTD acylitalisterase subletwork capturing co-expressing modules	84		
related to indold grycoside biosynthesis in shoots, stololis and roots			
4.2.5. Modelling and molecular docking of BAHD-acyltransferases to	85		
identify specific intermediates for last step modification of iridoid			
glycosides			
4.3. Identification and mapping of single nucleotide polymorphisms (SNPs)	89		
to the global co-expression networks			
4.3.1. Reads generation and De novo sequence assembly of GBS data	89		
4.3.2. Assembly of transcriptomes data	90		
4.3.3. Functional annotation of Picrorhiza kurroa transcriptomes	91		
4.3.4. Generation of gene co-expression networks	91		
4.3.5. Mapping GBS fragments of different populations on transcriptome	91		
dataset			
4.3.6. Gene co-expression network analysis of picroside % population in	93		
different tissues.			
4.3.7. Gene Ontology analysis of interacting nodes in the networks	93		

4.3.8. Identification of hubs with SNPs based on P-I % in shoots			
4.3.9. Network analysis of unique hubs for P-I% in shoots corresponding to			
different transcriptome samples			
4.3.10. Network analysis of common hubs of P-I% population	96		
4.3.11. Identification of hubs with SNPs based on P-I+P-II % content in	96		
stolons			
4.3.12. Network analysis of Unique hubs for P-I+P-II % in stolons	97		
corresponding to different transcriptome samples			
4.3.13. Network analysis of common hubs based on P-I+P-II % in stolons	97		
4.3.14. Identification of hubs with SNPs based on P-II % in roots	98		
4.3.15. Network analysis of unique hubs for P-II % in roots corresponding	98		
to different transcriptome samples			
CHAPTER 5	100		
DISCUSSION	100		
5.1. Co-expression networks of different organs/tissues and developmental	101		
stages of Picrorniza kurroa.	100		
biosynthesis in Picrorhiza kurroa	109		
5.3.Single nucleotide polymorphisms (SNPs) mapped to the global co-			
expression networks			
CONCLUSIONS	116		
Limitations of the Study	116		
FUTURE PROSPECTS			
LIST OF PUBLICATIONS			
REFERENCES	119		

ABSTRACT

Plants are repertoire of bioactive chemical entities. The current knowledge of biosynthetic routes in medicinal plants are mostly based upon target based molecular biology approaches, therefore, only partial information is available. The recent developments of Next Gene Sequencing (NGS) have generated high quality datasets that have provided key components related to various biological functions particularly in medicinal plant species. The last decade has resulted in development of some special computational tools to do high throughput analysis pin-pointing components representing functional modules. The gene co-expression networks are one such strategy based on the concept of graph theory that represents the relationship between genes based on gene expression in different circumstances i.e., tissue specific, temperature, disease phenotypes, etc. The current study focused on unravelling the complexity of the biosynthesis of iridoid glycosides in hepatoprotetive medicinal herb Picrorhiza kurroa. Furthermore, comparative gene coexpression-network analysis among transcriptomes derived from different tissues/ organs varying for iridoid glycosides pinpointed major hubs of acyltransferases belonging to BAHD-ATs. The study also developed a computational approach aimed at identification of SNPs through GBS analysis in a collection of 41 Picrorhiza kurroa populations, followed by construction of gene co-expression networks and mapping those SNPs to functional modules (hubs) capturing their functionality. Overall outcome of study has practical implications in designing genome engineering strategy for controlled production of iridoid glycosides.

LIST OF ABBREVIATION

PKS15	<i>Picrorhiza kurroa</i> tissue culture shoots grown at 15°C.		
PKS25	Picrorhiza kurroa tissue culture shoots grown at 25°C.		
PKSS	Picrorhiza kurroa shoots from plants grown at natural field conditions.		
PKSTS	Picrorhiza kurroa stolons from plants grown at natural field conditions		
PKSR	Picrorhiza kurroa roots from plants grown at natural field conditions.		
RNA-seq	RNA sequencing		
bHLH	basic helix loop helix		
MEP	non-mevalonate		
MVA	mevalonate		
TF	Transcription Factor		
ACT	Acyltransferase		
BAHD-AT	BAHD-acyltransferase		
SNP	Single Nucleotide Polymorphism		
Nr	Non-redundant		
CDS	CoDing Sequence		
RPKM	Reads Per Kilobase per Million Mapping Reads		
FPKM	fragments per kilobase of transcript per million fragments mapped		
ТРМ	Transcripts Per Million		
PPI	Protein-protein interaction		
TRN	Transcriptional regulatory network		
GRN	Gene regulatory network		
PI	Picroside I		
PII	Picroside II		
PDR	Pleiotropic drug resistance protein		
FCL	Farnesyl cysteine lyase		
СҮР	Cytochrome P450		
ССД	Carotenoid cleavage dioxygenase		
LcyE	Lycopene epsilon cyclase		
CAS	Cycloartenol synthase		

NRT	Nitrate transporter
STK	Serine-threonine protein kinase
IS	Isoprene synthase.
PRC	Pentatricopeptide repeat containing protein.
LRR-STK	LRR receptor-like serine-threonine protein kinase
ROS	Reactive oxygen species
PES1	Phytyl Ester Synthase 1
PES2	Phytyl Ester Synthase 2
UPGMA	unweighted pair group method with arithmetic mean
GBS	Genotyping by sequencing.
HPF-S	Fragments for high PI% in shoot
LPF-S	Fragments for low PI% in shoot
HPF-ST	Fragments for high PI+PII% in stolon
LPF-ST	Fragments for low PI+PII% in stolon
HPF-R	Fragments for high PII% in root
LPF-R	Fragments for low PII% in root
KEGG	Kyoto Encyclopaedia of Genes and Genomes
SND	Staphylococcal nuclease domain-containing protein
ABA	Abscisic acid
GhDREB2	Dehydration-Responsive Element-Binding Protein 2
AP2	APETALA 2
SQM	Squalene monooxygenase
DGAT	Diglycerol O-acyltransferase
CalS	callose synthase
MBOAT	Membrane bound O acyltransferase.
GABA	gamma-aminobutyrate
4CL	4-coumarateCoA ligase-like
PP2C	Protein phosphatase 2C 55-like
PEP	Plastid-encoded polymerase
FLN	Fructokinase-Like Protein

LIST OF FIGURES

Figure 2.1: Different classes of iridoid glycosides with catalpol backbone structure.

Figure 2.2: Different categories of biological networks

Figure 3.1: Flow diagram for the network generation and visualization.

Figure 3.2: Schematic workflow depicting generation and utilization of comparative coexpression networks leading to identification of acyltransferases transforming final modification of iridoid glycosides.

Figure 3.3: Structural transformations of acyl-moieties through BAHD-acyltransferases via catalpol towards formation of iridoid glycosides.

Figure 3.4: Systematic workflow for SNPs identification in GBS dataset, mapping to coexpression networks and identification of key components of various functional modules in *Picrorhiza kurroa*.

Figure 4.1. clustering of annotated transcripts of Shoots, Stolons and Roots of *Picrorhiza Kurroa* as per their length in base pair.

Figure 4.2: Top GO categories of transcripts of shoot, stolon and roots of *Picrorhiza kurroa* at Molecular function, biological process and cellular component levels of Gene Ontology.

Figure 4.3: Top KEGG categories of transcripts of shoot, stolon and roots of *Picrorhiza kurroa* at functional levels of KEGG Orthology.

Figure 4.4: Distribution of the dataset with TPM expression value greater than zero in the five transcriptome samples.

Figure 4.5: Distribution of nodes and edges for PKSS, PKS15, PKS25, PKSR, PKSTS followed by GO enrichment for identification of key component among the networks.

Figure 4.6: Common interactive sub-network modules between PKS15 and PKS25. Area of colours represent differential expression.

Figure 4.7: Unique interactive sub-network module of PKS15 when compared to PKS25.

Figure 4.8: Unique interactive sub-network module of PKS25 when compared to PKS15.

Figure 4.9: Common non-interactive sub-network module between PKS15 and PKSS. Area of colours represent differential expression.

Figure 4.10: Interactive unique sub-network module of PKS15 when compared to PKSS.

Figure 4.11: Interactive unique sub-network module of PKSS when compared to PKS15.

Figure 4.12: Common non-interactive sub-network module between PKS15, PKS25 and PKSS. Area of colours represent differential expression.

Figure 4.13: Common interactive terpenoid-glycoside sub-network module between PKS15 and PKSTS.

Figure 4.14: Interactive unique sub-network module of PKS15 when compared to PKSTS.

Figure 4.15: Interactive unique sub-network module of PKSTS when compared to PKS15.

Figure 4.16: Common interactive terpenoid-glycoside sub-network module between PKSS and PKSTS. Area of colours represent differential expression.

Figure 4.17: Common interactive terpenoid-glycoside sub-network module between PKSTS and PKSR. Area of colours represent differential expression.

Figure 4.18: Interactive unique sub-network module of PKSTS when compared to PKSR.

Figure 4.19: Interactive unique sub-network module of PKSR when compared to PKSTS.

Figure 4.20: Common interactive terpenoid-glycoside sub-network module between PKSS, PKSTS and PKSR. Area of colours represent differential expression.

Figure 4.20: Global co-expression network of acyltransferases extracted for *Picrorhiza kurroa* tissue culture shoots grown at 15°C (PKS15).

Figure 4.21: Global co-expression network of acyltransferases extracted for *Picrorhiza kurroa* tissue culture shoots grown at 25°C (PKS25).

Figure 4.22: Global co-expression network of acyltransferases extracted for *Picrorhiza kurroa* shoots grown in natural field conditions (PKSS).

Figure 4.23: Global co-expression network of acyltransferases extracted for *Picrorhizakurroa* stolons from plants grown at natural field conditions (PKSTS).

Figure 4.24: Global co--expression network of acyltransferases extracted for *Picrorhiza kurroa* roots from plants grown at natural field conditions (PKSR).

Figure 4.25: Comparative co-expression networks with linked nodes and edges of common acyltransferases between PKS15 and PKS25 transcriptomes.

Figure 4.26: Comparative co-expression networks with linked nodes and edges of common acyltransferases between PKS15 and PKSS transcriptomes.

Figure 4.27: Comparative co-expression networks of common acyltransferases with linked nodes and edges between PKS25 and PKSS transcriptomes.

Figure 4.28: Comparative co-expression networks of common acyltransferases with linked nodes and edges among PKS15, PKS25 and PKSS transcriptomes.

Figure 4.29: Comparative co-expression networks of common acyltransferases and linked nodes and edges between PKSS and PKSTS transcriptomes.

Figure 4.30: Comparative co-expression networks of common acyltransferases with linked nodes and edges between PKSTS and PKSR transcriptomes.

Figure 4.31: Comparative co-expression networks of common acyltransferases with linked nodes and edges among PKSS, PKSTS and PKSR transcriptomes.

Figure 4.32: Phylogenetic tree analysis for filtering potential BAHD acyltransferase from the transcriptomes for molecular modelling.

Figure 4.33: Distribution of SNPs based on the population and the type of category. (A): Distribution of SNPs among Low Population fragments. (B): Distribution of SNPs among High Population Fragments. (C): Distribution of HPF(Red) and LPF(Green) containing transcripts in global individual gene co-expression networks. (D): Subnetworks of transcripts containing HPF or LPF and their interacting.

Figure 4.34: Distribution of Gene Ontology modules interacting with SNPs containing transcripts. (A): Gene Co-expression network of transcripts with specific GO functions interacting with SNP containing nodes. (B): Distribution of transcripts mapping to specific GO functions.

LIST OF TABLES

Table 2.1: Comparison of various NGS Platforms.

Table 3.1: Transcriptome samples of *Picrorhiza kurroa* differing for growth, development and picrosides contents.

Table 3.2: Transcriptomes of *Picrorhiza kurroa* available for various geographical locations.

Table 3.3: Distribution and picrosides content of 37 populations of *Picrorhiza kurroa* representing different geographical regions of Himachal Pradesh.

 Table 4.1: Reads generation and de novo sequence assembly.

 Table 4.2: Distribution of nodes and edges in corresponding samples specific to terpene glycosides biosynthesis.

Table 4.3: Pairwise comparison of transcripts in PKS15, PKS25, and PKSS.

Table 4.4: Pairwise comparison of transcripts in PKSTS, PKSS, and PKSR.

Table 4.5: PKS15 vs PKS25 common modules.

Table 4.6: PKS15 vs PKSS common modules with respective TPM

Table 4.7: Distribution of TPM values of transcripts common in shoot samples PKS15 vsPKS25 vs PKSS.

Table 4.8: Distribution of TPM values of transcripts common in samples PKS15 vs PKSTS.

Table 4.9: Distribution of TPM values of transcripts common in samples PKSS vs PKSTS.

Table 4.10. Distribution of TPM values of transcripts common in samples PKSTS vs PKSR.

Table 4.11: Distribution of TPM values of transcripts common in samples PKSS vs PKSTS vsPKSR.

 Table 4.12: Distribution of nodes, edges and hubs of global and acyltransferase -specific co

 expression networks.

Table 4.13: Distribution of common nodes of acyltransferase subnetwork TPM values inPKS15 vs PKS25.

Table 4.14: Distribution of common nodes of acyltransferase subnetwork TPM values inPKS15 vs PKSS.

Table 4.15: Distribution of common nodes of acyltransferase subnetwork TPM values inPKS25 vs PKSS.

Table 4.16: Distribution of common nodes of acyltransferase subnetwork TPM values inPKS15 vs PKS25 vs PKSS.

Table 4.17: Distribution of common nodes of acyltransferase subnetwork TPM values inPKSS vs PKSTS.

Table 4.18: Distribution of common nodes of acyltransferase subnetwork TPM values inPKSS vs PKSTS.

Table 4.19: Distribution of common nodes of acyltransferase subnetwork TPM values inPKSS vs PKSTS vs PKSR.

Table 4.20: Outcomes of docking between BAHD-Acyltransferase Enzymes and their corresponding acyl-group donating compounds.

Table 4.21: Distribution of annotation, networks, and mapped SNP fragments on 8 transcriptome samples.

Table 4.22: Distribution of hubs specific for HPF-S and LPF-S identified for 8 transcriptome samples.

Table 4.23: Distribution of hubs specific for HPF-ST and LPF-ST identified for 8transcriptome samples.

Table 4.24: Distribution of hubs specific for HPF-R and LPF-R identified for 8 transcriptome samples.

CHAPTER 1 INTRODUCTION

Introduction

Secondary metabolites produced by plants include terpenoids, alcohols, alkaloids, acids, flavonoids, lignin, peptides, and others that are vital for growth and defence response processes [1] [2]. In response to external stimuli, the secondary metabolites biosynthesise occur through intricate interactions of several routes, modules, genes, enzymes, transcription factors, and transporters [3]. Every plant produces secondary metabolites of different classes, thus the biosynthetic machinery is also specialised according to distinct chemical entities [4]. Exploring each portion of the overall machinery is crucial to understand the overall biosynthetic process of a particular class of secondary metabolites. [5]. In order to address the complexity of biosynthetic machinery, several methods from molecular biology, biochemistry, genomics, genetics, and computational biology were utilized [6]. Recent developments in next generation sequencing in the fields of genomics and transcriptomics have made it possible to use genetic resources to unravel secondary metabolite biosynthesis complexity in ways that were previously unaddressed by conventional experimental methods. More crucially, it has provided chances for plant species with few genetic resources or no sequences [7]. Over the last decade, various diverse computational techniques have been developed for high-throughput analysis to pinpoint the components that constitute functional modules [8]. One such method based on the idea of graph theory is the gene co-expression networks, which depict the link between genes based on expression in different situations, such as tissue-specific, temperature, disease phenotypes, etc [9]. The fundamental idea behind this technique is to conceptualize global prespective of gene-gene interactions that take place to carry out overall system-level of functioning [10]. The "guilt-by-association" technique used in this strategy, which bases evaluation of the connectivity among gene relationships on the numerous functionalities, emphasises extremely relevant components in terms of interactions, differential expression, and functional relevance [11]; i.e., comparable gene expression profiles being observed across many RNA sequencing (RNA-seq) samples. This theory suggests that genes may link biological activity together. By comparing co-expression patterns across different plant species, computational analysis of gene co-expression networks can be used to determine important connections between regulators and targets, predict structural genes in metabolic pathways, and transfer gene functional annotations [12]. In a recent study, the biosynthesis of catechins, theanine, and caffeine in the tea plant, Camellia sinensis, was successfully evaluated using coexpression modules, and multiple hub genes that control the production of three metabolites [13], [14].

The creation of weighted gene co-expression networks from the transcriptomes of the medicinal plant Dioscorea nipponica allowed the identification of gene modules with characteristics related to dioscin control and production [15]. Using RNA-seq and co-expression network analysis in Soybean genotypes, potential genes for coursetrol biosynthesis and accumulation have been discovered [16]. New TFs and microRNAs were discovered as a result of the creation of composite networks of overlaying maps of co-expression of berry-specific regulators of the phenylpropanoid pathway [17]. A comparative study of 12 tissues from German and Roman chamomile was utilised to discover modules related to terpenoid and ester compounds using weighted gene co-expression networks [18]. Modules associated with variation in phenolics have recently been identified in barley transcriptome by weighted gene co-expression network analysis [19]. The transcriptomes of three grape species were analysed using a weighted gene co-expression network, which revealed 17 modules and two unique Anthocyanin levels, developmental phases, species, and regulation associated with the genes known as basic helix loop helix (bHLH) genes [20]. To comprehend the co-expression of various factors that occur collaboratively in the form of a network, capturing important hubs and interacting genes, the concepts of complex network theory are introduced in the current study. Thus, this approach will capture all potential important and interacting components at the complete system level of the medicinal herb *Picrorhiza kurroa*, not simply enabling comprehension of the system (in particular, morphological or biochemical phenotypic) components individually.

Medicinal herb, *Picrorhiza kurroa* endemic in the North-Western Himalayas is widely used in the preparation of various herbal drug formulations [21]–[23]. The medicinal and pharmacological value of herb has been reported as a hepatoprotective, neuroprotective, anti-oxidant, anti-tumor, anti-inflammatory and anti-diabetic activity, which have been attributed to iridoid glycosides (picrosides), primarily Picroside-I and Picroside-II [22], [24], [25]. Picroside-I is biosynthesized in shoots and Picroside-II in roots, thereafter both accumulate in stolons/rhizomes [26], which are major constituents of herbal drugs and also extracted as pure compounds. Kutkin, a compound made up of kutkoside and picrosides, is the plant's bitter and most active component [27]. P-I and kutkoside mixtures in certain ratios are typically necessary for hepatoprotective herbal medicine compositions [28]. *Picrorhiza kurroa* roots and rhizomes are used in the Ayurvedic herbal treatments Arogyavardhini, Tiktadya ghrita, Jatyadi ghrita,

Punarnavasava, and Nimbadi churna to treat skin conditions, ulcers, liver disorders, hyperacidity, and stomach issues [26]. Formulations based on picrosides, including Livplus, Livomyn, Livocare, Livotrit Forte, Tefroliv, and Picroliv, are commercialised and marketed by a number of biopharmaceutical companies, including BACFO Pharma, Dindayal Aushadhi Pvt. Ltd., TTK Pharma Pvt. Ltd., Zandu Pharma, and DIL Limited. Aucubin, geniopicroside, geniposide, verproside, and swertiamarin are among the various iridoid glycosides found in Picrorhiza kurroa [29]–[31]. This suggests that Picrorhiza kurroa is an important plant species for studies on the biosynthesis of iridoid glycosides. Picrorhiza kurroa is one of 242 plant species with a large yearly commerce, according to the National Medicinal Plants Board (Ministry of Ayush, Government of India). There is minimal room for recovery via vegetative reproduction in *Picrorhiza kurroa*'s natural habitats due to the steadily rising demand for herbal raw materials, particularly its roots and rhizomes [28]. The accumulation of these metabolites is influenced by environmental and genetic factors with the latter consisting of biosynthetic machinery components, kinases, transcription factors, pathway genes, and transporters. The primary goal of previous research on Picrorhiza kurroa has therefore been to create shoot culture platforms for the production of picrosides, either through metabolic engineering to increase and redirect metabolic flux to Picroside-I and Picroside-II biosynthesis or to optimise in vitro growth parameters of biomass that are similar to those found in natural habitat field conditions. [28], [32]-[35]. Trans-cinnamic acid individually shifts flux towards both pcoumaric acid and Picroside-I biosynthesis, whereas trans-cinnamic acid combined with catalpol directs maximal flux toward Picroside-I formation.[33]. These biochemical studies have shown that in order to increase endogenous picroside levels, Picrorhiza kurroa genes must undergo multistep engineering steps. Additionally, it has been suggested that it would be advantageous if P-II could be biosynthesized and stored in the shoots alongside P-I in order to lessen or even prevent the uprooting of Picrorhiza kurroa. [28]. Biosynthetic pathway of picrosides has been deciphered in *Picrorhiza kurroa* [7,8,13] wherein geranyl diphosphate formed from non-mevalonate (MEP) and mevalonate (MVA) pathways undergoes cyclization to form iridoid moiety which further condenses with a glucose molecule to form boschnaloside, which finally gets converted to catalpol by side chain modifications. P-I is formed by addition of a cinnamate moiety from the phenylpropanoid pathway to the sugar of catalpol, while P-II is formed by the addition of vanillic acid to the geranyl diphosphate backbone of catalpol [28], [33], [35]. Enzyme inhibitor and differential gene expression studies as well as molecular characterization approaches have deciphered crucial pieces of picroside biosynthetic

machinery, such as elucidation of route to geranioldiphosphate biosynthesis and characterization of two glucosyltransferases responsible for glucosylation of iridoid in Picrorhiza kurroa [35]-[37]. Reports on in-depth dissection of differential picrosides biosynthesis by organ-specific gene expression [32], [37] analysis of primary metabolism in picrosides accumulation [36], and identification of transcription factors (TFs) of picroside biosynthesis pathway [38] have further appended pieces of picrosides biosynthetic machinery. Hence, In the last decade approaches to pinpoint major components affecting the biosynthesis of Picroside-I and Picroside-II have been highlighted but leaving interactions among components undiscovered [28], [39], [40]. Apart from Picrosides, Boschnaloside, Aucubin, Bartsioside, Mussaenosidic acid, Deoxygeniposidic acid, and Geniposidic acid are among the iridoid glycosides that have been identified to be transitional substances in the catalpol biosynthesis pathway. Catalpol, an iridoid backbone of picrosides, is used to synthesise other iridoid glycosides by esterifying acyl groups (cinnamoyl, vanilloyl, p-coumaryl, benzoyl, etc.) to it. [41], [42]. In the instance of the biosynthesis of picrosides, blockage of the enzymes involved in the iridoid and shikimate/phenylpropanoid pathways has affected the total flux of picrosides by resulting in a shortage of the precursors for the iridoid backbone or acyl donors. [41], [43], [44]. Moreover, role of acylation in iridoid glycosides, including Picroside-I and Picroside-II is very important where acylation of catalpol via trans-cinnamoyl-CoA and vanillic acid occurs with the help of anthocyanin acyltransferase (ACT) leading towards picroside production [35]. The acylation of secondary metabolites biosynthesis has been considered as one of the most prominent and important steps in decorating final structures [45], [46]. This particular step is mainly executed by acyltransferases especially BAHD-acyltransferases [45]. The BAHD-acyltransferases (BAHD-ATs) particularly acylate using acyl-CoA thiosters (acyl donor), hence decorating last step modification of secondary metabolites [45]. BAHD acyltransferases are involved in the production of a number of secondary metabolites, including Montbretin, Spermidine, and other Phenolamides. Several molecular biology techniques have been used to identify and characterise BAHD-ATs, including molecular cloning, enzyme purification, transcriptome profiling, biochemical characterisation, and expression profiling [47], [48].

Furthermore, Single Nucleotide Polymorphisms (SNPs) specific towards secondary metabolites biosynthesis were extracted and mapped onto gene co-expression networks through a novel approach. The study has provided a unique strategy of gene co-expression networks that can also be implemented even in other species and domains of next generation sequencing

analysis. The variations in contents of Picroside-I and Picroside-II have highlighted the differences in their biosynthesis and accumulation under specific tissue, environment, and experimental conditions [32], [49]–[52]. Pandit et.al [32] highlighted different developmental stages and organs varying for picrosides . Furthermore, variations in picrosides contents have also been shown due to change in the geographical region [49]. As most of the herbal raw material is collected from wild or partly grown by farmers, thus warranting those genetic markers, preferably SNPs be developed from components of biosynthetic machinery. Molecular markers such as ISSR, AFLP and RAPD have been reported among populations of *Picrorhiza kurroa* using DNA finger printing strategies [49], [53]–[55]. however, such anonymous molecular markers can be present either in expressed or non-expressed regions of the genome [56]. Development of genetic markers such as SNPs could provide specificity provided derived from genes/transcripts with major role in the biosynthetic machinery of picrosides.

Therefore, to uncover interaction modules involved in the biosynthesis and accumulation of iridoid glycosides in a medicinal herb *Picrorhiza kurroa* three objectives were designed that have been discussed below. Firstly, we undertook approach based on gene co-expression networks analysis to unravel components of iridoid glycosides biosynthetic machinery of *Picrorhiza kurroa*, which were not captured through conventional molecular biology approaches. However, a coordinated visualization of transcriptional regulation of iridoid glycoside biosynthesis is still lacking, which can only be discerned through co-expression networks generated from whole-genome differential transcriptomics of different organs of *Picrorhiza kurroa*.

Secondly, we performed comparative gene co-expression-network analysis among transcriptomes derived from different tissues/ organs of *Picrorhiza kurroa* varying for contents of iridoid glycosides to pinpoint major hubs associated with BAHD-ATs. Our analysis also captured other components co-expressed with major Acyltransferases hubs, which provided us leads towards novel edges possibly contributing towards other components of biosynthetic machinery.

Third, we reported a computational approach aimed at identification of SNPs through GBS analysis in a collection of 37 *Picrorhiza kurroa* populations, varying for picrosides contents in roots, shoots, and stolons, followed by construction of gene co-expression networks and mapping those SNPs to functional modules (hubs) to capture their functionality viz-a-viz components of biosynthetic machinery. The effort of mapping specific SNPs on the transcripts

highlighted the presence of variation in the expressed regions of *Picrorhiza kurroa*. The key components containing specific SNPs were further highlighted using this novel approach. This resulted in identification of SNPs encompassing key components based of the high/low Picroside accumulation. Hubs containing population specific SNPs lying in the population specific gene co-expression network were shortlisted.

Research Gaps

- Which functional modules in networks are associated with the components of biosynthetic machineries of secondary metabolites biosynthesis in *Picrorhiza kurroa*?
- Can a novel strategy be developed to not only capture the functional modules in gene co-expression networks but also to map SNPs to functional modules/ hubs?

Research Objectives

- Build gene co-expression networks using NGS-transcriptome datasets of different organs/tissues and developmental stages of *Picrorhiza kurroa*
- Extract and prioritize co-expression modules contributing to secondary metabolites biosynthesis in *Picrorhiza kurroa*
- Identification and mapping of single nucleotide polymorphisms (SNPs) to the global co-expression networks

CHAPTER 2 REVIEW OF LITERATURE

2.1. Background

Plants have been utilized for diverse purposes to nourish the ecosystem by producing bioactive molecules with different chemical scaffolds [57]. Apart from being producers of a major share of food, the plants have essentially been consumed for pharmaceutical or nutraceutical purposes for ages [58]. These purposes are evident from the use of plants as medicine in the traditional methods of Ayurveda, Unani, Siddha, and traditional Chinese medicines [59]. In the modern scenario, the effectiveness of treatment approaches against diseases like cancer, diabetes, and fatty liver disease has led to the economic development of the herbal drug industry [60]–[62]. In India, almost 80% of the rural population has utilized traditional methods of treatment [63]. The global nutraceutical market has reached global trade in the billions of dollars with a rising growth rate every year [58]. The medicinal values of plant lies in the specialized biosynthetic machinery that yield various specialized secondary metabolites of diverse categories of terpenoids, alcohols, alkaloids, acids, flavonoids, lignin, peptides, etc [2]. The biosynthesis includes interaction of different biomolecule at various levels of pathways, genes, enzymes, transporters and transcription occurring in the systematic manner [3]. Therefore, to unravel the overall biosynthetic machinery such components are needed to be studied through various approaches of molecular biology, genetics, genomics, biochemistry, and computational biology [6]. The advances in Next generation sequencing analysis have enabled to pin-point the complexities by generating datasets of genomes, transcriptomes, and proteomes that cover everything globally [64][65]. Moreover, such approaches are highly compatible with those species in which less or no genomic resources are available [7]. Along with the combination of computational tools the researchers can reveal transcriptional complexities occurring for complex functionalities such as biosynthetic pathways [66]. The gene co-expression network analysis is widely used strategy that can highlight key components based on the linkage between genes and can also associate with functionally unknown genes through a network based identification [67]. This review of literature has widely been focused on the strategies uncovering key components in various functional modules in plants. The various studies in identification of functional system of medicinal herb Picrorhiza kurroa were also addressed in this section. Additional recent developments in Next Generation Sequencing analysis were taken into consideration, and they provided solutions to the complicated research problems in the contemporary scenarios. This section has also covered a quick discussion of network-based techniques and various resources utilized for the same.

2.1.1. Iridoid glycosides of Picrorhiza kurroa

As discussed previously, medicinal values of north-western Himalayan herb Picorhiza kurroa are in the specialized metabolites of various chemical classes. Iridoid glycosides are one such chemical class[68]. These basically monoterpene structures synthesised form 10-oxogernial that result in iridoid ring formation. The overall reaction is catalysed by iridoid synthase enzymes [69]. In Picrorhiza kurroa various specialized metabolites with catalpol backbone have been reported [28], [42] (Figure 2.1). Strategies to identify key components for such biosynthesis phenomena have been implied in the medical herb in last decade [28]. These strategies were mostly focused on two major iridoid glycosides namely Picroside-I and Picroside-II. Initially studies of HPLC analysis of tissue culture and field grown samples have shown differential accumulation in overall picroside content with change in temperature [51]. Furthermore, reporting of 15 pathway genes have shown direct or indirect relation with picroside accumulation [70]. These gene pathways were from two major biosynthetic pathways of MEP and MVA [70]. Difference in the protein expression have also been reported in the accumulating and non-accumulating conditions [71]. This lead to the designing of biosynthetic pathway with some gaps at last steps and intermediates through detection of biochemical intermediates by combination LC ESI and MS/MS techniques [42]. In another study expression analysis of key gene such as HMGR, PMK, DXPS, ISPE, GS, G10H, DAHPS and PAL were noted in differential tissue specific and picroside accumulating conditions [69]. Studies focusing on specific functional component such as transcription factors, miRNA and pathway enzymes have been implemented throughout [40], [72]. Various biomarker studies based on the. Furthermore, in past various NGS based studies have also been reported. This include generation of NGS-transcriptomes with differential picroside accumulation [73]. Comparative transcriptome analysis for pathway specific transcription factors have also been reported [38]. Furthermore recent studies of gene paralogues, transporters and transferases have shown potential in revealing the overall biosynthetic route in the plant [74]–[76]. Furthermore, the first genome of *Picrorhiza kurroa* has also been reported recently that can answer new questions that are not only related to iridoid glycosides but also for other growth and development functions [77]



Figure 2.1: Different classes of iridoid glycosides with catalpol backbone structure

2.1.2. Biomarker studies in *Picrorhiza kurroa*

The advances in plant breeding technologies have resulted in molecular breeding strategies that give low cost and high productivity[78]. Today, biomarker based diagnosis have led to crop improvement to breed species that sustain complex environmental conditions [79]. Molecular marker-based breeding has been achieved in model crop species that are associated with strong phenotype linkage the biomarkers. Some of the strategies of DNA finger printing have direct applications in the domain of molecular marker identification. Techniques such as RAPD, RFLP and SSR have direct applications in the biomarker identification [80]. These technique have also been implemented to identify the picroside specific biomarkers in *Picrorhiza kurroa* [49], [54], [55]. However, these techniques provide anonymous markers having chances of either present in expressed regions or not [81]. The development of Single Nucleotide

Polymorphism (SNP) based on the genome sequencing have highlighted the new ways to identify biomarkers related to phenotypic characteristics [82].

2.2. Developments in Next Generation Sequencing analysis

Previously developed sequences strategies of first generation were updated to more capable technologies known next-generation sequencing (NGS). These technologies were efficient enough to generate millions of reads without affecting the cost as compared to previous methods. In recent years these technologies have unveiled tremendous opportunities for data analytics that can aid the various fields of biological sciences [83]

2.2.1. Next generation sequencing

These technologies have gained immense popularity in the last decade due their cost effective and broad generation of datasets that can be analysed through customized analysis approaches and pipelines [83], [84]. Since, the single run generated millions of reads it is also important to care of error prone NGS runs through refinement strategies available in the form NGS-analytic pipelines. Furthermore, to compute such large amount dataset efficiently the role of high computing hard systems with parallel computing is also relevantly very important. Furthermore, it is also important to choose the NGS sequencing strategies wisely that should be based on availability of resources, datasets in public domain, the kind of reads required for the overall analysis subjected towards the problem [85].

2.2.2. Transcriptome sequencing

The complete set of RNA expressed to perform various functions is termed as transcriptome. The transcriptome is extremely complicated and contains several coding and noncoding RNA species. Historically, according to the fundamental tenet of molecular biology, RNA molecules were merely an intermediary between genes and proteins. Because they used the genetic code to encode proteins, messenger RNA (mRNA) molecules were the most widely investigated RNA species [86]. The transcriptome sequencing provide more insights at gene expressions and give assumptions of the final gene production in respect to external factors.

Advancement in the high-throughput sequencing technologies has transformed the field of transcriptomics by providing high quality transcriptome dataset through complementary DNA sequencing techniques [87]. Such tools are termed as RNA-seq studies that has successfully determined various questions that are sought to be answered. Successful applications of RNA-

Seq include mapping exon/intron boundaries, confirming or revising previously annotated 5' and 3' ends of genes, and precisely quantifying transcript levels. [88]. It describes a thorough bench-ready procedure for creating RNA-seq libraries for high-throughput pair- or single-end sequencing that is compatible with the Illumina sequencing platform [87]. The RNA-seq can be implemented through various sequencing platforms [89]. Some of the widely used commercially available platforms are Roche 454, PacBio, Illumina/ SOLiD, Nanopore and Helicos. These platforms used various sequencing approaches giving varying outcomes in term of reads length, quality, error, and cost effectiveness. Therefore, it is always important to choose the platform according to the study for which it has been utilized [89]. Among these Roche454 and PacBio are best known for longer reads generation when compared with illumina that generates reads only few hundred bases[89], [90].

Platform	Read length	Data Output	Run time
Illumina	150x2 PE	100 GB – 3 TB	29hrs- 4 days
Ion torrent	200-400 SE	60 MB- 50 GB	2.5 hrs-19 hrs
PacBio RS	3000 -15000	3 GB	20 min
454 GS	700	0.7 GB	23 hrs
FLX (Roche)			
SOLiD	85	15 GB	8 days
(Life)			

Table 2.1: Comparison of various NGS Platforms.

2.2.3. Computational analytics of NGS data

The dataset generated by such diverse strategies of NGS technologies have enabled the generation of new tools and software to analyse the dataset by using various computational implementations [91]. These tools are categorized based on the sequential usage to refine the dataset. The purpose of quality control, assembly/ reference sequence alignment, denovo annotation of unknown sequences, visualization, and quantification are included in these categories. Furthermore, tools are used in combination popularly known as pipeline in which process of providing subsequent output is executed [91]. Most of the NGS platforms also provide their in-house tools and software for processing NGS reads output for downstream processing. Some of the software representing a particular component of pipelines are discussed as under.

2.2.4. Quality control and trimming

The sequence raw reads generated by various NGS platforms are million in numbers therefore probability of being error prone is always a matter of concern [91], [92]. Quality score provided by the platforms are statistical values showing probability of error for each base in a particular raw read. The base-calling error probabilities, denoted by P, are used to calculate the Phred quality score, denoted by Q, which is specified as $10 \text{ Q P} = 10\log$. For instance, a Q30 value shows that accuracy is 99% and the likelihood of an inaccurate base call is 1 in 1000 [93]. The phred score criteria changes with respect the sequencing platform form which the raw reads are generated. The data filtering includes quality control and refinement based on the base call phred scores. Softwares like FastQC [94]and Trimmomatic [95]are widely used for this purpose. The refine goo quality read were taken further for the downstream analysis to get preferable outcomes.

2.2.5. De novo assembly

Following sequencing, millions of fragmented readings must be combined in accordance with the organism's chromosomes, requiring complex computations. With the advent of NGS platforms, read sizes shrunk, outperforming several current assemblers based on overlap graphs. A few of the significant assemblers are Trinity [96], ABySS [97], Velvet [71], and SOAPdenovo [71]

Due to overlapping graphs' inability to scale properly with rising read volumes, some of them use the well-known directed graphs known as de Bruijn graphs (advance overlap graphs). The reads are divided into fewer subsequences by k-mer in De Bruijn graphs. It adopts the strategy of joining together non-intersecting pathways into a single node. K-mer must be optimised for various values of k since it is not a fixed parameter for assembly. Depending on the size of k, different values of k result in various assemblies. To compare the quality of assemblies made with various k-mer, various measures are employed.

Two crucial metrics are as follows:

1. N50 value - The N50 value is the lowest size contig among larger contigs that can cover 50% of the transcriptome or genome.

2. Coverage - The proportion of nucleotides in the reference genome that are covered by assembled contigs. It can only be determined if a reference genome is available.

2.2.6. Functional annotation

Once the assembly is done it is important to annotate the assembled sequence by alignment strategies. Alignment is used to map the assembled nucleotide sequences in fasta format against various known databases[98]. Alignment tools such as BLAST [98], [99] and PLAST[100] are map assembled sequences against known databases. NCBI maintains the Nr (non-redundant) protein sequence database, which includes entries from a number of sources. Identical sequences from both curated and uncurated databases are combined into a single sequence in this extensive database. The prerequisites for merging two sequences are that they have the same length and contain the same residues throughout. A fasta signature (>) identifies each unique sequence, and control-A characters denote the separation of common sequences. By aligning the contigs or CDS to NCBI's non-redundant (nr) protein database, this database is most frequently utilised for functional annotation. The biological roles of freshly sequenced transcripts must be identified in detail in order to be used for downstream biological analysis.

2.2.7. In silico transcript abundance

We now have the benefit of deeply sequenced RNA-Seq data thanks to recent advances in nextgeneration sequencing technology (mRNA sequencing). Microarrays have been replaced by RNA-Seq, a technology that is in use today [91]. Through the use of massively parallel sequencing, cDNAs that correspond to an RNA fragment are translated into millions of short reads. For other transcriptome investigations, such as the de novo transcript assembly, these short reads can be used. It can directly take transcript sequences as an input for instance the transcripts produced by de novo transcriptome assembler. The counts to in silico expression profiling can be measured in:

- Reads Per Kilobase per Million Mapping Reads, or RPKM The formula for this level of measurement is RPKM= C/N*L. N stands for "Total number of mappable reads (in millions)," L represents for "Length of feature (in kb)," and C is for "Number of mappable reads on a feature (e.g., transcript, exon, etc.)."
- FPKM, or fragments per kilobase of transcript per million fragments mapped, is an acronym. It is similar to RPKM but uses transcript fragments rather than read counts.
- Transcripts Per Million is referred to as TPM. It is defined as :-TPM= (106)* Z *(C/N*L)
 Where additional Z parameter has been used to combat normalization factor [48].

2.3. Biological networks

The biological system consists of complex interaction between different entities of biomolecules such as gene, protein, mRNA, metabolite, etc. To understand such complication in globally the network-based approach is a widely accepted strategy to understand various levels of functions. The principles of graph theory are applied in understanding such complex systems [101]. The network or graph is a representation of components termed as nodes interaction with each other based on a particular relationship to form as linkage or an edge. The data analytics of high-throughput technologies like micro-arrays and RNA sequencing have lead toward system level identification of participating components [101], [102]. Such technologies have assured annotation of different levels of biological functions [103]. Networks are one of such approaches that identifies target based on the concept of "guilt by association" [102]. Therefore, it is always important to identify which component among the network module explains the behaviours of the system represented. Such key components are considered to be hubs for the network as they are frequently connected with most of the components by forming an edge [104]. Some of the types of biological networks are proteinprotein interaction (PPI) network [105], metabolic network [105], genetic interaction network[106], gene/ transcriptional regulatory network [107] and cell signalling networks [108]. These categories were organized on different level of biological molecule such as gene, mRNA, protein, and metabolite interacting to perform functional modules that can be further categorized.



Figure 2.2: Different categories of biological networks

2.3.1 Protein-Protein Interaction networks

PPI networks act as the overall framework of signalling route, that is regulated by environmental relationship toward cellular and genetic response [105]. The availability of large scale protein interaction information have enabled the study of protein interactions [109]. The study of the interaction between proteins have enabled us to deal with the three dimensional structural complexity by focusing on the interaction again stimuli [110]. Capturing interaction module can enable the modular design toward mechanism of action i.e. transport, enzymatic function, signalling and regulation [111], [112]. This information can be retrieved to various data resources designed especially for PPI networks. Some of the examples are STRING [113], GeneMANIA [114], FunCoup[115], I2D[116] and ConsensusPathDB[116].

2.3.2 Metabolic networks

Metabolic network is a network that comprises of interconnected biological pathways of chemical reactions performing metabolic activities necessary for cellular actives [117]. The biosynthesis of various chemical entities are end products of such metabolic networks[118]. The metabolites are specialize chemical entities that varying structurally, functionally, quantitatively and qualitatively from different species therefore found to be specialized [119]. Therefore, structural design of such metabolic network is important to understand the overall system of the organisms[65]. Some of the known information of the template metabolic pathways can be retrieve from various data resources such as KEGG[120], Reactome [111] and MetaCyc.

2.3.3 Gene interaction networks

It consist of a groups inter-connecting through functional relationship such as co-expression, ontology, similar function etc [121]. The gene interaction is an important aspect of understanding the relationship between genotype and phenotype [122]. The immense generation of Next Generation Sequencing dataset have created new questions in the varying genetic interactions with change in external stimuli [9]. Gene interaction network that represents interactions based on similar series of gene expression are known as gene co-expression network [9].

2.3.4 Gene/ transcriptional regulatory networks

Instructions for biological development and physiological reactions are encoded by transcriptional regulatory networks (TRNs). Recent improvements in computational modelling and genomic technology have transformed our ability to create models of TRNs [107]. Responses to intrinsic and environmental cues are closely regulated by a number of transcription factors (TFs). Gene regulatory networks (GRNs), which serve as a blueprint for the transcriptional controls driving development and environmental responses, are composed of these transcription factors (TFs) and their regulatory links [123].

2.3.5 Cell signalling networks

Networks of signalling pathways that connect receptors to various cellular machinery are created when signalling pathways come together. These networks process information in addition to receiving and transmitting signals. To comprehend how information is processed and how input-output linkages are established, computational models must be used because of the complexity of these networks [108]
Chapter 3 MATERIALS AND METHODS

The materials and methods have been described in following heads:

3.1. Dataset information

To achieve the objectives proposed in current study various datasets of transcriptomes and GBS raw reads of multiple accessions of *Picrorhiza kurroa* were utilized. The obtained datasets were processed through different NGS assembly pipelines, data visualization tools and software that would be discribed in the forthcoming sections. In short, a total number of 13 transcriptomes with different picroside contents, geographical regions, tissues, and experimental conditions were use in the study. The various NGS pipelines and protocols for library preparation, quality control, assembly, annotation, and quantification produced various outcomes. These outcomes were taken as inputs for further downstream analysis that includes generation and visualization of gene co-expression networks, identification of key components based on differential expression and functions of genes, sub-network extraction, and mining and mapping SNPs.

3.1.1. Transcriptomes dataset

Initially five transcriptomes from different tissues and experimental conditions were considered for the objective 1 and 2. These sample were previously quantified by HPLC for Picroside-I and Picroside-II concentrations [51]. These five samples were namely PKS-15 (Shoots grown at 15°C), PKS-25 (Shoots grown at 25°C), PKSS (Field grown shoots), PKSTS (Field grown stolons) and PKSR (Field grown roots) (Table 3.1). Shoot samples of both experimental and field grown conditions PKS-15, PKS-25 and PKSS were exclusive for Picroside-I and root sample PKSR only accumulates Picroside-II. whereas, both Picroside-I and Picroside-II were found to be reported in the stolon sample, PKSTS. In addition to that 8 transcriptomes reported by Varun et al [28] represented variation in accessions for Picroside-I and Picroside-II accumulation. The location of these samples was Hudan Bhatori, Moral Danda, Teita, Pattal Tissa, Dhel, Moral Danda, Salam Tith and Sural Bhatori annotated with unique accession code of PKS-1, PKS-5, PKS-4, PKS-21. PKST-3, PKST-5, PKST-16 and PKS-18, respectively.

S.No	Sample Name	Sample Condition	Picroside-I (%)	Picroside-II (%)	NGS High quality reads	Picrorhiza kurroa	Transcripts	Annotated Sequences
1	PKS-25	Shoots at 25°C	0.001	-	23,213,562		105,475	44,958
2	PKS-15	Shoots at 15°C	0.6	-	24,920,439	and the second s	129,865	47,726
3	PKSS	Shoots in field grown	2.7	-	21,211,113		150,566	40,117
4	PKSR	Roots in field grown	-	0.4	22,857,993	X	167,453	55,578
5	PKSTS	Stolons in field grown	1.77	0.99	20,910,870)))	244,558	66,979

Table 3.1: Transcriptome samples of Picrorhiza kurroa differing for growth, development and picrosides contents.

Table 3.2: Transcriptomes of *Picrorhiza kurroa*

 available for various geographical locations

S. No.	Picrorhiza kurroa	Location
	Accession ID	
1.	PKS-1	Hudan Bhatori
2.	PKS-5	Moral Danda
3.	PKS-4	Teita
4.	PKS-21	Pattal Tissa
5.	PKST-3	Dhel
6.	PKST-5	Moral Danda
7.	PKST-16	Salam Tith
8.	PKST-18	Sural Bhatori

3.1.2. Library preparation of transcriptome dataset

TRIzol Reagent (Invitrogen), RNeasy Mini Kit (Qiagen) and other kits were used to extract whole RNA from each sample. Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), NanoDrop (Thermo Fisher Scientific Inc.), and a 1 percent agarose gel were used to measure and quantify the total RNA in each sample. For the subsequent library preparation,

1 g of total RNA with RIN value higher than 7 was employed. The NEBNext® UltraTM RNA Library Prep Kit for Illumina® was used to create the next-generation sequencing libraries in accordance with the manufacturer's instructions. According to its effective concentration and anticipated data volume, the eligible libraries were fed onto HiSeq 2500 sequencer. This resulted in the generation of transcriptome raw reads, which were taken further for quality check analysis.

3.1.3. De novo transcriptome assembly

Transcriptome raw reads generated in the form of Fastq were further evaluated based on the quality control using FastQC [94]. Raw reads with adapters, N>10% and Qscore ≤ 5 were eliminated from further process of assembly using trimmomatic [95]. BinPacker[124], IDBA[125], and rnaSPAdes [126] were used to assemble cleaned raw reads from each unique transcriptome. The normalisation pipeline utilised was BBNorm from the BBMAp[127] package. The kmer lengths of 31, 25 and 60 were considered for BinPacker, rnaSPAdes and IDBA respectively. To filter and identify authentic transcripts, concatenated assemblies from of each these tools were taken. The transfuse programme (https://github.com/cboursnell/transfuse) was used to get transcript quantification by aligning clean reads with assembled transcriptome data after the true transcripts were initially identified using the evidential gene packages:tr2aacds.pl tool [128]. These outcomes were assembled sequence with larger sequence length considered to be termed as transcript.

3.1.4. Functional annotation of assembled transcriptomes

In order to find similarities between query sequences and huge databases, the annotation of the assemblies was carried out using PLAST [100], a tool based on seed-based heuristic algorithm against primary database libraries of NCBI, Uniprot etc. For gene, evolutionary, and functional annotation of the transcriptome, the eggNOG [129] database was utilised. GO [130] and KEGG [131] updated versions were utilized for functional enrichment of each dataset based on structural and functional annotations.

3.1.5. Extraction of expressed transcripts

The Transcript expression calculated in the form of quantification values according to Transcript per million reads (TPM) were caried out with the help of Salmon tool [132]. Each transcriptome sample transcripts with TPM value greater than zero were considered as expressed for the study. Transcripts sequence identifiers were extracted from the spreadsheet

of complete assembly using MS Excel. To extract the sequence from the list of identifiers of transcripts sequences Samtools faidx [133] was used to pooled out desired batch of sequences of interest (Figure 3.1)

3.1.6. Mapping transcripts to other corresponding transcripts

Each set of expressed transcripts was mapped with all other transcriptome sets by local standalone version of BLASTn[98]. The threshold criteria for mapping were 1-e09 evalue, greater than 75% query coverage and greater than 95% identity. Each individual set of transcripts was taken as query and all other sets were considered as subject and included in pooled form. The resulting tabular outcomes were further verified by match identity for each matched transcripts pairs of different transcriptomes samples (Figure 3.1).



Figure 3.1: Flow diagram for the network generation and visualization.

3.1.7.Generation of gene expression matrix

Matched pairs of transcripts of each set of transcript sequence were represented in the form coexpression matrix of each individual transcriptome. The matrices generated from each transcriptome represented differential expression profiles (TPM) of each expressing transcript of that sample. The name of rows represents transcript identifier and column names consist of sample ID. Five gene expression matrices with TPM values greater than 0 were the outcome of this step.

3.1.8. Network generation and visualizations

The networks of co-expression were generated using GENIE3[134] Package of R. GENIE is based on tree approach for regulatory network having multifactorial values of genes ranking the probable interaction in the form of link-list. Here the multifactorial information of coexpression from the mapping of transcript pairs were taken as input. For generation of link list default tree method "ensemble" with K (Number of candidate regulators randomly selected at each tree node) value 7, and the number of ensembles were set at 50. This generated a link list showing complex pairs of interaction ranked based on a connectivity score. The link list of connectivity score >0.005 were extracted from the generated link list. This resulted in generation of individual link list representing individual transcriptome same. The filtered linked list with optimum value of connectivity score, were considered as pairs of transcripts showing co-expression or interactions. These interactions were visualized in the form of co-expression network using the Cytoscape Network visualization tool. Network analyser was utilized to calculate number of interactions of each transcript represented in the network in the nodes, hence degree of freedom of each node was calculated. For representation of differential gene expression on the network, donut and pie chart were used as style showing relative expression among transcriptome samples. Nodes size and colours were styled based on the degree of freedom and expression in individual transcriptome condition. The generated network represents overall global system of transcriptomes therefore in order to analyse the networks toward the function of interest, subnetworks were generated.

3.1.9. Generation of subnetworks of gene co-expression.

Gene enrichment based on various levels of ontologies using GO enrichment [130] weas taken for the generation of subnetwork. The subnetwork based functional class of annotation for iridoid glycoside biosynthesis and acyltransferases were used as key to extract similar functional annotation. Nodes showing interaction with such functional interest were also extracted. This resulted in the generation of sub network representing key functions of interest.

3.2 Visualization of acyltransferases related networks

The subnetworks representing acyltransferase function were extracted from each global coexpression network to achieve the analysis. Apart from this similar style based on the degree of freedom with node size and pie chart representation based on differential gene expression were represented for these subnetworks. The comparative subnetworks analysis was done among the transcriptome samples to determine the commons with different interactions (Figure 3.2). The potential candidates were captured based on the function as a acyl group donor proposed to be decorating final structures of iridoid glycosides.

3.2.1 Capturing acyltransferases based on involvement in co-expression networks

The individual and comparative network analyses served as the foundation for network visualisations. Transcripts with the greatest degree of flexibility and expression in the network were found using individual network analysis, which examined the total interactions and expression of each node. Contrary, in a comparative study, the same network was examined for common nodes with a comparable class of acyltransferase function across various transcriptome samples based on their differential expression and distinctive relationships. The nodes involved in the biosynthesis of iridoid glycosides and those with distinctive interactions were chosen, and MEGA was used to compare their sequences for further elimination of redundancies. [135]. Based on alignment score, a phylogenetic tree was created, and sequences with longer sequences were chosen from each branch. The potential acyltransferases were taken forward for molecular modelling and docking analysis.



Figure 3.2: Schematic workflow depicting generation and utilization of comparative coexpression networks leading to identification of acyltransferases transforming final modification of iridoid glycosides.

3.2.2 Molecular modelling and docking

Transeq (https://www.ebi.ac.uk/Tools/st/emboss transeq/) was used to translate the selected transcripts' nucleotide sequences, and Pfam (http://pfam.xfam.org/) was used to identify the frames that included acceptable acyltransferase domains. The *ab-initio* technique of I-Tasser was utilised to construct 3D structures utilising sequences composed of acyltransferase domains (https://zlab.umassmed.edu/bu/rama/). The models that showed the best Z-score, Qmean, and C-value were chosen as the final candidates. Additionally, in the Swiss PDB viewer, the energy reduction of most of the optimal structures was done. Following that, acyl group-donating ligands for the special iridoid glycoside (Figure 3.3) were acquired from PubChem, and Marvin sketch was used to create their three-dimensional (3D) conformations. AutoDock Tools (http://autodock.scripps.edu/resources/adt) were used to construct the structures of proteins and ligands with acyl groups that were shortlisted. Using AutoDock vina[136], six modelled proteins were individually tested against each of the acyl-group ligands (Figure 3.2). The outcomes were attained using several protein-ligand complex conformations that were rated according to their binding affinities. More than -7 kcal/mol was classified as bad binding affinity, while conformations with binding affinities of less than -9 and -7 kcal/mol were rated excellent and good, respectively. So, for each acyltransferase that was modelled, possible ligands were chosen.



Figure 3.3: Structural transformations of acyl-moieties through BAHD-acyltransferases via catalpol towards formation of iridoid glycosides.

3.3 De novo GBS assembly and library preparation

Picrorhiza kurroa tissue samples were obtained from the nursery of the Himalayan Forest Research Institute in Jagatsukh, Kullu, Himachal Pradesh, India. Here, populations gathered from various North-Western Himalayan geographical regions were planted and maintained to to minimise the impact of environment and reflecting only the genetic differences. Data from each population's HPLC analysis were received from previous studies [52], [137]. For the purpose of building co-expression networks and identifying SNPs, HPLC data populations were divided into high and low picroside content populations. For the GBS study, the tissue samples from each group were also taken into consideration (Table 3.3). TES and C-tab techniques were used to extract the whole genomic DNA from each sample. Each sample's total DNA was measured and qualitatively analysed using nanodrop and a 0.8 percent agarose gel. DNA samples were added to individual adapter-containing tubes at a concentration of 10 ng/l. Following that, the materials in a 20-litre container were digested with ApeKI enzyme (New England Bio Labs, Lipswitch, MA). To ligate adapters to sticky ends, T4 ligase was employed in each well at the proper concentration. Double-purified digested DNA samples with unique barcode adapters were created using AMPure XP beads. The pooled DNA fragments from each library were then amplified in a 50-liter container that also contained 25 pmol of each common PCR primer, 1x Taq Master Mix (New England Bio Labs), and 10 ng of pooled DNA fragments (these primers contained complementary sequences for amplifying restriction fragments with ligated adapters, binding PCR products to oligonucleotides that coat the Illumina sequencing flow cell and priming subsequent DNA sequencing reactions). The library was cleaned with AMPure XP beads to get rid of unused dNTPs, enzymes, and other impurities. The PCRenriched library was quantified using Qubit 3.0 and evaluated in an Agilent Technologies 4200 tape station system using high sensitivity d1000 screen tape in accordance with the manufacturer's instructions. The SE illumina library was loaded onto NextSeq 500 for cluster creation and sequencing using 1X 150 bp chemistry after the Agilent Tape Station profile was used to determine the Qubit concentration for the libraries and the mean peak size. The quality of the raw readings generated from the samples was examined using FastQC. The data quality was assessed according to the base and read phred scores.

3.3.1. Identification of Single Nucleotide Polymorphisms (SNPs)

The picroside content of each unique raw read library of populations was further processed for SNP detection. Using Process radtags, the restriction site and barcodes were verified. Reads were demultiplexed and trimmed using the following parameters: a final read length of 120 bp, a phred33 quality score, a score limit of 10, the apeKI enzyme, the adaptor "ACACTCTTTCCCTACACGACGCTCTTCCGATCT," and an illumina filter to reject reads that was unacceptable. In a de novo pipeline, each stack component was sequentially run using default settings for the ustacks, cstacks, and sstacks. [138], [139]. Consequently, a library of fragments was created with a probable SNP frequency based on their existence in many



Figure 3.4: Systematic workflow for SNPs identification in GBS dataset, mapping to co-expression networks and identification of key components of various functional modules in *Picrorhiza kurroa*.

populations. Based on the amount of picrosides in each sample, two population groups were created. The total samples were split into two groups of populations with high and low picrosides concentration. This served as the starting point for the SNP population analysis, which produced a segment of SNP frequency found in high/low populations. For further

mapping of SNPs on the transcriptome dataset, fragments specific to high and low populations were aligned to different transcriptome datasets.

3.3.2 Mapping SNPs to gene co-expression networks

Fragments from the de novo assembly of populations' GBS data were aligned to the transcriptome dataset using BLASTn with more than 95% query cutoff and identity. A high or low score, indicating the population of the fragment, was assigned to each transcript mapping with SNP encompassing fragment. Each mapped node was assigned a unique colour. The global gene co-expression network's transcripts and associated edges were retrieved together with the GBS segments to which they were linked (figure 3.4).

3.3.3 Analysis of functional module from gene co-expression network

Transcripts with certain GO keywords were produced from the transcriptome dataset. These transcripts were used to extract functional modules, including edges of various functions, using the prey-bait approach. In samples containing different concentrations of picroside, these functional modules were contrasted (Figure 3.4).

3.3.4 Categorization of the GBS sample based on the PI and PII concentration in Shoot, roots and Stolons.

GBS samples were categorized in two different groups of high and low Picroside content in root, shoot and stolon. The rationale for grouping the population is to divide the dataset uniformly into two categories. The two populations of relatively high and low picroside content were divided for each set of tissue whereas the GBS sample with relatively moderate picroside content was not considered for further network analysis network analysis. The reason for not considering the moderate picroside content containing sample was to divide the samples in to two distinct populations with no intermediate relationship. (Table 3.3).

Table 3.3: Distribution and picrosides conten	nts of 37 populations	of Picrorhiza	kurroa representing	different
geographical regions of Himachal Pradesh.				

	D Lin Chaste		P-II in Roots				
Location	P-I In Shoots	Stolone (%)	(%)			Altitude	GBS raw
	(70)	3(010115 (70)		Accession Code	District	(meters)	reads
Hudan Bhatori	0.53	1.15	0.04	PKS-1	Chamba	3620	826564
Bhuri	5.38	0.75	0.05	PKS-2	Kinnaur	3330	693021
Dhel	1.28	2.35	0.23	PKS-3	Kullu	3597	655025
Teita	0.13	2.11	0.08	PKS-4	Chamba	3590	931418
Moral Danda	1.86	0.84	0.5	PKS-5	Shimla	3354	464062
Kundaghat	0.23	1.09	0.12	PKS-8	Kullu	3200	547961
Pulag Nath	1.76	1.75	0.25	PKS-9	Kinnaur	3435	866676
Dalau Pathar	2.53	1.85	0.14	PKS-10	Shimla	2703	653392
Rohtang	3.91	1.52	0.12	PKS-11	Kullu	3979	620324
Gue (Spiti)	2.33	0.87	0.13	PKS-13	Lahaul & Spiti	3671	963188
Yungpa	1.82	0.41	0.11	PKS-14	Kinnaur	3440	571912
Seri	1.3	1.36	0.02	PKS-15	Chamba	2135	465872
Salam Tith	1.72	0.65	0.039	PKS-16	Chamba	3440	987351
Chander Khani	0.64	0.58	0.33	PKS-17	Kullu	2354	720238
Sural Bhatori	1.59	1.26	0.35	PKS-18	Chamba	3323	765793
Banjar	0.28	0.57	0.03	Banjar_Site	Kullu	2866	509460
Katgaon	0.5	0.23	0.03	PKS-20	Kinnaur	3115	772934
Pattal (Tissa)	0.71	1.82	0.02	PKS-21	Chamba	3245	1057704
Tinnu Gaon	1.47	0.74	0.16	PKS-22	Lahaul & Spiti	3238	516258
Granfu	0.54	1.19	0.05	PKS-23	Lahaul & Spiti	3100	594021
Bhagi	0.93	0.1	0.13	PKS-24	Shimla	3100	566464
Bhrigu	0.24	3.69	0.06	PKS-25	Kullu	4170	862580
Shringul Tung	0.37	0.59	0.03	PKS-26	Shimla	3307	1145338
EXISTING STOCK (1 YEAR)	1.68	1.85	0.46	Existing_Stock-1_yr	-	-	963336
EXISTING STOCK (2 YEAR)	3.43	0.93	0.05	Existing_Stock-2_yr	-	-	750029
EXISTING STOCK (3 YEAR)	2.09	0.83	0.06	Existing_Stock-3_yr	-	-	616973
EXISTING STOCK (4 YEAR)	1.43	0.42	0.13	Existing_Stock-4_yr	-	-	490285
HAMTA	2.89	1.16	0.59	HAMTA	Kullu	4270	1024295
UDAYPUR	1.48	2.1	0.35	Udaypur	Lahaul & Spiti	2743	898944
MAYAR	0.74	3.59	0.24	Mayar	Lahaul & Spiti	2469	642869
PULGA	1.68	1.9	0.2	Pulga	Kullu	2895	1787390
BANSHERE	0.23	0.34	0.17	Bansheru	-	-	654585
CHITKAL SITE-2	0.27	0.85	0.42	Chitkal_Site-2	-	-	646347
JAMARI TOP	1.04	2.01	0.04	Jamari_Top	-	-	988001
KUNDAGHAT TOP	3.7	1.31	0.26	Kundaghat_Top	Kullu	3200	444717
TINUGAON SITE-2	0.14	0.255	0.35	Tinu_Gaon_Site-2	Lahaul & Spiti	3238	543182
KATGAON SITE-2	0.69	2.14	0.02	Katgaon_Site-2	Kinnaur	3115	911520

Medium

Low

High

CHAPTER 4 RESULTS

4.1. Gene co-expression networks using NGS-transcriptome datasets of different organs/tissues and developmental stages of *Picrorhiza kurroa*

4.1.1. Reads generation and De novo sequence assembly

Experimentally cultured samples of shoots at temperatures of 15° C (PKS-15) and 25° C (PKS-25) showing lesser Picrosides content compared to field-grown shoots suggested that in vitro stress response could be identified in these transcriptomes [140]. The tissues of shoots (PKSS), stolons (PKSTS) and roots (PKSR) grown in natural conditions contained higher amount of Picrosides [3–5]. In accordance with RNA-seq results for PKS15, PKS25, PKSS, PKSTS, and PKSR Illumina paired-end sequencing produced 51,756,778, 45,410,214, 44,340,806, 43,672,124, and 47,689,140 raw reads respectively (Table 4.1). For PKS15, PKS25, PKSS, PKSTS, PKSTS, and PKSR, the number of raw readings that were trimmed and cleaned and qualified for further processing was 49,670,218, 43,726,396, 42,261,006, and 41,707,042 respectively, (Table 4.1). PKS15, PKS25, PKSS, PKSTS, and PKSR final assembled transcript counts were 129,865, 105,475, 150,566, 244,558, and 167,453, respectively (Table 4.1).

Sample ID	Transcripts	annotated	Predicted gene	GO terms	KEGG Terms	eggNOG annotation
			name			
PKSS	150556	40117	6774	21774	18095	37352
PKS15	129865	47726	8526	27019	22069	44064
PKS25	105475	44958	8030	25453	20627	41622
PKSTS	244558	66979	7532	23868	19726	62166
PKSR	167453	55578	6713	21526	17808	51720

 Table 4.1: Reads generation and de novo sequence assembly.

4.1.2. Functional annotation of P. kurroa transcriptomes

EggNOG [129] was used as a resource database to construct orthology predictions for transcriptome assemblies that had been annotated using PLAST [100] at various taxonomic levels. From latest versios GO and KEGG pathway databases, groups of orthologs were subsequently functionally annotated [130], [131]. This led to the actual transcript annotation of PKS15, PKS25, PKSS, PKSTS, and PKSR (Figure 4.1). From hits in the Uniprot database,



Figure 4.1: clustering of annotated transcripts of Shoot, Stolon and Roots of *P.Kurroa* as per their length in base pair.

unigenes from PKS15, PKS25, PKSS, PKSTS, and PKSR, were retrieved in the numbers 8526, 8030, 6774, 7532, and 6713 respectively. Following, GO annotation, assembled sequences from PKS15, PKS25, PKSS, PKSTS, and PKSR, respectively, yielded 27019, 25453, 21774, 23868 and 21526 transcripts (Figure 4.2).



Figure 4.2: Top GO categories of transcripts of shoot, stolon and roots of *P. kurroa* at Molecular function, biological process and cellular component levels of Gene Ontology.

Additionally, KEGG database annotation identified 22,069, 20,627, 18,095, 19,726, and 17,808 transcripts, from PKS15, PKS25, PKSS, PKSTS, and PKSR respectively, (Figure 4.3).



Each of the five transcriptomes under analysis contained roughly 16% of annotated transcripts per million reads (TPM) values greater than zero. (Figure 4.4). Thus, only a small fraction of the RNA-seq dataset accounted for expression above zero TPM value.



Figure 4.4: Distribution of the dataset with TPM expression value greater than zero in the five transcriptome samples.

Workflow of network formation and visualisation, illustrating global visualisation of organspecific transcriptional regulation of picrosides production and biomass discussed briefly in Chapter 3. Finally, five distinct global co-expressed network modules were created using 8176, 8335, 8206, 7789, and 8402 unique transcripts from the RNA-seq data of PKS15, PKS25, PKSS, PKSTS, and PKSR, respectively. (Figure 4.4, Figure 4.5)



Figure 1 **Figure 4.5:** Distribution of nodes and edges for PKSS, PKS15, PKS25, PKSR, PKSTS followed by GO enrichment for identification of key component among the networks.

The number of nodes in these global co-expressed network modules were 2779, 2657, 2929, 2712, and 2972 for PKS15, PKS25, PKSS, PKSTS and PKSR, respectively (Figure 4.5). Further, the number of edges in these network modules were 16,165, 16,156, 14,527, 15,360, and 15,923 for PKS15, PKS25, PKSS, PKSTS and PKSR, respectively (Figure 4.5).

4.1.3. Comparative co-expression network analysis

For comparing co-expressed genes between any two or three transcriptomes discussed in this study, iridoid glycosides mainly from monoterpenoid background were considered, and GO terms aptly describing terpenoids and glycosides were extracted and mapped among the

networks, thus generating a distribution of nodes and edges (Table 4.2). Pair-wise comparisons describing unique and common nodes in these five transcriptomes have been mentioned in Tables 4.3 and 4.4. Co-expressed gene sub-networks of terpenoid glycosides generated from the global networks (Figure 4.5) aided us in assessing links between biosynthesis of iridoids at the organ-specific transcriptome and metabolome levels.

Table 4.2: Distribution of nodes and edges in corresponding samples specific to terpene glycosides biosynthesis.

Sample	Nodes	Edges
PKS15	148	253
PKS25	189	253
PKSS	236	235
PKSTS	152	271
PKSR	132	117

Sample for Comparison	Unique		Common		
PKS15 v/s PKSS (336*)	109 in PKS15	199 in PKSS	28		
PKS25 v/s PKSS (367*)	140 in PKS25	189 in PKSS	38		
PKS15 v/s PKS25 (288*)	110 in PKS15	151 in PKS25	27		

Table 4.3: Pairwise comparison of transcripts in PKS15, PKS25, and PKSS.

* Transcripts present in either or both samples.

Table 4.4: Pairwise com	parison of trans	cripts in PKSTS	, PKSS, and PKSR.
			, ,

Sample for Comparison	Unique	Common
PKSTS v/s PKSS (337*)	110inPKSTSPKSS	31
PKSR v/s PKSS (318*)	91 in 186 in PKSR PKSS	41
PKSTS v/s PKSR (245*)	113in104 inPKSTSPKSR	28

* Transcripts present in either or both samples.

4.1.4. Comparative co-expression network analysis between the shoot-only systems

The interactions that were common between PKS15 and PKS25 have been pictorially presented in (Figure 4.6), while the interactions found unique and specific to either PKS15 or PKS25 have



Figure 4.6: Common interactive sub-network modules between PKS15 and PKS25. Area of colours represent differential expression.

been presented in Figure. 4.7 and 4.8, respectively. Table 4.5 is descriptive listing of Figure 4.6 Thirty-six common interactions have been observed between PKS15 and PKS25 terpenoid glycoside sub-networks. The major hubs detected based on intense swarming of connected interactions are a group of TFs and the auxin response factor. The common node-wise interactions between PKS15 and PKS25 have highlighted auxin responsive interactions (Auxin response factor, Transport inhibitor response proteins, and Scarecrow-like proteins),

Comparison of averaged differential gene expression of nodes present in the common network interactions between PKS15 and PKS25 revealed 8.5 fold higher expression of Endoglucanase



Figure 4.7: Unique interactive sub-network module of PKS15 when compared to PKS25.

in PKS15 compared to in PKS25, while expression of Cellulose synthase-like genes was 12.2 fold higher in PKS25 compared to PKS15. The major interactions from terpenoid glycoside sub-network that were unique to PKS15 and PKS25 were compared. The major unique hubs in PKS15 that were identified in the decreasing order of the number of interactions were SND > Pleiotropic drug resistance protein (PDR) > Farnesyl cysteine lyase (FCL) > CYP > Carotenoid cleavage dioxygenase (CCD) > Lycopene epsilon cyclase. A unique interacting hub in PKS15 with a significantly lower expression (ash coloured) of Cycloartenol synthase (CAS) as the common node was also detected. Interactions that were present in PKS15 but separated from the other co-expressed nodes included Phytoene synthase and E3 Ubiquitin-protein ligase (RHA1B-like). The major unique hubs in PKS25 that were identified in the decreasing order by the number of interactions were CAS > Serine-threonine protein kinase (STK) > SND > CCD > LrgB-like family> Uncharacterized protein family> Asparagine synthetase> V-type proton ATPase catalytic subunit. A separate non-linked Nitrate transporter (NRT) domain was also identified in the unique interactions of PKS25.]. Thus, co-expression of *PDR* as a major hub in

PKS15 unique sub-network might missing in PKS25 co-expressed unique interactions. Further, a unique co-expressed interaction connecting a Terpene/Isoprene synthase (*IS*) to Tocopherol cyclase (terpenoid-methylated phenol interactions) and *CAS* have been noted in PKS15 while any interactions involving *IS* was missing in PKS25. On an average *STK*s were over four-fold upregulated and identified as the second major unique interacting co-expressed hub in PKS25.



Figure 4.8: Unique interactive sub-network module of PKS25 when compared to PKS15

Transcript	PKS15	PKS25
ABC transporter B family member	3.202414	8.100121
auxin response factor	7.650815	5.206076
Beige/BEACH domain	1.19136	2.106626
BTB POZ domain-containing protein	6.959855	3.749734
cellulose synthase-like protein	8.00497	97.53869
cytochrome P450	27.63615	10.94947
E3 SUMO-protein ligase	5.575157	5.050733
elongator complex protein	3.592034	4.218303
Endoglucanase	29.93674	3.527784
enhancer of mRNA-decapping protein	6.535803	4.646109
expressed protein	5.006141	2.398221
F-box kelch-repeat protein	5.629032	3.525826
FYRN	3.927243	3.921058
Glycosyltransferase	4.113697	4.817498
HELICc	3.328089	2.437997
homeobox-leucine zipper protein	6.291664	5.015253
homeodomain protein	2.511032	7.566019
serine threonine kinase	5.109808	21.18872
LRR receptor-like serine threonine-protein kinase	5.088693	2.907978
Pentatricopeptide repeat-containing protein	4.470103	3.901911
peroxisomal fatty acid beta-oxidation multifunctional protein	13.19421	3.431917
potassium transporter	3.316704	3.790901
protein EXECUTER 1, chloroplastic-like	4.897749	4.679127
pyrophosphate-energized vacuolar membrane proton	10.20231	16.47162
resistance protein	4.904869	8.639097
response regulator	4.198561	6.231456
Scarecrow-like protein	4.02178	3.64239
serine threonine-protein kinase	7.415561	4.686767
Splicing factor 3B subunit	5.892997	9.321721
Staphylococcal nuclease domain-containing protein	5.121077	6.891662
Transcription factor	13.74189	7.362943
transport inhibitor response	19.35024	23.45225
WD domain, G-beta repeat	4.574745	3.685862
zinc finger	1.59262	7.98217
zinc finger CCCH domain-containing protein	22.05254	16.55951

 Table 4.5: PKS15 vs PKS25 common modules.

4.1.5. Comparative co-expression network analysis between PKS15 vs PKSS

We could not assess any common interactions between PKS15 and PKSS except for a common co-expressed interaction between *PDR* to RNA recognition motif. Common nodes (not common interactions between nodes) were included as a basis to compare PKS15 and PKSS. The common non-interactive nodes and respective unique interactive nodes upon comparison between PKS15 and PKSS have been pictorially represented in Figure. 4.9, and Figures. 4.10 and 4.11, respectively. Table 4.4 is descriptive listing of Figure 4.9.

The common nodes of interest were PDR > IS > FCL > MYB family TF > CCD > CYP. Thus,



PKS15 PKSS

Figure 4.9: Common non-interactive sub-network module between PKS15 and PKSS. Area of colours represent differential expression.

the common nodes involved are those functional in terpenoid biosynthesis. As expected, *STK* promoting downstream ABA signalling has been downregulated 6.8-fold in PKSS compared to PKS15. *FCL* has been upregulated 2.2-fold in PKSS, Tocopherol cyclase has been downregulated 2.1-fold, *CAS* has been downregulated 1.8-fold, and *PDR* has been upregulated in PKSS. A GDSL esterase lipase was upregulated 3.6-fold in PKSS. Some of the hubs in

decreasing order of intensity of unique interactions in PKS15 are STK > PDR > FCL >Tocopherol cyclase while the major hubs in PKSS unique interactions are *ABA 8' -hydroxylase* > IS > PDR > TF > a Gibberellin related protein. The PKS15 unique interactions comprise 2 connected but separate co-expressed interacting hubs.



Figure 4.10: Interactive unique sub-network module of PKS15 when compared to PKSS

The smaller network proceeds via *SND* to a *STK* to *CCD* to Lycopene beta cyclase to *PDR* or *CYP* and completes the link back to *SND*. The major interacting network has a closed loop via a *NRT* to a TF to *CYP* to *CAS* to a *IS* to Tocopherol cyclase to the *NRT*. In an extended closed loop, *NRT* is connected to *CAS* to *SND* to the *NRT* via a *STK* and an *LRR- kinase*. For an analytical cross-comparison, PKS25 unique interactions consisted of a single closed co-expressed loop between a *STK* to a *CCD* to *FCL* to *SND* back to the *STK*. Thus, we identified unique interaction in PKSS with Tocopherol cyclase, *CAS*, *CYP*, *CCD*, and Cellulose synthase moved out from its closed main co-expressed loop, and major new hubs like *ABA* 8'

hydroxylase, *PDR*, Gibberellin related protein, and a Beta carotene hydroxylase imported into the main closed co-expressed loop of PKSS. Phytoene synthase has formed a separate coexpressed loop/ network in both PKSS and PKS15 unique co-expressed sub-networks denoting chlorophyll/photosynthetic metabolism while a unique Phytoene synthase co-expressed network was missing in PKS25 unique co-expressed sub-networks. Thus, while in PKS25 unique interactions, a prominent co-expressed network reveals ABA biosynthesis and growth cessation (*CAS*, *CCD*, *STK* as major hubs), in PKS15 the unique network negatively regulates



Figure 4.11: Interactive unique sub-network module of PKSS when compared to PKS15

ABA biosynthesis (*FCL* as a major hub) and has a Phytoene synthase and Cellulose synthase in interactive networks which highlights its promoted growth in comparison to PKS25. Contrary to both PKS15 and PKS25, the field transcriptome of PKSS manages to degrade ABA (*ABA 8' -hydroxylase* as the major hub in PKSS unique sub-networks) while at the same time brings in gibberellin signaling in co-expressed unique networks.

Transcript	PKSS	PKS15
armadillo beta-catenin-like repeat family protein	4.084408	1.19136
Carotenoid cleavage dioxygenase	6.871375	6.10379
Cycloartenol synthase	5.94054	10.78402
cytochrome P450	6.886614	7.401911
DNAJ heat shock N-terminal domain-containing protein	4.977611	11.5432
E3 ubiquitin-protein ligase RHA1B-like	16.46655	12.90126
farnesylcysteine	9.289215	4.277345
GDSL esterase lipase	18.9391	5.306687
interconversion of serine and glycine (By similarity)	2.247597	0.681495
Mitochondrial 2-oxoglutarate malate carrier	6.330201	4.692952
MYB family transcription factor	2.614725	5.287064
Nucleolar protein, Nop52 containing protein	6.352651	19.46438
Pentatricopeptide repeat-containing protein	3.924682	9.509451
Phytoene synthase	3.9439	5.195932
Pleiotropic drug resistance protein	7.483761	4.868731
resistance protein	1.050992	5.735146
Serine threonine protein phosphatase 2A 59 kDa regulatory subunit B'	2.675286	8.07069
serine threonine-protein kinase	1.617979	11.00489
Staphylococcal nuclease domain-containing protein	7.726296	16.34593
synthase	5.223625	6.666108
tocopherol cyclase	4.676084	9.588065
Transcription factor	22.9974	7.32575
WD domain, G-beta repeat	3.31076	4.89148
zinc finger	4.92195	11.91645
RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	7.36493	6.143348

Table 4.6: PKS15 vs PKSS common modules with respective TPM

4.1.6. Comparative co-expression network analysis between PKSS > PKS15 > PKS25

The common non-interactive co-expressed sub-network module between PKS15, PKS25, and PKSS have been pictorially represented in Figure 4.12, and listed in Table 4.7. Twenty-seven common nodes have been observed between PKS25, PKS15, and PKSS based on presence of nodes and organ-specific expression (no interactions). It can be assumed from above that the global co-expressed terpenoid glycoside specific sub-networks could be explained nicely based on interactions but were sufficiently non informative just based on the presence of co-expressed

nodes. Some of the prominent nodes were *STK*, Pentatricopeptide repeat containing protein (*PRC*), TF, *CYP*, *IS*, etc.



Figure 4.12: Common non-interactive sub-network module between PKS15, PKS25 and PKSS. Area of colours represent differential expression.

Transcripts	PKS15	PKS25	PKSS
E3 ubiquitin protein ligase	3.743413	3.070088	1.159328
Pentatricopeptide repeat-containing protein	9.338263	3.157428	3.583392
Transcription factor	4.843492	9.439253	85.21287
serine threonine protein kinase	15.33364	3.106397	4.278974
adaptor-related protein complex 5, zeta 1 subunit	6.262103	8.64577	3.987752
auxin response factor	6.329332	2.223681	6.421098
zinc finger CCCH domain-containing protein	4.60228	6.919893	3.738881
expressed protein	3.551046	8.296954	2.282987
zinc finger	4.932892	7.841449	2.974443
receptor-like protein kinase	2.065877	2.244506	0.609209
cytochrome P450	27.19445	11.52195	2.515298
zinc finger protein	2.306862	8.464385	4.016024
ATP-dependent RNA helicase	6.460118	6.86257	3.255702
phosphatase 2C	1.489639	9.859771	10.51774
calcium-dependent protein kinase	8.328569	31.13009	9.268935
LRR receptor-like serine threonine-protein kinase	4.984257	7.584049	3.568045
Retrotransposon protein	2.345301	5.329669	3.580926
resistance protein	1.006652	5.147355	2.245813
tubby-like F-box protein	6.127738	7.563347	4.95621
AAA-type ATPase family protein	5.650143	12.95124	3.296491
RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	17.14815	20.64412	18.24598
ubiquitin carboxyl-terminal hydrolase	15.44891	7.915596	8.493643
Ubiquitin fusion degradation protein UFD1	4.677805	8.886397	9.969154
synthase	44.12977	3.008815	4.011906
T-complex protein 1 subunit	4.064845	2.275716	4.609685
FBOX	4.783053	9.390411	3.781475

Table 4.7: Distribution of TPM values of transcripts common in shoot samplesPKS15 vs PKS25 vs PKSS

4.1.7. Comparative co-expression network analysis between PKS15 vs PKSTS

Common and respective unique interactions between PKS15 and PKSTS have been represented in Figures 4.13 - 4.15 respectively, while their descriptive listings can be obtained from Table 4.8. The common interactions between PKS15 and PKSTS were rare revealing massive differences between signalling in two organs grown under distinct conditions. Only two interactions, one involving a methyltransferase and the *SND* and the other between a *STK* and a *RING* domain protein were common between PKS15 and PKSTS. With almost all the common nodes revealing upregulation in PKS15, only the *PDR* was 3.3-fold upregulated in PKSTS. The unique PKS15 network reveals a closed loop from a *NRT* to a TF to *CYP* to *SND* to *CAS* to an *IS* to Tocopherol cyclase closing in on the *NRT. CYP* is linked through coexpression to a *FCL*. A LRR receptor-like serine-threonine protein kinase (*LRR-STK*), a *CCD*, and a Lycopene beta cyclase are linked to the *NRT* outside the closed network seemingly implying the prominent regulatory role of the co-expressed *NRT* in maintaining the metabolic balance between tetraterpenoids and apocarotenoids (eg., ABA) and other terpenoids such as the mono and triterpenoids. The PKSTS unique network has a closed co-expressed interacting loop connecting Lycopene beta cyclase via a highly expressed F-Box protein to *LRR-STK*, *SND*, *CCD* and closing the loop in Lycopene beta cyclase. While the closed loop in unique PKS15 only had a *SND* specifying fine tuning of mRNA regulation, PKSTS additionally has a F-Box like protein associated to *CCD* specifying protein degradation as a major fate of the coexpressed loop. In PKSTS unique sub-network, a co-expressed *FCL* is connected in series to a *NRT* that controls two wings, with co-expressed *ABA 8' -hydroxylase*, and a Beta carotene



Figure 4.13: Common interactive terpenoid-glycoside sub-network module between PKS15 and PKSTS.

hydroxylase on one wing and an *IS*, a Tocopherol cyclase, and a highly expressed aldehyde dehydrogenase on the other wing. Thus, *NRT* was eventually linked to *FCL*, *CCD*, and *IS* in both PKS15 and PKSTS unique sub-networks, but in PKSTS unique sub-network, *NRT* was additionally linked to a *ABA 8' -hydroxylase* indicating ABA degradation coupled to mono and triterpenoid biosynthesis. Stolons have been identified as the repertoire for P-II and P-I. *PDR* transporter was not directly linked to the terpenoid networks in PKSTS, suggesting products formed via terpenoid networks are not transported but stored. Further, unlike *IS* in unique

PKS15 networks, the *IS* in PKSTS unique networks was linked to a Respiratory burst oxidase suggesting possible involvement of reactive oxygen species (ROS) in inducing biosynthesis of mono and triterpenoids in stolons.



Figure 4.14: Interactive unique sub-network module of PKS15 when compared to PKSTS



Figure 4.15: Interactive unique sub-network module of PKSTS when compared to PKS15

in samples PK515 vs PK515.		
Transcripts	PKS15	PKSTS
Carotenoid cleavage dioxygenase	5.121077	4.248048
farnesylcysteine	5.163374	9.992099
LRR receptor-like serine threonine-protein kinase	8.858808	1.133794
lycopene beta cyclase	16.72553	1.956582
Methyltransferase	4.692952	3.327606
Pentatricopeptide repeat-containing protein	9.509451	2.262529
Phytoene synthase	5.673829	5.219277
Pleiotropic drug resistance protein	3.529382	11.59771
resistance protein	8.07069	2.353269
Retrotransposon protein	1.095883	1.336321
RING	11.17907	2.959283
serine threonine-protein kinase	5.484575	5.645493
Staphylococcal nuclease domain-containing protein	19.7393	4.31855
synthase	7.76275	6.884878
tocopherol cyclase	12.35878	4.97656
Transcription factor	7.32575	17.72097

Table 4.8: Distribution of TPM values of transcripts common in samples PKS15 vs PKSTS

4.1.8. Comparative co-expression network analysis between PKSS vs PKSTS

The common interactive sub-network module between PKSS and PKSTS has been pictorially depicted in Figures 4.16 and listed in Table 4.9. The compact closed common interactive network between PKSS and PKSTS consists of an *IS* linked in a closed loop series to a TF, a *STK*, a *FCL*, and a *ABA 8' -hydroxylase* closing the loop at the *IS*. The PDR transporter as expected was connected though not common to the closed loop. The *IS* in both PKSS and PKSTS have a conglomeration of co-expressed enzymes linked, suggesting biosynthesis of mono and triterpenoids in both shoots and stolons in field samples. The basic differentiation between PKSS and PKSTS unique terpenoid networks is the linkage mode of the PDR transporter, ie., whether the transporter is directly linked in the closed co-expressed network (in PKSS) or is co-expressed but not linked in the closed network (in PKSTS).


Figure 4.16: Common interactive terpenoid-glycoside sub-network module between PKSS and PKSTS. Area of colours represent differential expression.

Transcripts	PKSS	PKSTS
Abscisic acid 8'-hydroxylase	1.746285	2.746312
Beta-carotene hydroxylase	15.46688	4.690096
Cbl-interacting protein kinase	15.08495	11.27708
dna binding protein	2.579402	2.03751
DUF246 domain-containing protein	4.101885	4.994395
farnesylcysteine	9.289215	9.992099
Galacturonosyltransferase	1.770531	4.931684
Pentatricopeptide repeat-containing protein	3.800065	2.262529
Pleiotropic drug resistance protein	7.493578	11.59771
resistance protein	1.050992	2.353269
serine threonine-protein kinase	1.486694	3.828183
Staphylococcal nuclease domain-containing protein	7.43018	4.31855
synthase	5.0238	6.884878
tocopherol cyclase	4.676084	4.97656
Transcription factor	26.87006	21.72524
zinc finger	4.92195	2.482071

Table 4.9: Distribution of TPM values of transcripts common insamples PKSS vs PKSTS.

4.1.9. Comparative co-expression network analysis between PKSTS vs PKSR

The common and respective unique interactive sub-network modules between PKSTS and PKSR have been pictorially depicted in Figure 4.17 and in Figures 4.18 and 4.19 as well as listed in Table 4.10. The common network between PKSTS and PKSR consists of a *LRR-STK* linked in a co-expressed closed loop with a U-box domain containing protein, a TF, a retrotransposon protein linked to the *LRR-STK*. A non-linked co-expressed methyltransferase with sulfate transporters, and peroxidase was found to be a unique commonality between stolons and roots. The non-linked co-expressed sulfate network in both stolons and roots suggests diminished ABA signaling as a common network characteristic. A single network of six co-expressed genes connected in series constitute a linked network in the unique interactions of PKSR. These six genes are *CAS*, Receptor-like protein kinase, *STK*, an *IS*, Phosphatidylinositol 4-phosphate-5-kinase, and a *PDR*. A Cation-chloride co-transporter and a ABC transporter B family member was identified to be co-expressionally linked to this *IS*. The transporters directly linked to the *IS* might be required for the import of terpenoid and



Figure 4.17: Common interactive terpenoid-glycoside sub-network module between PKSTS and PKSR. Area of colours represent differential expression

phenylpropanoid moieties into the vacuole for the biosynthesis of picrosides. The PDR transporter linked to the series network might be essential to the export of picrosides to other organs, after biosynthesis in the roots. The unique interactions in PKSTS represent a NRT in three different series connected interactions. One of these series connections has a ABA 8' -

hydroxylase, and a Beta carotene hydroxylase. The other connected series has an Aldehyde dehydrogenase, Tocopherol cyclase, Respiratory burst oxidase and an *IS*. The other connected series has *FCL*, *SND*, a *CCD*, a Lycopene beta cyclase and a *ABCG* transporter among others. It is quite probable that since the *IS* is not directly co-expressionally linked to any transporters in PKSTS unique sub-network, the *NRT* involved in the network transports the terpenoid moiety to roots for further modification of the terpenoids (acylation) and is transported back to stolons



Figure 4.18: Interactive unique sub-network module of PKSTS when compared to PKSR.

for storage. Farnesyl diphosphate synthase, *NRT*, Ent-kaurene oxidase, Tocopherol cyclase, Beta carotene hydroxylase, *CCD*, and several other hubs were identified to exist as non-linked separate co-expressed small interactions.



Figure 4.19: Interactive unique sub-network module of PKSR when compared to PKSTS.

common in samples r KSTS vs r Ks	JK	
Transcripts	PKSR	PKSTS
E3 ubiquitin protein ligase	1.818831	7.659606
F-Box protein	2.056076	5.878629
U-box domain-containing protein	4.310822	5.609737
Retrotransposon protein	0.667123	1.336321
Methyltransferase	1.141686	3.327606
Nitrate transporter	2.181961	2.74321
transcription factor	4.895627	25.72951
zinc finger	6.14119	2.52747
cytochrome P450	2.332459	23.67505
synthase	5.246414	2.009604
Galacturonosyltransferase	4.397389	4.931684
Cbl-interacting protein kinase	5.983873	11.27708
pumilio homolog	2.495526	1.837802

Table 4.10. Distribution of TPM values of transcripts
common in samples PKSTS vs PKSR

4.1.10. Comparative co-expression network analysis between PKSS, PKSTS, and PKSR.

The common interactive sub-network modules between PKSS, PKSTS and PKSR have been pictorially depicted in Figure 4.20 and listed in Table 4.10. A comparative common interaction analysis between PKSS, PKSTS, and PKSR has revealed a TF as the common major hub. Apart from genes responsible for normal maintenance of any organ, glycosyltransferases, *IS*, Callose synthase, and *CYP* have been identified as the common hubs (Figure 4.20). It is thus proposed that biosynthesis of isoprene units, its condensation and cyclization as well as addition of glucose moiety to the terpenoid unit are integral to all the organs under comparison. Transporter specific common interactions are absent in these terpenoid glycoside sub-networks, revealing isoprene units are mostly exported via long distance transporters after glucosylation for further modifications like addition of phenylpropanoid moiety, etc. Likewise, after their biosynthesis they must be exported to other organs or to the vacuoles via transporters for storage or for other functions.



Figure 4.20: Common interactive terpenoid-glycoside sub-network module between PKSS, PKSTS and PKSR. Area of colours represent differential expression.

Transcripts	PKSS	PKSTS	PKSR
transcription factor	85.21287	1.885059	6.226975
cytochrome P450	2.515298	10.14081	2.074169
synthase	4.011906	16.05539	2.396971
serine threonine protein kinase	4.278974	1.06753	6.071904
DUF246 domain-containing protein	3.230478	4.208159	1.313394
signal peptide peptidase-like	8.576918	2.741929	5.399903
STYKc	3.93669	1.632868	0.874649
FBOX	3.781475	8.461979	1.824926
domain-containing protein	2.695508	1.351698	3.932251
resistance protein	2.245813	2.94954	2.450717
Pentatricopeptide repeat-containing protein	3.583392	1.355426	3.300776
UDP-glycosyltransferase	2.311334	39.70881	1.962242
LRR receptor-like serine threonine-protein kinase	3.568045	4.736941	2.682493
protein ethylene insensitive	12.65049	21.46408	4.772541
expressed protein	2.282987	5.748525	1.553024
DEAD-box ATP-dependent RNA helicase	5.012573	4.977827	3.312039
WD domain, G-beta repeat	3.202788	0.750093	2.458402
cyclin-dependent kinase	1.923536	7.897478	17.95435
KH domain	2.68719	1.262614	2.384842
Retrotransposon protein	3.580926	1.505254	1.107837
heat shock	7.778552	0.710258	47.9187
U-box domain-containing protein	6.856266	8.534279	4.509818
callose synthase	2.840461	2.063748	7.709904
Breast carcinoma amplified sequence 3	1.397153	3.864689	1.879397
ubiquitin carboxyl-terminal hydrolase	8.493643	23.26001	1.721892
RING	6.120659	2.99533	1.749573
glycosyltransferase	3.246409	3.591555	1.639171
UBX domain-containing protein	11.68704	2.408162	0.87147

Table 4.11: Distribution of TPM values of transcripts common in
Samples PKSS vs PKSTS vs PKSR.

4.2. Co-expression modules contributing to secondary metabolites biosynthesis in *Picrorhiza kurroa*

4.2.1. Structural analysis of individual co-expression networks highlighting acyltransferase function in different samples

The prey-bait strategy was used to extract acyltransferase networks with identified potential functions from global networks. This resulted in individual networks primarily representing acyltransferases and other transcripts, forming co-expression nodes. Eventually, 5 subnetworks were visualized, PKS-15 was observed least complex with only 60 nodes and 47 edges, whereas PKSR showed the highest complexity with 173 nodes and 167 edges in comparison with other samples. Among PKSTS, PKSS and PKS25, 125 and 142 nodes with 121 and 134 edges were observed in PKSS and PKS-25 respectively, whereas PKSTS had the lesser no. of edges (104) and nodes (115) (Table 4.10). The nodes with the maximum degree of freedom in PKS15, PKS25, PKSS, PKSTS, and PKSR ranged from 8 to 12 for above samples, deemed as hubs (Table 4.12). Based on the expression of interactive nodes in individual acyltransferase network highly expressing nodes other than acyltransferase were also captured, possibly playing significant role in the individual system. The results from this analysis revealed following observations in different samples. In PKS15, it was observed that "transketolase" and "chromosomal maintenance protein" were showing 3-fold higher expression whereas, "Universal Stress protein", "ATP-dependent Clp protease", "ATP-binding subunit clpA homolog", "transcription factor (WRKY24)", "ion channel" showed 2-fold higher expression than the average mean of 7.08 TPM (Transcripts Per Million) of nodes in the network (Figure 4.21).



Figure 4.20: Global co-expression network of acyltransferases extracted for *P. kurroa* tissue culture shoots grown at 15°C (PKS15).

In the case of PKS25, average TPM value was 8.08 and it was observed that edges annotating as "Heat shock cognate 70 kDa", "1- deoxy-D-xylulose 5-phosphate reductoisomerase", "Two-component response regulator-like", "PB1, amine oxidase", "ATP-dependent Clp protease ATP-binding subunit", "Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD", "delta1-



pyrroline-5-carboxylate", "G-type lectin S-receptor-like serine threonine-protein kinase",

Figure 4.21: Global co-expression network of acyltransferases extracted for *P. kurroa* tissue culture shoots grown at 25°C (PKS25).

"RNaseH family protein", and "subtilisin-like" had 2-fold higher expression (Figure 4.21). The average TPM for PKSS was 8.27, and the "late embryogenesis abundant protein", "beta-glucosidase", "glutamine synthetase", "protein transparent testa", "fructose bisphosphate aldolase", and "glyoxylate" showed >3-fold expression to average (Figure 4.22).



Figure 4.22: Global co-expression network of acyltransferases extracted for *P. kurroa* shoots grown in natural field conditions (PKSS).

In stolon sample, the average TPM value for transcripts was 8.01, "Polyphenol oxidase", "amine oxidase", "hydroquinone", and "cytochrome P450" connecting edge trasncripts showed 3-fold higher expression to the average. On the other hand, "Cbl-interacting protein kinase", "peptide methionine sulfoxide reductase", and "NAC domain" showed >2-fold expression (Figure 4.23).



Figure 4.23: Global co-expression network of acyltransferases extracted for *P.kurroa* stolons from plants grown at natural field conditions (PKSTS).

The overall average TPM value for the PKSR sample was 5.9, and the edge transcripts such as "Glutamate decarboxylase", "basic region leucine zipper motif 53", "Heat shock cognate 70 kDa", "EIN3-binding F-box protein", and "4-hydroxy- 3-methyl but-2-en-1-yl diphosphate" showed >3 fold expression compared with the average whereas "ZnF_C2HC", "phosphoenolpyruvate carboxykinase", "UDP-glucuronate 4-epimerase", "Digalactosyldiacylglycerol synthase 1", "tryptophan synthase", and "L-ascorbate oxidase" showed 2-fold higher expression to the average (Figure 4.24)



Figure 4.24: Global co--expression network of acyltransferases extracted for P. Kurroa roots from plants grown at natural field conditions (PKSR).

Table 4.12: Distribution	of nodes edges ar	d hubs of global	and acyltransferase	-specific co-ex	pression networks
Table 4.12. Distribution	or nouces, cuges ar	ia nuos or grobai	and acynnansierase	-specific co-cx	pression networks.

Tissue Sample	Description	Nodes & Edges in global network	Acyltransferases captured	Nodes & Edges in Acyltransferase Network	Major Acyltransferases in network
	Shoots at 15° C (PKS15)	Nodes: 2779 Edges: 16165	15	Nodes: 60 Edges: 47	lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex acyltransferase- like At1g54570 diacylglycerol O- acyltransferase lysophospholipid acyltransferase LPEAT1 (LOC105166172) acyltransferase- like At3g26840 S-acyltransferase 7 (LOC111408808) O-acyltransferase WSD1-like ubiquitin- conjugating enzyme
	Shoots at 25° C (PKS25)	Nodes: 2657 Edges: 16156	30	Nodes: 142 Edges: 134	O-acyltransferase acyltransferase- like At3g26840 acyltransferase- like At1g54570 phospholipid 1-acyl-sn-glycerol- 3-phosphate acyltransferase 4 (LOC105159207) acyltransferase- like S-acyltransferase 22 (LOC105160003)

					O-acyltransferase WSD1 (LOC105172709) S-acyltransferase O-acyltransferase (WSD1-like) membrane-bound O-acyltransferase C24H6.01c-like
	Field grown Shoots (PKSS)	Nodes:2929 Edges:14527	28	Nodes:125 Edges:121	acyltransferase S-acyltransferase BAHD acyltransferase O-acyltransferase S-acyltransferase At2g14255-like acyltransferase- like At1g54570 1-acyl-sn-glycerol- 3-phosphate acyltransferase long-chain-alcohol O-fatty- acyltransferase O-acyltransferase WSD1-like
)))	Field grown Stolons (PKSTS)	Nodes:2712 Edges:15360	23	Nodes:115 Edges:104	acyltransferase- like At1g54570 BAHD acyltransferase lysophospholipid acyltransferase LPEAT1-like (LOC105963377) S-acyltransferase 23 (LOC105169327) S-acyltransferase 7 (LOC111408808) O-acyltransferase 7 (LOC111408808) O-acyltransferase 3 (LOC105954659) S-acyltransferase 8 (LOC105179299)

					1
					Pentatricopeptide
					repeat-containing
					protein
					membrane-bound
					O-acyltransferase
					C24H6.01c-like
					Phospholipid
					diacylglycerol
					acyltransferase
					S-acyltransferase
					acyltransferase
					O-acyltransferase
				BAHD	
		Field grown Nodes:2972	40	Nodes:173 Edges:167	acyltransferase
1 1 1 1 1 1					Acyltransferase
XIIIS	Field grown				membrane-bound
11.480	Roots (PKSR) Edges:15923	Edges:15923			O-acyltransferase
ALLAN		2460010720			C24H6.01c-like
					long-chain-alcohol
					O-fatty-
					acyltransferase
					Argonaute
					Phospholipid
					diacylglycerol
					acyltransferase

4.2.2. Comparison of acyltransferase networks of experimentally cultured shoot samples (PKS-15 and PKS25 versus field-grown shoots (PKSS)

Comparative analysis of the shoot samples was done among individual samples since all the shoot samples showed different phenotypic characteristics as well as variable content of compounds in P. kurroa (Sharma et al., 2021). Network comparison between PKS15 and PKS25 resulted in the generation of a network having common nodes, and surprisingly, the common interactions were not observed. A total of 13 common nodes with potential TPM expression in both samples were observed, namely "1-acyl-sn-glycerol-3-phosphate acyltransferase 4", "acyltransferase-like At1g54570" (Phytyl synthase-2), ester At3g26840" (PES1), "callose synthase", "diacylglycerol O-"acyltransferase-like acyltransferase", "DNA-directed RNA Polymerase", "expressed protein", "F-box LRR-repeat protein", "lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex", "membrane-bound O-acyltransferase C24H6.01c-like", "O-

acyltransferase" (WSD1- like), "Pentatricopeptide repeat-containing protein", and "U-box domain-containing protein" (Figure 4.25) (Table 4.13).



Figure 4.25: Comparative co-expression networks with linked nodes and edges of common acyltransferases between PKS15 and PKS25 transcriptomes.

Transcript	PKS15	PKS25
acyltransferase	8.166106	6.426799
Acyltransferase-like protein	5.259263	5.982342
acyltransferase-like protein At1g54570	13.93166	9.994469
callose synthase	3.238465	4.426359
diacylglycerol O-acyltransferase	3.416613	8.780628
DNA-directed RNA Polymerase	2.1557	6.256761
expressed protein	2.99	5.762326
F-box LRR-repeat protein	3.847903	4.756498
Lipoamide acyltransferase component of branched-chain alpha-keto acid		
dehydrogenase complex	3.354092	7.445896
membrane-bound O-acyltransferase C24H6.01c-like	6.927565	6.305654
O-acyltransferase (WSD1-like)	5.595803	16.32642
O-acyltransferase WSD1-like	16.10202	5.043076
Pentatricopeptide repeat-containing protein	1.896695	4.944018
Phospholipid diacylglycerol acyltransferase	6.800648	10.74189
S-acyltransferase	10.33211	7.947102
Transcription factor	8.233841	3.053742
U-box domain-containing protein	9.694073	4.197774

 Table 4.13: Distribution of common nodes of acyltransferase subnetwork TPM values in PKS15 vs PKS25.

Most of the class had putative acyltransferase functions, whereas other functions such as synthase, polymerase, were also observed. Although, both the samples were grown in tissue culture stress conditions, the comparative networks presented common nodes with unique interactions in both the samples, for example, the common hub (related to acyltransferase) interacted with "transketolase", and "structural maintenance of chromosome protein" in PKS15, whereas, with serine/threonine protein kinase in PKS25 that may be obvious because PKS25 shoots are grown at a higher temperature than to its counterpart PKS15. Furthermore, networks of both experimental and field grown shoot samples were compared to pinpoint potential elements playing role in biosynthetic machinery. Common expressing transcripts in PKS15 and PKSS, namely "acyltransferase-like At1g54570 (PES2)", "callose synthase, DNAdirected RNA Polymerase", "E3 ubiquitin-protein ligase", "expressed protein", "Oacyltransferase WSD1-like", and "U-box domain-containing protein" were detected, wherein common transcripts network comprised 41 nodes and 35 edges specific to transcriptome samples. It was observed that the lack of overlapping edges in both the samples might be due to the differential expression of transcripts. Furthermore, the expression of nodes was comparatively higher in PKS15 than PKSS (Figure 4.26) (Table 4.14).

=		
Row Labels	PKS15	PKSS
acyltransferase	8.166106	5.665133
acyltransferase-like protein At1g54570	13.93166	3.861277
DNA-directed RNA Polymerase	2.1557	4.250608
E3 ubiquitin-protein ligase	1.996924	1.733923
expressed protein	2.99	8.110639
O-acyltransferase WSD1-like	16.10202	11.21269
Phospholipid diacylglycerol acyltransferase	6.800648	3.172053
S-acyltransferase	10.33211	4.528512
structural maintenance of chromosomes		
protein	21.86458	2.04762
Transcription factor	8.233841	7.673844

Table 4.14: Distribution of common nodes of acyltransferase

 subnetwork TPM values in PKS15 vs PKSS.

Moreover, the degree of freedom of PKS15 specific edges was much higher than PKSS, proving the relationship of higher expression of acyltransferase to the co-expression edges. However, in the case of the unique nodes of both the samples, the opposite was observed in which number of nodes and edges in PKSS were higher than PKS-15. Additionally, an independent cluster of acyltransferases was observed of major hubs, namely "O-acyltransferase", "BAHD acyltransferase", "S-acyltransferase", and "1-acyl-sn-glycerol-3-phosphate acyltransferase" in PKSS. These acyltransferases were mainly present in the end-product biosynthesis of terpeneglycoside moieties. On the other hand, PKS15 was observed to have only 24 nodes and 20 edges that showed a smaller number of hubs, pinpointing towards less no. of co-expressions (Figure 4.26) (Table 4.12).



Figure 4.26: Comparative co-expression networks with linked nodes and edges of common acyltransferases between PKS15 and PKSS transcriptomes.

Network comparison of PKS25 and PKSS presented 10 common nodes in both the samples; which were "mitogen-activated protein kinase", "Phospholipid diacylglycerol acyltransferase", "acyltransferase-like protein At1g54570" (*PES2*), "BEL1-like homeodomain protein", "S-acyltransferase", "DNA-directed RNA Polymerase", "serine threonine-protein kinase", "expressed protein", "O-acyltransferase WSD1-like", "acyltransferase-like", "O-acyltransferase", "methyltransferase", "Retrotransposon protein", "fructose-bisphosphate aldolase", and "Cycloartenol synthase". S-acyltransferase was major hub for PKS25, whereas O-acyltransferase was for PKSS network. Other smaller nodes such as "Fructose bisphosphate aldolase", "cycloartenol synthase", "MAP kinase", and "methyltransferase" were common with unique interactions (Figure 4.27) (Table 4.15).



Figure 4.27: Comparative co-expression networks of common acyltransferases with linked nodes and edges between PKS25 and PKSS transcriptomes.

Transcripts	PKS25	PKSS
mitogen-activated protein kinase	12.31908	5.292614
Phospholipid diacylglycerol acyltransferase	10.74189	3.172053
acyltransferase-like protein At1g54570	9.994469	3.861277
BEL1-like homeodomain protein	9.951031	3.003643
S-acyltransferase	7.947102	4.528512
acyltransferase	6.426799	5.665133
DNA-directed RNA Polymerase	6.256761	4.250608
serine threonine-protein kinase	6.000065	3.687288
expressed protein	5.762326	8.110639
O-acyltransferase WSD1-like	5.043076	11.21269
acyltransferase-like	4.825192	8.659518
O-acyltransferase	4.811779	3.913051
Galacturonosyltransferase	3.155811	1.002988
Transcription factor	3.053742	7.673844
methyltransferase	2.440912	7.52863
Retrotransposon protein	2.39492	5.554626
fructose-bisphosphate aldolase	2.01389	21.85238
Cycloartenol synthase	1.493032	5.94054

Table 4.15: Distribution of common nodes of acyltransferasesubnetwork TPM values in PKS25 vs PKSS.

Additionally, it was observed that PKS25 consists of a larger network (98 nodes and 87 edges) whereas PKSS showed smaller network (27 nodes and 23 edges) (Figure 4.21 and Figure 4.22) (Table 4.15). Common nodes mutually expressing in both PKS25 and PKSS showed similar differences noticed earlier in PKS15 v/s PKS25. In PKS-25, co-expression of callose synthase and MBOAT indicated the function of callose formation in the plasma membrane layer. On the other hand, this interacted with E3 ubiquitin-protein ligase that activates ubiquitin-conjugating enzyme indicating the presence of certain protein degradation occurring in PKSS. Other essential enzymes such as long-chain acyl CoA synthetase were observed only in PKSS. Common nodes expressing in PKSS, PKS15, and PKS25 showed differential interaction in their corresponding conditions, where "acyltransferase-like At1g54570 (*PES2*)", "callose synthase", "DNA-directed RNA Polymerase", "expressed protein", and "U-box domain-containing protein" were found expressed in all 3 transcriptomes (Figure 4.28) (Table 4.16).

Transcripts	PKS15	PKS25	PKSS
acyltransferase	8.166106	6.426799	5.665133
Acyltransferase-like			
protein	5.259263	5.982342	na
acyltransferase-like			
protein At1g54570	13.93166	9.994469	3.861277
protein ritige ie yo	1000100	,,,,,,,,,,	0.001277
callose synthase	3.238465	4.426359	na
diacylglycerol O-			
acyltransferase	3.416613	8.780628	na
DNA-directed RNA			
Polymerase	2.1557	6.256761	4.250608
1	2.00	5 7 (000)	0.110(20
expressed protein	2.99	5.762326	8.110639
F-box LRR-repeat	2.047002	175(100	
protein	3.84/903	4./36498	na
acyltransferase			
component of			
branched-chain			
alpha-keto acid			
denydrogenase	2 254002	7 4 4 5 9 0 6	
complex 1 1 1 0	3.354092	7.445896	na
membrane-bound O-			
acyltransferase	(0275(5	(205(54	
C24H6.01C-like	6.92/365	0.303034	na
(WSD1 Litra)	5 505902	16 22642	
(WSD1-like)	3.393803	10.32042	na
WSD1 like	16 10202	5 042076	11 21260
NSD1-IKC Dentetricementide	10.10202	5.045070	11.21209
remaincopeptide			
nrotain	1 806605	1 011018	20
Phoenholinid	1.890095	4.944010	lla
diacylalycerol			
acultransferase	6 800648	10 7/180	3 172053
	0.000040	10./4109	5.172055
S-acyltransferase	10.33211	7.947102	4.528512
Transcription factor	8.233841	3.053742	7.673844
U-box domain-			
containing protein	9.694073	4.197774	na

Table 4.16: Distribution of common nodes of acyltransferasesubnetwork TPM values in PKS15 vs PKS25 vs PKSS.

Overall, comparative analysis of shoot-derived networks was done based on individual pairs since all shoot samples showed different phenotypic characteristics where "callose synthase" and "U-box containing protein" were expressing highest in PKSS. In contrast, "acyltransferase-

like At1g54570" (*PES2*) expressed higher in PKS15 and PKS25, which mainly maintains integrity caused by a stress-related condition in chloroplast membrane by abiotic factors since PKSS shoots are grown in the natural field conditions with lower expression. In PKS25, the expression of a common node was found moderate except "DNA-directed RNA polymerase", highlighting the overall transcription activity possibly higher in this sample (Figure 4.28) (Table 4.13).



Figure 4.28: Comparative co-expression networks of common acyltransferases with linked nodes and edges among PKS15, PKS25 and PKSS transcriptomes.

4.2.3. Comparison of acyltransferase networks among different tissues of field grown plant samples (PKSS, PKSTS and PKSR)

Network comparison of PKSS and PKSTS showed differential expression and interactions of common nodes. The common nodes expressing in shoots and stolons were "acyltransferase-like protein At1g54570" (*PES2*), "autophagy protein", "BAHD acyltransferase", "expressed protein", "heat shock 70 kDa protein", "Hydrolyzes glycerol-phospholipids at the terminal phosphodiesteric bond" (By similarity), "O-acyltransferase WSD1-like", "Phospholipid diacylglycerol acyltransferase", "Retrotransposon protein", and "S-acyltransferase" (Figure 4.29) (Table 4.17).



Figure 4.29: Comparative co-expression networks of common acyltransferases and linked nodes and edges between PKSS and PKSTS transcriptomes.

Transcripts	PKSS	PKSTS
acyltransferase	5.665133	5.336695
acyltransferase-like	8.659518	38.6021
acyltransferase-like protein At1g54570	3.861277	3.301555
autophagy protein	4.346463	5.049633
BAHD acyltransferase	7.57448	14.86969
expressed protein	8.110639	20.65944
Heat shock 70 kDa protein	8.071508	9.199704
Hydrolyzes glycerol-phospholipids at the terminal phosphodiesteric bond (By		
similarity)	2.396875	0.856055
O-acyltransferase WSD1-like	11.21269	9.697264
Phospholipid diacylglycerol acyltransferase	3.172053	4.60297
Retrotransposon protein	5.554626	2.007269
S-acyltransferase	4.528512	9.957728
S-acyltransferase At2g14255-like	7.922346	2.449905
transcription factor	7.673844	6.292024

Table 4.17: Distribution of common nodes of acyltransferase subnetwork TPM values in PKSS vs

 PKSTS

It was observed that S-acyltransferase comparatively expressed less in PKSS but still formed more no. of co-expressing interactions. In addition, heat shock protein and transposon protein were also having different interactions and expression in samples resulting in exclusive interaction for both PKSS and PKSTS. On the other hand, 9 nodes and 6 edges were observed only in PKSTS and the presence of "MBOAT" as a major hub. In the case of PKSS, connecting links were observed between "O-acyltransferase" and "long-chain-alcohol-O-fatty acyltransferase" via "fructose bisphosphate aldolase" moiety (Figure 4.29) (Table 4.14). Analysis of PKSTS and PKSR acyltransferase networks resulted in identification of common nodes mutually expressing differently in both the samples; which were "26S proteasome non-ATPase regulatory subunit", "AAA-type ATPase family protein", "Acyltransferase", "Acyltransferase-like protein", "acyltransferase-like protein At1g54570" (PES2), "aspartic proteinase", "BAHD acyltransferase", "cytochrome P450", "DEAD-box ATP-dependent RNA helicase", "expressed protein", "Hydrolyzes glycerol-phospholipids at the terminal phosphodiesteric bond" (By similarity), "Lectin-domain containing receptor kinase", "LRR receptor-like serine threonine-protein kinase", "membrane-bound O-acyltransferase C24H6.01c-like", "O-acyltransferase WSD1-like", "Pentatricopeptide repeat-containing protein", "phosphatase 2C", "Phospholipid diacylglycerol acyltransferase", "receptor-like protein kinase", "ribonuclease P family protein", and "S-acyltransferase", factor" (Figure 4.30) (Table 4.15).



Figure 4.30: Comparative co-expression networks of common acyltransferases with linked nodes and edges between PKSTS and PKSR transcriptomes

Major hubs for both the samples were "S-acyltransferase", "BAHD acyltransferase" forming different co-expression interactions. Further, some connecting nodes such as "pentatricopeptide repeat-containing protein", "aspartic proteinase", "receptor-like protein kinase", "AAA-type ATPase family protein", "DEAD-box ATP dependent RNA helicase", "lectin domain receptor kinase" were present in both the samples with differences in expression. Furthermore, despite of higher expression of CYP 450, interactions with other nodes were absent in PKSTS, but in PKSR, the same node was found to connect all the 3 major hubs via similar co-expression profile. The "acyltransferase-like protein At1g54570" (*PES2*) was exclusive hub for PKSTS and majorly formed interactions and higher expression (Figure 4.30) (Table 4.18).

Tuble 1.10. Distribution of common nodes of acylitansienase suchetwork 1114 value		STRETD.
Transcripts	PKSR	PKSTS
26S proteasome non-atpase regulatory subunit	3.715405	8.130582
AAA-type ATPase family protein	2.561353	2.931542
Acyltransferase	3.980942	5.336695
Acyltransferase-like protein	2.687091	5.92144
acyltransferase-like protein At1g54570	3.065248	3.301555
aspartic proteinase	5.902519	2.109266
BAHD acyltransferase	7.665313	14.86969
cytochrome P450	3.487211	23.67505
DEAD-box ATP-dependent RNA helicase	7.735477	6.651131
expressed protein	13.28977	20.65944
Hydrolyzes glycerol-phospholipids at the terminal phosphodiesteric bond (By		
similarity)	3.581558	0.856055
Lectin-domain containing receptor kinase	5.173803	3.729279
LRR receptor-like serine threonine-protein kinase	1.271733	4.292554
membrane-bound O-acyltransferase C24H6.01c-like	2.296254	3.241296
O-acyltransferase WSD1-like	8.950824	9.697264
Pentatricopeptide repeat-containing protein	2.746941	1.95773
phosphatase 2C	4.818019	10.32369
Phospholipid diacylglycerol acyltransferase	5.042476	4.60297
receptor-like protein kinase	1.723092	4.735057
ribonuclease P family protein	3.71623	4.650993
S-acyltransferase	3.611732	9.957728
transcription factor	8.293427	6.292024

Table 4.18: Distribution of common nodes of acyltransferase subnetwork TPM values in PKSS vs PKSTS.

Shoot, root, and stolon acyltransferase networks were compared, resulting in the identification of components in all of them with a difference of expression profile. These were namely "Acyltransferase-like protein At1g54570" (*PES2*), "BAHD acyltransferase", "expressed protein", "Hydrolyzes glycerol-phospholipids at the terminal phosphodiesteric bond (By similarity)", "O-acyltransferase WSD1-like", "Phospholipid diacylglycerol acyltransferase", and "S-acyltransferase" (Figure 4.31) (Table 4.19).



Figure 4.31: Comparative co-expression networks of common acyltransferases with linked nodes and edges among PKSS, PKSTS and PKSR transcriptomes.

Table 4.19: Distribution of common nodes of acyltransferase subnetwork TPM va	lues in PKSS vs PKSTS vs
PKSR .	

Transcripts	PKSR	PKSS	PKSTS
Acyltransferase	3.980942	5.665133	5.336695
acyltransferase-like protein At1g54570	3.065248	3.861277	3.301555
BAHD acyltransferase	7.665313	7.57448	14.86969
expressed protein	13.28977	8.110639	20.65944
Hydrolyzes glycerol-phospholipids at the terminal phosphodiesteric bond			
(By similarity)	3.581558	2.396875	0.856055
O-acyltransferase WSD1-like	8.950824	11.21269	9.697264
Phospholipid diacylglycerol acyltransferase	5.042476	3.172053	4.60297
S-acyltransferase	3.611732	4.528512	9.957728
transcription factor	8.293427	7.673844	6.292024

4.2.4. BAHD acyltransferase subnetwork capturing co-expressing modules related to iridoid glycoside biosynthesis in shoots, stolons and roots

The BAHD acyltransferase-specific subnetwork was only found in the field-grown samples. Further analysis of connecting nodes highlighted the involvement of unique edges for every specific sample. In PKSR, "HSF-type DNA-binding", "serine carboxypeptidase-like", "glycogen synthase", "Isoleucyl-tRNA synthetase", "OPT oligopeptide transporter protein", "cytochrome P450", "SNARE associated Golgi protein", "element-binding protein", "E3 ubiquitin-protein ligase", "Phytochrome A-associated F-box" and "ribonuclease P family protein" (Figure 4.31) were observed forming edges with BAHD-ATs. In PKSS, "resistance protein", "STYKc", "ribosomal protein S1", "serine threonine-protein kinase", "DUF1771", "Molybdopterin guanine dinucleotide synthesis protein B", "autophagy protein", "smr domaincontaining protein", "BEL1-like homeodomain protein", "WD repeat-containing protein", "ENTH domain" were observed as corresponding edges (Figure 4.31). CBL-interacting serine threonine-protein kinase is a stress tolerance protein and involved in shoot biomass development (Suzuki et al., 2007). In PKSTS, "BRASSINAZOLE-RESISTANT 1-like", "Transcription factor" (bZIP and bHLH), "aspartic proteinase", "F-box protein", "Retrotransposon protein", "polyphenol oxidase", "cellulose synthase-like protein, and histone acetyltransferase", "5' -nucleotidase SurE-like", "DUF4206" were detected (Figure 4.31). Coexpression linkage with polyphenol oxidase (PPO) is strikingly important for phenylpropanoid pathway including biosynthesis of acyl group containing acids such as cinnamic acid and 4coumaric acid that are major functional groups in last step modifications of iridoid glycosides. The differences in the edge formation in every sample highlighted the involvement of specific functional changes with expression among different tissues. BAHD-ATs expression and interaction in different samples indicated their presence in last step modification of iridoid glycosides. Therefore, sequences of BAHD-ATs were further investigated with multiple sequence alignment approach. Four transcripts of BAHD were generated using Multiple Sequence Alignment by UPGMA approach (Figure 4.32).





other iridoid biosynthesis

Figure 4.32: Phylogenetic tree analysis for filtering potential BAHD acyltransferase from the transcriptomes for molecular modelling.

Transcripts with larger length were shortlisted from each clad of the tree. This resulted in total of 6 transcript sequences used for further modelling and molecular docking against acyl-group ligands to check the specificity for iridoid glycoside compounds.

4.2.5. Modelling and molecular docking of BAHD-acyltransferases to identify specific intermediates for last step modification of iridoid glycosides

Overall, a total of 13 major compounds belonging to iridoid glycoside class are reported in P. kurroa. These are Picroside-I, Picroside-II, Picroside-III, Picroside-IV Picroside-V, kutkoside, pikuroside, 6-ferulloyl catapol, vernicoside, minecoside, verminoside, specioside and 6vanniloyl catalpol. Among them, the presence of common catalpol structure with different acyl groups forming esterification bond with one of its oxygen atoms has been noticed. In our study, Picroside-II, kutkoside, pikuroside and 6-vanniloyl catalpol have been found to have vanilloyl functional group whereas in picroside-III and 6-ferruloyl catalpol, the coniferyl moiety has been observed. P-coumaroyl functional group was specific for Picroside-V and specioside. Moreover, Picroside-I, Picroside- V, vernicoside and minecoside, presence of cinnamoyl, methyoxybenzoyl, benzoyl and 3-hydroxy-4-methoxycinnamoyl functional groups discussed in Chapter 3. Such additions are mainly performed by various Acyl-CoAs, which donate their acyl groups to catalpol.Six transcript sequences, which have been shortlisted through comparative analyses as detailed above were named as SS_3469, STS_4084, STS_4241, STS_8424, SR_4494, and SR_4510 (Table 4.17), where alphabet code represents identity to a particular sample (PKSS, PKSTS, and PKSR) and number denotes transcript_id in our dataset. In docking results of BAHD-ATs (receptors) and acyl donors (ligands) most of the interactions were of H-bonds and hydrophobic interactions. The binding affinity was observed in the good and excellent binding affinity ranges. Thus, the compounds with binding affinity less than -7.5 have been considered as potential counterparts. The results showed that the cinnamoyl, vanilloyl, feruloyl and benzoyl moieties containing acyl donors had excellent binding affinities for their respective BAHD-ATs specific to different tissue samples of *P. kurroa*. The interactions with their respective BAHD-ATs also pinpointed towards their biosynthetic sites.

Protein ID	Modelled structure	Acyl donor compounds	Binding
			affinity
			(kcal/mol)
SS_3469	E.	Cinnamoyl_CoA	-9.1
(PKSS)		p_coumaroyl_COA	-9
		6-ferulloylcatalpol	-8.8
	CON STREET	vanilloyl_CoA	-8.6
		ferruoyl_COA	-8.4
		6-vanilloylcatalpol	-8.4
		6-cinnamoyl catalpol	-8
STS_4084		6-cinnamoyl catalpol	-9.7
(PKSTS)		p_coumaroyl_COA	-8.7
	Maps.	6-vanilloylcatalpol	-8.7
		benzoyl_COA	-8.3
		ferruoyl_COA	-8.3
		Cinnamoyl_CoA	-8.2
	ALL STORES	6-ferulloylcatalpol	-8
STS_4241		p_coumaroyl_COA	-8.8
(PKSTS)	S FOR BY S A	Cinnamoyl_CoA	-8.7
		6-cinnamoyl catalpol	-8.6
		caffeoyl_COA	-8.5
		6-ferulloylcatalpol	-8.5
		6-vanilloylcatalpol	-8.3
	Se cere	3_hydroxy_4_methoxycinnamoyl_COA	-8.2
	<u> </u>	benzoyl_COA	-8.2
		vanilloyl_CoA	-8.2
STS 8424		6-cinnamoyl catalnol	
(PKSTS)		Cinnamovl CoA	-8.1
		6-ferullovlcatalpol	_7 9
		6-vanillovlcatalpol	-7.7
		Catalnol	-7.5
			1.5
SR 4494		3 hydroxy 4 methoxycinnamoyl COA	-10.2
(PKSR)	And the second	benzoyl_COA	-9.5
		6-cinnamoyl catalpol	-9.2
	A A A A A A A A A A A A A A A A A A A	ferruoyl COA	-9.1
		6-ferulloylcatalpol	-8.9
	A Start Start	caffeoyl COA	-8.8
		Cinnamoyl CoA	-8.8
		vanilloyl CoA	-8.7
		p_coumaroyl_COA	-8.3

Table 4.20: Outcomes of docking between BAHD-Acyltransferase Enzymes and their corresponding acyl-group donating compounds.

		6-vanilloylcatalpol	-8.3
SR_4510	2 ~	6-cinnamoyl catalpol	-8.9
(PKSR)		Cinnamoyl_CoA	-8.9
		caffeoyl_COA	-8.8
	6-ferulloylcatalpol	-8.8	
		6-vanilloylcatalpol	-8.8
	THE BAR	p_coumaroyl_COA	-8.6
		Catalpol	-8.2
	\sim	vanilloyl_CoA	-8

4.3. Identification and mapping of single nucleotide polymorphisms (SNPs) to the global co-expression networks

4.3.1. Reads generation and De novo sequence assembly of GBS data

For each population, the results of GBS sequencing were raw reads, which were then further cleaned by the radtags procedure. 161834 fragments of assembled sequences were catalogued by the cstacks after the cleaned readings were processed through the stacks pipeline. Based on HPLC quantifications of P-I and P-II in shoots, stolons, and roots, the P. kurroa populations were split into high and low populations. From the outcomes P-I percent in shoots, P-I percent +P-II percent in stolons, and P-II percent in roots, the GBS dataset was divided into 3 types of populations. Populations with concentrations greater than 1.6 % was considered as high whereas those with less than 0.6 % were taken as low content population set for P-I % in shoots. Overall, 14 populations were considered as high P-I population and 11 were considered as low P-I populations. Population-wise SNP analysis resulted in the identification of 26,7186 SNPs in overall fragments. Fragments unique for only high and low populations were filtered from the total catalogue resulting in 3,483 fragments for high P-I% in shoots (HPF-S) and 4,677 for low P-I% population (LPF-S). Similarly, based on total P-I + PII% in stolons, GBS datasets were divided into high and low populations. 13 populations with greater than 1.75 PI+PII % content was considered as high and 16 populations with less than 1.0 PI+PII % were considered as low content. Fragments unique to only high and low populations were filtered from the total catalogue resulting in 6,102 fragments for high PI+PII % in stolons (HPF-ST) and 1,878 for low PI+PII % population (LPF-ST). Likewise, the dataset was divided into low and high populations for PII% content in roots. 16 populations with greater than 0.16 PII% were considered as high and 24 population with less than 0.15 PII% were considered as low. Fragments unique to only high and low populations were filtered from the total catalogue resulting in 3,141 fragments for high PII % in root (HPF-R) and 2298 for low PII % population (LPF-R). These sets of fragments based on differential picroside contents in different tissues were considered for mapping against transcriptomes followed by co-expression network analysis.

4.3.2. Assembly of transcriptomes data

Illumina paired-end sequencing resulted in generation of 29903780, 26053444, 37034218, 41298148, 20413038, 20064888, 27790024, and 31291454 raw reads from RNA-seq analysis of of PKS1, PKST3, PKS4, PKS5, PKST5, PKST16, PKST18, and PKS21 respectively (Table S1). Number of trimmed and cleaned raw reads that qualified for further processing were 28912390, 25021970, 35896124, 39998026, 19532924, 19182054, 26739904, and 30119848, for PKS1, PKST3, PKS4, PKS5, PKST5, PKST16, PKST18, and PKS21, respectively. The number of final assembled transcripts for PKS1, PKST3, PKS4, PKS5, PKST5, PKST6, PKST3, PKS4, PKS5, PKST16, PKST18, and PKS21, respectively. The number of final assembled transcripts for PKS1, PKST3, 199520, 121842, and 241041, respectively. (Table 4.21)

	PKS1	PKST3	PKS4	PKS5	PKST5	PKST16	PKST18	PKS21
Raw Reads	29903780	26053444	37034218	41298148	20413038	20064888	27790024	31291454
Cleaned Raw Read	28912390	25021970	35896124	39998026	19532924	19182054	26739904	30119848
Assembled	140777	152527	200074	148973	209785	199520	121842	241041
Annotated transcripts	36317	47066	48985	45502	44763	49024	43531	52698
Gene Ontology annotated	17118	21402	21897	23299	18157	22051	23176	22384
KO annotated	14769	18080	18483	19407	15582	18376	19166	18758
Uniprot	5625	6762	7135	7546	5851	6916	7252	7238
Global Network Nodes	6290	7411	6677	3823	6874	6989	6239	7004
Global Network Edges	65944	69570	22404	58312	63023	61946	60990	64046
Mapped Fragments	802	900	1022	1213	729	889	1106	1026
High Population	361	389	444	513	357	368	476	428
Low Population	388	456	492	613	309	456	548	522
SNP network nodes	1389	2280	918	1870	1766	1880	1991	2170
SNP network edges	2075	4299	926	3475	3002	3244	3832	4092

Table 4.21: Distribution of annotation, networks, and mapped SNP fragments on 8 transcriptome samples
4.3.3. Functional annotation of *Picrorhiza kurroa* transcriptomes

Transcriptome assemblies were annotated with PLAST and were computed for orthology predictions at different taxonomic levels using eggNOG [129] resource database. Groups of orthologs were thereafter functionally annotated from recently updated GO [130] and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway databases [131]. This resulted in the actual annotation of 36317, 47066, 48985, 45502, 44763, 49024, 43531 and 52698 transcripts of PKS1, PKST3, PKS4, PKS5, PKST5, PKST16, PKST18, and PKS21, respectively. 5625, 6762, 7135, 7546, 5851, 6916, 7252, and 7238, unigenes from PKS1, PKST3, PKS4, PKS5, PKST5, PKST5, PKST16, PKST18, and PKS21, respectively were extracted from hits in UniProt database. Annotation by GO resulted in 17118, 21402, 21897, 23299, 18157, 22051, 23176, and 22384, transcripts of assembled sequences from PKS1, PKST3, PKS4, PKS5, PKST16, PKST18, and PKS21, respectively. In addition, annotation using the KEGG database identified 14769, 18080, 18483, 19407, 15582, 18376, 19166, and 18758, transcripts, respectively, from PKS1, PKST3, PKS4, PKS5, PKST5, PKST5, PKST16, PKST18, and PKS21, RS54, PKS5, PKST5, PKST16, PKST18, and PKS21, CTable 4.18).

4.3.4. Generation of gene co-expression networks

Annotated transcripts with transcripts per million reads (TPM) values greater than zero were although a smaller fraction of the total transcriptome dataset but still covered large entities in the form of a global network representing the overall system. The number of nodes in the 8 network modules were 6290, 4259, 6677, 3823, 6874, 6989, 6239, and 7004 from PKS1, PKST3, PKS4, PKS5, PKST5, PKST16, PKST18, and PKS21, respectively. Furthermore, the number of edges from these networks were 65944, 71007, 22404, 58312, 63420, 62353, 61376, and 64521 for PKS1, PKST3, PKS4, PKS5, PKST5, PKST

4.3.5. Mapping GBS fragments of different populations on transcriptome dataset

Fragments containing SNPs were mapped to transcriptome datasets covering modules of transcripts among transcriptomes. Overall, 802, 900, 1022, 1213, 729, 889, 1106, and 1026 fragments for P-I % were mapped against PKS1, PKST3, PKS4, PKS5, PKST5, PKST16, PKST18, and PKS21 transcriptomes, respectively. Out of these 361, 389, 444, 513, 357, 368, 476, and 428 were HPF-S and 388, 456, 492, 613, 309, 456, 548, and 522 were LPF-S whereas the rest were mapped to both HPF-S and LPF-S.

Similarly, 1012, 1208, 1295, 1471, 881, 889, 1174, 1423 and 1287 transcripts were mapped to the unique fragment library based on P-I +PII % content in stolons for PKS1, PKST3, PKS4, PKS5, PKST5, PKST16, PKST18, and PKS21 transcriptomes, respectively. Out of these 826, 1011, 1085, 1221, 988, 1204, and 1059 were HPF-ST and 79, 104, 104, 131, 84, 98, 117, and 120 were LPF-ST. Likewise, for PII% in root specific population 479, 505, 635, 1470, 437, 525, 629, and 1101 transcripts of PKS1, PKST3, PKS4, PKS5, PKST5, PKST16, PKST18, and PKS21 transcriptome samples were mapped respectively. Out of these 274, 260, 356, 1222, 155, 280, 343, and 518 were HPF-R and 169, 183, 240, 132, 246, 209, 243 and 369 were LPF-R whereas, rest of the transcripts mapped to both. These mapped fragments were taken as bait to extract interacting edges from the global co-expression network (Figure 4.33).



Figure 4.33: Distribution of SNPs based on the population and the type of category. (A): Distribution of SNPs among Low Population fragments. (B): Distribution of SNPs among High Population Fragments. (C): Distribution of HPF(Red) and LPF(Green) containing transcripts in global individual gene co-expression networks. (D): Subnetworks of transcripts containing HPF or LPF and their interacting.

4.3.6. Gene co-expression network analysis of picroside % population in different tissues.

The gene co-expression network was extracted from global co-expression network using transcripts matching with SNP encompassing fragments of high/ low populations. These subnetworks represented as nodes having SNPs specific for high or low Picroside-I in shoots were relatively smaller in number of nodes in the range of 918-2280 nodes and 926-4299 edges. Similarly, subnetwork consisting of nodes based on SNPs specific for Picroside content in stolons were in the range of 1272-2747 nodes and 1498-5752 edges. Likewise, Sub-networks specific for SNPs based on P-II % content in roots were in the range of 518-2372 nodes and 518-4737 edges. Among extracted networks, the nodes were divided based on the type of population and the type of SNPs namely transitions and transversions.

4.3.7. Gene Ontology analysis of interacting nodes in the networks

The functions mostly related to secondary metabolites biosynthesises were used to functionally annotate the interactive edges of SNPs containing nodes of each transcriptome sample. These functional modules were extracted using keywords such as vacuoles, transporter activity, transferase, isoprenoid, secondary metabolite biosynthesis, defence response, transcription factor, root, starch, protein kinase, carbon fixation, chloroplast, photosynthesis, shoot development, and signal transduction as most of these GO terms are directly or indirectly related to secondary metabolites biosynthesis and storage in plants. Overall average of 1655 transcripts were covered for transcriptome sample. Out of these, 839 transcripts were covered in PKS1 that was least out of all whereas, 2212 transcripts were covered in PKST3 that was the highest among all the transcriptome samples. Among the function, the transcripts with function related to carbon fixation were covered least by the interactive network whereas chloroplast was found to be covered by the highest number of transcripts. (Figure 4.34)



Figure 4.34: Distribution of Gene Ontology modules interacting with SNPs containing transcripts. (A): Gene Co-expression network of transcripts with specific GO functions interacting with SNP containing nodes. (B): Distribution of transcripts mapping to specific GO functions.

4.3.8. Identification of hubs with SNPs based on P-I % in shoots

Hubs from each co-expression network were extracted individually based on the categories of high/low population SNPs and the occurrence of transversion or transition type of SNPs. Each individual sample showed a wide variety of transcripts representing various putative functions covering interactive nodes of different GO annotations mentioned above. In context of Picroside-I concentration, PKS1 and PKS4 were lying in the range of low concentrations therefore significance of hubs comprising LPF is of high importance. Similarly, PKS5, PKST5 and PKST16 were populations of high Picroside-I concentrations therefore hubs comprising HPF are high prevalence. However, PKST3, PKST18 and PKS21 were lying in the moderate range of P-I concentration therefore, importance of hubs with both HPF-S and LPF-S should also be considered. (Table 4.22)

Populatio n ID	PI % in shoot	Population	Total Number of Hubs	Hubs comprising HPF	Hubs comprising LPF	Hubs comprising both HPF and LPF
PKS1	0.53%	Low	12	6	6	0
PKST3	1.28%	Moderate	17	6	7	4
PKS4	0.13%	Low	15	6	6	3
PKS5	1.68%	High	14	4	7	3
PKST5	1.68%	High	18	8	7	3
PKST16	1.72%	High	15	5	10	0
PKST18	1.59%	High	16	6	8	2
PKS21	0.71%	Moderate	18	9	9	0

Table 4.22: Distribution of hubs specific for HPF-S and LPF-S identified for 8 transcriptome samples.

4.3.9. Network analysis of unique hubs for P-I% in shoots corresponding to different transcriptome samples

Hubs unique to each set of a transcriptome based on the presence of populations were identified thus resulting in the probable SNPs containing transcripts corresponding to P-I concentrations. In case of low P-I% population, 8 hubs were identified uniquely having LPF-S which were namely "protein transport protein SEC61 subunit", "actin filament-based movement", "4-coumarate--CoA ligase-like", "domain protein", "aldehyde oxidase", "protein phosphatase 2C

55-like", "pyrophosphate-energized vacuolar membrane proton" and "leucine aminopeptidase". On the other hand, 9 hubs namely "NADP-dependent alkenal double bond reductase", "Transmembrane emp24 domain-containing protein", "arginine Nmethyltransferase", "Helix-loop-helix DNA-binding domain", "Ankyrin repeat domain 24", "30S ribosomal protein", "Cold acclimation protein", "Uroporphyrinogen decarboxylase" and "DJ-1/PfpI family" were found specifically in high picroside population transcriptome. (Table 4.19)

4.3.10. Network analysis of common hubs of P-I% population

Shortlisted SNPs containing nodes were categorized as transition and transversion based on the presence of SNPs. Furthermore, another criteria of population was also consistent with the network where nodes of high and low population were considered. Nodes with transversion type of SNPs were identified from each transcriptome, followed by sorting based on occurrence and interaction among different transcriptome samples. Total 50 and 61 hubs were identified for HPF and LPF, respectively, Whereas 16 hubs were identified for both high and low populations. Among these, 10 hubs were found to be present for at least 3 transcriptomes. Some of the important hubs specific for low population SNPs were "AdoMet-dependent rRNA methyltransferase", "Serine theonine protein kinase", "auxin responsive protein", "3-deoxy-Darabino-heptulosonate 7-phosphate", "actin filament-based movement", "UDP-Glycosyltransferase", and "leucine aminopeptidase". On the other hand, hubs such as "clavaminate synthase-like protein", "fructose-bisphosphate aldolase", "Pyruvate decarboxylase", "2-oxoglutarate malate translocator", "5-hydroxyisourate hydrolase", "arginine N-methyltransferase" were specific only for high population SNPs. Some of the important hubs present for both the high and low population were "Nuclear matrix constituent protein", "4-coumarate-CoA ligase" (Table 4.22)

4.3.11. Identification of hubs with SNPs based on P-I+P-II % content in stolons

Based on the mapped fragments for HPF-ST and LPF-ST fragments, hubs were identified. Based on the differential P-I+P-II content in stolons, these hubs were further filtered. Among the eight transcriptomes, PKST3, PKS4 and PKS21 were thought to have a high picroside content population in stolons, so the presence of hubs containing HPF-ST was given more attention in the analysis. Similarly, hubs with LPF-ST were taken into consideration as PKS5, PKST5, and PKST16 were in the lower range of PI+PII content. Like PKS1, PKST18 has hubs with LPF-ST and HPF-ST concentrated in their somewhat sized populations. (Table 4.20)

4.3.12. Network analysis of Unique hubs for P-I+P-II % in stolons corresponding to different transcriptome samples

As a result of the identification of hubs specific to each set of a transcriptome based on the existence of populations, stolons with P-I+PII contents in their transcripts were likely SNP-containing. The E3 ubiquitin-protein ligase, CBL-interacting serine threonine protein kinase, Zinc finger protein CONSTANS-LIKE, glutathione, homeobox-leucine zipper protein, and phosphoserine aminotransferase were identified as 7 hubs specifically having LPF-ST in the case of low P-I+PII content population. However, six hubs were found to be specifically high in the picroside population transcriptome, including "E3 ubiquitin-protein ligase," "glutathione," "kinesin-like calmodulin-binding protein," "serine threonine-protein phosphatase 2A 65 kDa regulatory subunit A," "cell division cycle protein 48," and "peptide nitrate transporter At1g22540-like". (Table 4.23)

Table 4.23: Distribution of hubs specific for HPF-ST and LPF-ST identified for 8 transcriptome samples.

Populatio n ID	PI + PII % in stolon	Population	Total Number of Hubs	Hubs comprising HPF	Hubs comprising LPF	Hubs comprising both HPF and LPF
PKS1	1.15%	Moderate	13	5	5	3
PKST3	2.35%	High	19	7	7	5
PKS4	2.11%	High	14	7	4	3
PKS5	0.84%	Low	19	7	6	6
PKST5	0.84%	Low	20	7	8	5
PKST16	0.65%	Low	20	8	7	5
PKST18	1.26%	Moderate	16	6	5	5
PKS21	1.82%	High	15	6	5	4

4.3.13. Network analysis of common hubs based on P-I+P-II % in stolons

Shortlisted SNP-containing nodes were classified as transition or transversion based on the presence of SNPs. Another population criterion that took into account the nodes with high and low populations was compatible with the network. Each transcriptome's nodes containing transversion-type SNPs were discovered, and then the nodes were sorted based on how frequently they appeared and how they interacted with other samples. For HPF-R and LPF-R,

a total of 50 and 47 hubs, respectively, were found. In contrast, 23 hubs with both high and low populations were found. 19 hubs were discovered to be present for at least 3 transcriptomes. Some of the important hubs specific for low population SNPs were "ubiquitin carboxyl-terminal hydrolase", "Pfam:IstB", "one of the two reaction center proteins of photosystem II", "S-acyltransferase", "Histone H1", "WRKY transcription factor", "ABC transporter B family member", "siroheme" and "tubby-like F-box protein". On the other hand, hubs such as "UDP-Glycosyltransferase", "E3 ubiquitin-protein ligase", "CBL-interacting serine threonine-protein kinase", "Zinc finger protein CONSTANS-LIKE", "glutathione", "homeobox-leucine zipper protein", "phosphoserine aminotransferase", and "Protein of unknown function (DUF1644)" were specific only for high population SNPs. (Table 4.20)

4.3.14. Identification of hubs with SNPs based on P-II % in roots

The mapped fragments associated to HPF-R and LPF-R fragments were used to identify hubs. Based on the occurrence of distinct PII content in roots in the transcriptomes, these hubs were further filtered. PKS5, PKST5, and PKS18 were three of the eight transcriptomes that were thought to have a high picroside content population for roots, hence the investigation was more heavily focused on hubs that included HPF-R. Similarly, hubs with LPF-R were considered since PKS1, PKST4, and PKST16 were in the lower range of PII content. Similarly, PKST3, PKS21 also had a moderate population, and its hubs with LPF-R and HPF-R were concentrated there. (Table 4.21)

4.3.15. Network analysis of unique hubs for P-II % in roots corresponding to different transcriptome samples

Hubs unique to each set of a transcriptome based on the presence of populations were identified and resulted in probable SNPs encompassing transcripts corresponding to PII% content in roots. In case of low PII % populations, 8 hubs were identified uniquely having LPF-R and these were namely "Ethylene-overproduction protein", "HEAT repeat", "PfkB-type carbohydrate kinase family protein", "receptor-like protein kinase", "Rhamnose biosynthetic enzyme", "sucrose transporter", "TBCC domain-containing protein 1-like", and "transmembrane 9 superfamily member". On the other hand, 17 hubs namely "5-hydroxyisourate hydrolase", "5methyltetrahydropteroyltriglutamate—homocysteine", "BTB POZ domain-containing protein", "calcium-dependent protein kinase", "Cysteine-rich receptor-like protein kinase", "DJ-1/PfpI family", "Domain of Unknown Function (DUF1086)", "galactinol synthase", "Methionyl-tRNA", "phosphoribosylformylglycinamidine synthase", "Splicing factor 3B subunit", "splicing factor U2af large subunit", "synthetase", "translational activator", "U-box domain-containing protein", and "zinc finger CCCH domain-containing protein" were found specifically high picroside population transcriptome. (Table 4.24)

Table 4.24: Distribution of hubs specific for HPF-R and LPF-R identified for 8 transcriptome samples.

Population ID	PII % in root	Population	Total Number of Hubs	Hubs comprising HPF	Hubs comprising LPF	Hubs comprising both HPF and LPF
PKS1	0.04%	Low	20	9	6	5
PKST3	0.23%	Moderate	24	8	6	10
PKS4	0.08%	Low	14	7	7	3
PKS5	0.5%	High	32	16	8	8
PKST5	0.5%	High	23	9	8	6
PKST16	0.04%	Low	26	11	8	7
PKST18	0.35%	High	24	10	6	8
PKS21	0.02%	Moderate	25	11	8	6

CHAPTER 5 DISCUSSION Initially the study was sought to design strategies to build gene co-expression networks explaining overall global system in *Picrorhiza kurroa*. Furthermore, it included identification of key modules of different classes of genes contributing to biosynthesis and accumulation of iridoid glycosides. This lead to designing of study into three objectives that were achieved sequentially. First, objective was to build co-expression networks using NGS transcriptome datasets of different organs/tissues and developmental stages. Second objective included extraction and prioritization of co-expression modules related to the secondary metabolites biosynthesis. Last objective of identification and mapping single nucleotide polymorphism (SNPs) to global co-expression networks were enrichments of the findings of previous two objectives. These objectives are discussed briefly in following sections.

5.1. Co-expression networks of different organs/tissues and developmental stages of *Picrorhiza kurroa*.

The generated gene co-expression networks showed differences in interaction and degree of freedoms with change in the transcriptome sample. In transcriptomes created under laboratory conditions, it is simple to identify factors that influence how plants express their genes, but in the field, the dynamics of the transcriptome are more complicated and controlled by endogenous nocturnal rhythms, ambient temperature, the age of plants, and solar radiation [142]. The significance of a node (gene) in the network as a key hub gene increases if it participates in more number of pathways. TopGO, an R-bioconductor utility, was used to retrieve GO annotations for enrichment studies. Co-expressed gene sub-networks were created from the global co-expressed gene network modules that had undergone GO enrichment. These modules were simple to evaluate compared to the main network modules because of their greatly reduced complexity. Additionally, it may be possible to determine if a co-expressed gene sub-network that is specific to a given organ can help in the manufacture of a particular class of terpenoids, such as phytosterols or monoterpenoid iridoid glycosides. Different, organ-specific phytohormonal signaling networks were identified that may be in charge of the enhanced accumulation of P-I or P-II. Furthermore, these global sub-networks correctly anticipated the transfer of isoprene units between organs as well as their storage and decomposition.

Additionally, different connections in shoot-only systems' terpenoid glycoside sub-networks that lack a specific organ for storing picrosides were observed.

Based on the comparison of shoot-only systems of PKS15, PKS25, and PKSS various key components were noted and reported in Chapter 4. Some of the key observations revealed their importance in the system are discussed briefly as followed. In comparison to PKS25, PKS15 displayed noticeably altered leaf shape, higher shoot biomass growth, a 10-fold increase in leaf size, and an increase of 1.8-fold in the weight of a single shoot [34]. P-I and P-II have both been detected, but only P-I was found in PKS15 in substantial amount (0.6%) and PKS25 in minimal amounts (0.01%) [72]. A set of TFs and the auxin response factor are the majors hubs discovered as a result of an intense swarming of linked interactions. It has been discovered that F-box kelch-repeat proteins govern the turnover of phenylalanine ammonia-lyase to control phenylpropanoid production [143]. Cytochrome P450s (CYPs) are the primary enzymes that catalyse the decoration of the basic terpenoid skeletons in downstream terpene production pathways [144]. The common node-wise interactions between PKS15 and PKS25 have highlighted auxin responsive interactions (Auxin response factor, Transport inhibitor response proteins, and Scarecrow-like proteins), the activation of mRNA, protein, and genomic DNA degradation pathways (E3 SUMO-protein ligase, Enhancer of mRNA-decapping protein), Staphylococcal nuclease domain-containing protein (SND) and some secondary metabolitespecific interactions (F-box kelch-repeat proteins, CYP). Comparison of averaged differential gene expression of nodes present in the common network interactions between PKS15 and PKS25 revealed 8.5-fold higher expression of Endoglucanase in PKS15 compared to in PKS25, while expression of Cellulose synthase-like genes was 12.2 fold higher in PKS25 compared to PKS15. Sequence similarities between cellulose synthase-like genes and cellulose synthase genes lead to the hypothesis that they could produce additional plant cell wall polymers as glycosyl transferases [145]. The hydrolysis of lignocellulosic biomass, on the other hand, requires the glycosyl hydrolases known as endoglucanases [146]. SNDs have been implicated to be up accumulated and constitutively expressed in cells requiring a fine tuning of gene expression, regulation of mRNA stability and degradation, protein sequestration, or in redistribution of nucleic acid derivatives like nitrogen, phosphorus and nucleotide base during plant programmed cell death [147], [148]. To put it simply, many proteins with SND domains are part of the class of proteins known as liquid-liquid phase separation proteins, which may be

necessary for the processing, storage, and transportation of less lipophilic monoterpenoids like P-I or P-II [147]. According to reports, the PDR transporters transfer phytoalexin, camalexin, and other unknown compounds give resistance to a number of fungal infections [149] and petunia transportation of strigolactone short-distance strigolactone transport needs PDR1[150]. Strigolactones, which operate as auxin secondary messengers and are sesquiterpene lactones, have been shown to limit shoot branching and to promote favourable interactions between roots and arbuscular mycorrizal fungus through secretion in root exudates. Strigolactones are recognised to be produced by the oxidative cleavage of carotenoids by CCD, as is the plant hormone abscisic acid (ABA) [151]. Mutants lacking ABA are also lacking in strigolactones, indicating that ABA may operate to positively control the production of strigolactones [152]. Therefore, PDR's co-expression as a key hub in PKS15's special sub-network may aid in the preferential removal of sesquiterpenoids like strigolactones and tetraterpenoids or apocarotenoids like ABA for effective P-I or P-II storage, as well as resulting in increased shoot growth in PKS15 because of their transport away from the shoot system. However, it is impossible to completely rule out the potential of P-I and P-II being transported by the PDR transporter to stolons or roots. FCL alleviates dormancy in plant regeneration sections and negatively modulates ABA signalling in plants [153], [154]. Seed germination has been found to be delayed by the concentration-dependent administration of ABA to plants [153], [154]. Farnesyl pyrophosphate, which is a crucial branch point in the MVA pathway and serves as a precursor of several terpenes including sesquiterpenes, sterols, and triterpenes, is formed when farnesyl diphosphate synthase catalyses the condensation of dimethylallyl diphosphate with two units of isopentenyl pyrophosphate [155]. FCL only reacts with farnesylcysteine, and the resulting products, farnesal or geranyl geranial, are further reduced to farnesol and geranylgeraniol [153], [154]. These prenyl alcohols are then phosphorylated to produce farnesyl pyrophosphate or geranylgeranyl pyrophosphate, which serve as building blocks for the production of terpenes [153], [154]. FCL is a significant hub in the monoterpenoid glycoside sub-networks that are specific to P-I in shoot-only systems like PKS15, suggesting that recycling of farnesylcysteine from prenylated proteins may be necessary for the manufacture of bicyclic monoterpenoid iridoid P-I. The ε -branch of the carotenoid biosynthesis pathway is controlled by lycopene epsilon cyclase, whose downregulation has been shown to increase the amount of β -carotene in plants [156]. To create phytosterols, CAS catalyses the conversion of 2,3-oxidosqalene to cycloartenol [157]. In Panax notoginseng cells, RNA interference of CAS and subsequent overexpression of Farnesyl pyrophosphate synthase led to a decrease in phytosterol levels and an increase in triterpene saponin levels [157]. As a result, our coexpression analyses of distinct sub-networks showed increased phytosterol synthesis in PKS25 when CAS overexpression was taken into account as a key co-expressing and interacting hub. Cellulose synthase was absent in PKS25, which co-expressed unique contacts but did not have the unique interaction involving cellulose synthase to PDR and cellulose synthase to SND. The mechanics and development of the shoot apical meristem have been reported to be impacted by cellulose production by the enzyme cellulose synthase [158]. Further, a unique co-expressed interaction connecting a Terpene/Isoprene synthase (IS) to Tocopherol cyclase (terpenoidmethylated phenol interactions) and CAS have been noted in PKS15 while any interactions involving IS was missing in PKS25. IS was identified with similarity to a terpene synthase like Kaurene synthase (a diterpenoid synthase). Abiotic stress tolerance is increased in transgenic sweet potato plants by tocopherol cyclase, which is necessary for the production of the methylated phenols γ and δ -tocopherol [159]. As a result, both IS and FCL in PKS15 may be responsible for the supply of isoprene units for iridoid biosynthesis. On average, STKs were almost four times upregulated, and they were also found to be PKS25's second most important hub of specifically interacting coexpression. ABA sensitive gene regulation in Arabidopsis has been linked to STKs such as SnRK2.2, SnRK2.3, and SnRK2.6 [160]. Abscisic acid signalling and drought tolerance have been predominantly attributed to STK, whose silencing receptorlike kinase class is known to impart lower ABA sensitivity and drought hypersensitivity [161]. In PKSS, a GDSL esterase lipase was 3.6 fold increased. Plant GDSL esterase lipases are versatile hydrolases that can act as thioesterases, proteases, arylesterases, and phospholipases, among other things [161]. Acetylajmalan esterase, a GDSL esterase lipase from Rauvolfa verticillata, can catalyse the creation of the terpenoid indole alkaloid ajmaline from acetylajmaline [162]. It is known that beta carotene hydroxylase converts beta carotene to zeaxanthin, and overexpression of the beta carotene hydroxylase chyB gene in Arabidopsis boosted the pigments found in xanthophyll (oxygen-containing carotenoids), delayed lipid peroxidation, and accelerated photosynthesis [163]. Co-expressed linkages to PDR in PKSS, which indicate adequate antioxidative protection in PKSS, were found to link antheraxanthin production as well. Gibberellin, a phytohormone, is essential for a number of plant developmental processes, including germination, root elongation, blooming transition, and flower formation [164]. Because of this, co-expression of proteins related to Gibberellin in specific interactions in PKSS indicates control of shoot development in PKSS.

Grape cuttings have been shown to have improved dehydration tolerance and adventitious rooting when ABA 8' -hydroxylase is inhibited by the synthetic inhibitor Abz-E3M [165]. Thus, whereas in PKS25 unique interactions a prominent co-expressed network reveals ABA biosynthesis and growth cessation (CAS, CCD, and STK as major hubs), in PKS15 unique interactions negatively regulate ABA biosynthesis (FCL as a major hub) and have Phytoene synthase and Cellulose synthase in interactive networks which highlights its promoted growth in comparison to PKS25. The PKSS field transcriptome, in contrast to PKS15 and PKS25, is able to degrade ABA (ABA 8' -hydroxylase is the primary hub in PKSS unique sub-networks), while also bringing in gibberellin signalling in co-expressed unique networks. Its shoot development is further aided by the distinct PKSS connections between phytoene synthase and tocopherol cyclase. The enzyme known as phytoene synthase is the first step in the carotenoid pathway, producing phytoene from geranylgeranyl diphosphate[166]. Although the carotenoid biosynthesis pathway's "bottle-neck" enzyme is mostly involved in photosynthetic pathways, carotenoids support the formation of pigment-protein complexes, aid in energy absorption, and transport electrons [166]. In PKS15 and PKSS, phytoene synthase was put in a distinct network without any connections to protein breakdown pathways, indicating continuous shoot development and photosynthesis. Overall Comparative co-expression network analysis between PKSS, PKS15 and PKS25 revealed common nodes have been observed between PKS25, PKS15, and PKSS based on presence of nodes and organ-specific expression (no interactions). Assumed from the foregoing, the global co-expressed terpenoid glycoside specific sub-networks were sufficiently non-informative solely based on the presence of coexpressed nodes but could be nicely described on the basis of interactions. The Pentatricopeptide Repeat Containing Protein (PRC), TF, CYP, IS, and others were some of the key nodes.

Furthermore, comparison of co-expression network PKSTS with PI accumulating shoot samples PKS15 and PKSS showed various findings reported in the chapter 4. Some important components are discussed as followed. Stolons have been identified as the repertoire for P-II and P-I. In plants, respiratory burst oxidase is regarded as the hub of the ROS network [167]. According to reports, triggering ROS enhances the formation of secondary metabolites such as lignans [168]. A collection of common enzymes, including a desaturase, a reductase, a galacturonosyltransferase, a flavin-containing monooxygenase, a GDSL esterase lipase, as well as additional lipases and phospholipases, have also been co-expressedly connected to the IS in the PKSTS unique sub-network. Thus, it appears that PKSTS is the primary location for the

production and storage of certain mono- and triterpenoids. Whereas, while Comparing with PKSS coexpression network it was found that an IS is connected in a closed loop series to a TF, a STK, an FCL, and an ABA 8' -hydroxylase, which closes the loop at the IS, to form the compact closed common interaction network between PKSS and PKSTS. The PDR transporter, as anticipated, was connected to the closed loop even though it wasn't common. A cluster of co-expressed enzymes connected to the IS in both PKSS and PKSTS suggests the production of mono- and triterpenoids in both shoots and stolons in field samples. The PDR transporter's linkage mode, or whether it is directly linked in the closed co-expressed network (as in PKSS) or co-expressed but not linked in the closed network, is the primary difference between the distinct terpenoid networks of PKSS and PKSTS.

The pairwise co-expression network analysis of field grown tissues with differential picroside accumulation was also reported in the chapter 4. In contrast to other similarities between stolons and roots, a non-linked co-expressed methyltransferase with sulphate transporters and peroxidase was discovered. In the early phases of water stress, sulphate transport via sulphate transporters is recognised as xylem-borne chemical cues that come before expression of ABA biosynthesis genes [169]. The non-linked co-expressed sulfate network in both stolons and roots suggests diminished ABA signalling as a common network characteristic. A single network of six co-expressed genes connected in series constitute a linked network in the unique interactions of PKSR. These six genes are called CAS, STK, an IS, phosphoinositide 4phosphate-5-kinase, and a PDR. A member of the ABC transporter B family and a cationchloride co-transporter were found to be co-expressed with this IS. The import of terpenoid and phenylpropanoid moieties into the vacuole to produce picrosides may require transporters that are directly connected to the IS. After picrosides are produced in the roots and exported to various organs, the PDR transporter connected to the series network may be crucial. A NRT is represented by the particular interactions in PKSTS in three different sequences of related interactions. A beta carotene hydroxylase and an ABA 8' -hydroxylase are present in one of these series connections. Aldehyde dehydrogenase, Tocopherol cyclase, Respiratory burst oxidase, and an IS are present in the other connected series. FCL, SND, a CCD, a Lycopene beta cyclase, and an ABCG transporter are among the proteins found in the other connected series. It is highly likely that the NRT in the network carries the terpenoid moiety to roots for additional terpenoid modification (acylation) and is transported back to stolons for storage because the IS is not directly co-expressed related to any transporters in PKSTS unique subnetwork. Farnesyl diphosphate synthase, NRT, Ent-kaurene oxidase, Tocopherol cyclase, Beta carotene hydroxylase, *CCD*, and several other hubs were identified to exist as non-linked separate co-expressed small interactions.

The common main hub between PKSS, PKSTS, and PKSR has been identified as a TF by a comparative common interaction analysis. Thus, it is hypothesised that all the organs under comparison require the biosynthesis of isoprene units, as well as their condensation, cyclization, and attachment of a glucose moiety to the terpenoid unit. These terpenoid glycoside subnetworks demonstrate that isoprene units are primarily exported via long distance transporters following glucosylation for additional modifications such as insertion of phenylpropanoid moiety, etc. by the absence of transporter specific common contacts. Similar to this, after their production, they must be exported via transporters to other organs or to the vacuoles for storage or other purposes

All the TFs in terpenoid-glycoside sub-networks were also identified based on their higher relevance in the biosynthesis [38]. The shared interaction of PKS15 and PKS25 is a key hub for a collection of TFs, including the auxin response factor Ethylene-responsive TF (ERF), bHLH, WRKY, WRKY59, WRKY33, and bZIP-23. Even while the auxin response factor directly interacted with any or all of these TFs, it was crucial that the auxin response factor also actively reciprocated by activating or repressing downstream genes based on the particular TF among the group in charge of the co-expressed circuit. Auxin response factors are often bound by the auxin co-receptors transport inhibitor response 1/AFB-Aux/IAA protein when auxin levels are low, preventing them from activating or inhibiting downstream auxin-responsive genes [170] [60]. Auxin response factors can activate or repress downstream genes when SCFTIR1/AFBE3 ligases bind the auxin co-receptors for destruction at increased auxin concentrations. Rumex palustris under flooding conditions is claimed to induce petiole elongation as an escape mechanism when ethylene builds up inside submerged tissues[171]. Petiole elongation is delayed by ABA. By converting ABA into phaseic acid and downregulating a CCD needed for ABA biosynthesis, ethylene buildup in the plants prevented the synthesis of the amino acid ABA. Furthermore, ABA applied externally prevented the up-accumulation of gibberellin A1 and petiole elongation. During apple ripening, a MYC2 TF bound ERF3 to activate the ethylene pathway gene ACS1 [172]. MYC2 and ERF2 cooperated to prevent ACS1 suppression. In order to fine-tune or decrease ABA sensitivity, bHLH TFs are known to bind E-box elements in the promoters of ABA-responsive genes, controlling a protracted delay or dormancy in plants [173] [174]. By directly interacting with WRKY59 in cotton plants, DEHYDRATION-

RESPONSIVE ELEMENT-BINDING PROTEIN 2 (GhDREB2) produces ABA-independent drought tolerance [175]. In Arabidopsis, WRKY33 is known to adversely control ABA production by acting upstream of CCDs [176]. A mediator of ABA signalling, bZIP-23 controls brassinosteroid signalling while minimising growth arrest brought on by ABA activation [177]. When specialised receptors detect ABA, they transmit the signal to various Ser/Thr kinase groups, which phosphorylate the bZIP TFs. After such post-translational modification, the transcription factors (TFs) become active and bind to particular cis-acting sequences known as abscisic-acid-responsive elements or GC-rich coupling sites, affecting the expression of downstream target genes [178]. Banana fruits have been shown to feature WRKY TFs that bind W-box elements in the promoters of CCDs to activate ABA signalling and cold stress resistance [179]. Thus, in P. kurroa, the fate of other components of the co-expressed linked circuits in terpenoid-glycoside metabolism is actually determined by the co-expression of a particular TF in the primary hub of TF common to PKS15 and PKS25. Unique PKSS's connected coexpressed loop had either a WRKY40 or an AP2D23 TF. WRKY40 overexpression has been linked to the production of anthocyanins triggered by injury [180]. Triple mutants of GOLDEN2-LIKE1 and 2, WRKY40, and ABA have been shown to exhibit ABA hypersensitivity, and WRKY40-related transcription modules have been shown to negatively regulate ABA response [181]. It has been demonstrated that ABA and gibberellin signals are directly regulated antagonistically by APETALA 2 (AP2) TFs with an AP2 domain [182]. A WRKY40 TF in PKSR unique sub-network in Fig. S13 was linked to a Gibberellin related protein via a E3 Ubiquitin-protein ligase probably indicating reduced ABA activation as well as reduced gibberellin response. Together, TFs were missing in the closed-linked loops in unique co-expressed terpenoid glycoside sub-network of PKS25. On the other hand, three TFs alternated the closed circuitry in PKS15 unique co-expressed sub-network at the same hubpoint. In PKSS unique co-expressed sub-network, two TFs alternated the same hub-point in the closed loop. These TFs in PKSS have been reported earlier to negatively regulate ABA and positively regulate secondary metabolites biosynthesis. In PKSTS unique co-expressed subnetwork, two TFs negatively regulated the CCD-ABA linked co-expressed circuit. Although no TFs were linked to the IS linked series in PKSR unique co-expressed network, a single coexpressed TF separately and negatively regulated ABA response. These co-expressed closed links probably act like electronic circuits controlled by a group of TFs. A specific TF from this group of TFs might strengthen specific part/s of the co-expressed loop/s as and when required.

5.2. Co-expression modules contributing to secondary metabolites biosynthesis in *Picrorhiza kurroa*

The primary objective of current study was to identify acyltransferases that are involved in structural modifications of major iridoid glycosides of P. Kurroa as well as to capture what other components interact under specific growth conditions. The next question was to identify cofactors indirectly involved in the biosynthetic machinery of important iridoid glycosides in response to external/internal stimuli. In this direction, we applied gene co-expression approach followed by comparative network analysis of different transcriptome samples. The findings confirmed that every transcriptome sample had unique outcomes that represented their overall system. BAHD-ATs were prominently exclusive possibly towards the last step modifications. The observations from experimentally cultured shoot specific network mainly addressed the role of stress and stress response component in the biosynthesis of secondary metabolites whereas field grown tissue samples mainly included the involvement of BAHD-ATs class for last step modification by donating different acyl groups to the catalpol structure. Interestingly, the transcripts with putative function of squalene monooxygenase (SQM) were observed throughout all the transcriptomes. Since SQM is important component in catalpol biosynthesis from aucubin [35] it can be concluded that catalpol is present throughout the plant (all transcriptome samples). From the network analysis of field grown tissue samples, it was observed that BAHD-AT, and S-acyltransferase were differentially expressed and showed higher levels of expression in PKSTS comparative to others. In case of PKSS, "Oacyltransferase WSD1-like" showed the highest expression, whereas intermediate in PKSR. Furthermore, it was observed that although, BAHD was noticed as a common hub in all three samples, its interactions were different, highlighting the fact that the change in co-expression depends on the tissue and environmental conditions. All interacting nodes have been elaborated in the results section. Since the contents of Picroside-I and Picroside-II have been reported to vary in shoots, roots, and stolons [26] the findings confirmed the correlation of BAHD expression with Picroside-I and II biosynthesis. Furthermore, co-expressed linkage of "E3 ubiquitin-protein ligase" in PKSR was also observed in PKSS which might be causing some proteasomal degradation, due to which the picroside biosynthesis be possibly affected in both tissues. Therefore, it can be hypothesized that picroside biosynthesis occurs throughout the plant tissues, however, the final steps of picrosides modifications occur in the stolon. Apart from this, in PKSR co-expressed linkage of BAHD were "Serine carboxypeptidase like protein"

(SCPL), "Glycogen synthase", and "Cytochrome P450" specialized for other secondary metabolite biosynthesis. For example, SCPLs are involved in the biosynthesis of other metabolites by performing trans-acylation function via alternative route, Glycogen synthase (UGTs) is responsible for cucurbitacin biosynthesis, Cytochrome P450 is involved in the biosynthesis of other iridoid glycosides. In PKSS, co-expressed node "CBL-interacting serine threonine-protein kinase" is involved in stress tolerance whereas in PKSTS co-expressed linkage PPO is observed to be significant in providing substrates to BAHD-ATs. To corroborate findings, the molecular docking has been performed where results of 6 different BAHD-ATs showed specificity with respect to transcriptome samples. SS 3469 derived from shoots showed highest binding affinity with Cinnamoyl CoA that is possible candidate for addition of cinnamoyl moiety to catalpol to form Picroside I whereas, STS 4084, STS 4241, STS 8424, SR 4494, and SR 4510 derived from stolons and roots showed variety of binding affinities with other acyl-group donors. Therefore, docking study suggested the potential BAHD-ATs for final step modifications in more than one iridoid glycoside, further suggesting that major organs of secondary metabolites biosynthesis are mainly stolons and roots. Another important outcome based on presence of hubs and mostly expressing transcripts among transcriptome samples indicated that, "Acyltransferase-like At1g54570" (PES2) has been found as the most prominent hub among all the samples, PES2 genes belong to esterase/ lipase/thioesterase acyltransferases family having wide role in employing various classes of acyl donors for various metabolic activities. Further to that, it has role in abiotic stress response involving maintenance of photosynthetic membrane [183]. Furthermore, Both PES1 and PES2 were found expressing in PKS15 and PKS25 with differential expression, where PES2 was higher in PKS25 and PES1 in PKS15, the role of up-accumulation in expression of both PES1 and PES2 affects the chlorophyll degradation and senescence in the leaf [184] hence may be affecting shoot biomass in PKS15 and PKS25. Therefore, presence of such acyltransferase as hub not only have importance in secondary metabolites biosynthesis but also in stress specific condition occurring in tissue culture growth environment of PKS15 and PKS25. In PKS15, overexpression of transketolase highlighted the rehydration in leaves as the function is mainly involved in photosynthesis. Since erythrose-4-phosphate, a product of the pentose phosphate pathway, is involved in the first stage of picroside biosynthesis, transketolase's participation in the pentose phosphate pathway may have an indirect impact in picroside accumulation [35], [185] Moreover, the role of chromosomal maintenance protein is important in stress-related

conditions [186]. *PES2* that is a type of diglycerol O-acyltransferase (DGAT) involved in TAG biosynthesis, is important for energy storage required in stress

hence controlling overall biomass of shoots [187]. Overexpression of transcription factor "WRKY" enhances the MVA pathway that play role in secondary metabolites biosynthesis. Likewise, other double-fold expression transcripts identified in PKS15 network (mentioned in results) are mostly stress-tolerant components . In PKS25, overexpression of "Heat shock cognate 70 kDa" signified the drought stress condition occurring in the plant. The presence of "1- deoxy-D-xylulose 5-phosphate reductoisomerase" (DXPR), a component of nonmevalonate pathway, indicated its role in picroside biosynthesis [35], [188]. The higher expression of "G-type lectin S-receptor-like serine threonine-protein kinase", "RNase H family protein", and "subtilisin-like" showed that transcripts are mostly involved in the plant stress tolerance condition, clearly highlighting the conclusion that PKS25 had abiotic stress condition that ignited expression level to the higher limit [189]–[191]. The stress in PKS25 shoots was clearly visible through reduced growth and biomass compared to shoots grown in PKS15 (Supplementary Fig. 2). In PKSS, "late embryogenesis protein" indicated plant tolerance to dehydration [192]. Beta-glucosidase works as a chemical defence mechanism by producing glucose moieties against herbivores [193], thus important in the secondary metabolites process. Glutamine synthetase is also necessary for various metabolic processes of growth and development activity in plants [194]. Fructose bisphosphate aldolase is an important component of glycolysis in plants [195]. Similarly, "momilactone A synthase-like" and "ABC transporters" had >2 fold higher expression thus, presenting a positive sign of secondary metabolites biosynthesis as both the components are part of specialized biosynthetic gene clusters in terpenoid biosynthesis [196], [197]. Moreover, comparative analysis among all shoot samples showed following observations. The "callose synthase" (CalS) has been observed to have highest expression and significant component in PKSS followed by PKS25 and PKS15 indicating the native response to abiotic stress in respective samples. According to reports, callose is important in innate immunity; therefore, the role of *CalS* in the network should be considered as a key component in samples since all samples were grown in different environmental conditions [198]. Furthermore, the presence of interacting node "ubiquitinconjugating enzyme" indicated degradation of callose that may also increase cell-cell movement indicating the necessity for signalling activities, evidential observations highlighted the similar conclusion [199]-[201]. The differences in the co-expressing linkage were also clearly noticed in the nodes, where CalS was found to interact with MBOAT in PKS25, DGAT

in PKS15, and S-acyltransferase in PKSS. The interaction with MBOAT indicates that CalS activation in the membrane protects the overall structure of the cell. Similarly, interaction with DGAT might indicate the callose stimulation in the membrane since DGAT is also a class of MBOAT [202]. Moreover, the interaction of S-acyltransferase also indicated the same conclusion as it also targets the membrane proteins. Hence, the role of CalS has been considered essential in overall structural maintenance and biosynthesis of secondary metabolites. In PKSTS, PPO is a specialized metabolizing gene in plants, its higher expression is crucial for phenylpropanoid pathways [203]. Amine oxidase also works in the specialized defence mechanism of wound healing in many plants [204]. Furthermore, cytochrome P450 is a key component in iridoid glycosides biosynthesis that catalyses hydroxylation reactions [205]. Overexpression of "Cbl-interacting protein kinase" has indicated a stress response in Stolon [206]. Overall, production of secondary metabolites was one of the stress response activities that was emphasised in individual network of PKSTS as this is the prime metabolizing and storage organ for most of secondary metabolites in P. Kurroa, including iridoid glycosides. In PKSR- the root-derived network, Glutamate decarboxylase showed the synthesis of gammaaminobutyrate (GABA), which is essential for plants in various growth and development activities like cytosolic pH regulation, carbon movement in the TCA cycle, transport, and storage [207]. The overexpression of transcription factor Bzip53 in roots has been reported as essential for reducing salt stress conditions by affecting the primary metabolic function of gluconeogenesis and amino acid catabolism; furthermore, it also orchestrates the lateral root formation in some plant species [208], [209]. Heat shock cognate 70 kDa is a drought stress response component in normal conditions and helps plants to adapt to stress conditions such as high temperatures; therefore, higher expression was observed in sample [188]. EBF protein is an important component of ethylene response and signalling of it corresponds towards the proteasomal degradation and ubiquitination of EIN3, resulting in ethylene response necessary for root elongation [210]. Apart from this, Calcium-binding proteins, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate ispG, and Phosphoenolpyruvate carboxykinase are responsible for root hair formation, component in non-mevalonate pathway, and involved in root elongation in lateral stages respectively [211], [212]. Overall individual network study as well as supporting docking results revealed that major sites of secondary metabolites biosynthesis are stolon and roots as most of associated enzymes were present in individual networks of these tissues.

5.3. Single nucleotide polymorphisms (SNPs) mapped to the global coexpression networks

The main objective of this study was to develop a strategy to capture SNPs mapped to functional modules through combination of GBS and gene co-expression networks. Existing approaches utilizing genome wide studies for scanning high density markers has enabled good quality outcome for plants and animals [213]. Such studies utilized anonymous techniques based on DNA-finger printing i.e. RFLP, RAPD, AFLP and SSRs that had less chances of targeting the expressed regions [214]. The SNP markers from expressed regions were mostly captured through the availability of expression sequence tags databases that are mostly available for model organisms [213]. GBS techniques are ultimately a cost effective solution for plant breeding of such species that are not widely studied [214]. Such technique have been utilized to identify a larger number of SNP biomarkers in various plant species [81], [215], [216]. Although, it is important to identify and categorise such biomarkers on the basis of functional and structural role in the overall system. Hence, our study has proposed a novel approach utilizing mining and mapping of SNPs containing genomic fragments coupled with transcriptome-enabled gene co-expression networks.

GBS data of 37 P. kurroa populations were analysed individually through Denovo strategies for identifying SNPs specific towards high versus low P-I content. The mapping of SNPs containing HPF and LPF on transcriptomes showed favourable probabilities of molecular markers in the expressed regions for identifying relationships among diversified populations. The outcomes of initial analysis in combination with gene co-expression network analysis showed large density of interactions, representing the overall system in Picrorhiza kurroa. The transcriptomes have previously shown a wide range of Picrosides accumulation under various experimental conditions, that might be due to effect on biological process, cellular components, and molecular function. In this context, certain modules such as chloroplast, photosynthesis, signal transduction, root development, defence response, shoot development, and transferases were predominantly captured among the interacting nodes of SNPs containing transcripts. The driving components of secondary metabolites production are signal transduction mechanisms impacted by environmental complexity therefore, crucial for commercial production [217]. The shoot development, photosynthesis and chloroplast development and growth are mainly affected by light intensities influencing the overall secondary metabolites accumulation in plants [218]. Protein kinases modulate series of defence responses therefore leading towards

specialized metabolic pathways [219]. Similarly, relevance of other functional modules such as transferases, transporter activity, root development and vacuoles is important in capturing SNPs containing transcripts. The combined network of the SNPs containing transcripts showed potential coverage of interactions among each functional module. The categories of SNPs were classified as transition and transversions based on the point mutation it possesses, such categories have been reported in various model crop species [214] as both have significant effects in plant breeding strategies. The hubs among specific populations of high/low P-I with similar fragments, that is LPF for low populations and HPF for high populations were considered as favourable outcomes. "protein transport protein SEC61 subunit" a part of Sec61 ER protein that control the susceptibility against fungus infection in plants, therefore considered to be an important factor for defence response [220]. "4-coumarate--CoA ligase-like" (4CL) is an important component of iridoid glycoside pathway, further studies for gene paralogues have reported that the gene makes it crucial to be studied at SNP level for variation and its relationship with picroside accumulation [74]. Aldehyde oxidases are enzymes that oxidises aromatic and non-aromatic specialized aldehyde, are important response components against virus infection in plants therefore considered an important factor for growth and development [221]. Protein phosphatase 2C 55-like (PP2C) are key components of signal transduction in higher plants [222] are negative modulators of various protein kinase activities specific towards stress response especially in ABA signalling [223] therefore, SNP identification in such component is considered of high importance. Leucine aminopeptidase is also a component of the defence response signalling [224] hence is considered to be a key component of the study. Overall, these hubs were specific for LPF fragments specifically consisting of transversions and were only found in co-expression network of low P-I content transcriptome. In case of high P-I population co-expression network hubs mapping with HPF were primarily focused. In particular arginine N-methyltransferases are mostly reported to influence the plant growth and development in model plants species [225]. Helix-loop-helix DNA-binding domains are transcription factors, that have been previously reported in regulation of picrosides biosynthesis [38]. Ankyrin repeat domain 24 are part of ankyrin repeat genes studied for stress tolerance in other plant species [226]. Cold acclimation protein was also one of the major hubs in high population transcriptome coexpression network with specific HPF, show supporting adaptations of Picrorhiza kurroa in cold alpine regions [227]. Uroporphyrinogen decarboxylase a key enzyme for biosynthesis of cholorophyll and heme in plants [228] was also a major hub specifically for high P-I accumulation.

In case of populations based on the PII% in roots, unique hubs specific to HPF-R and LPF-R were focused. The hubs corresponding to LPF-R and mapping only with transcriptome of low PII% were shortlisted. "HEAT repeat" is an important component of innate immunity in plants [228]. "PfkB-type carbohydrate kinase family protein" consists of FRUCTOKINASE-LIKE PROTEIN (FLN) that is reported to have direct effect on plastid-encoded polymerase (PEP) [229] is an important chloroplast gene, therefore, could be crucial for growth and development related functions. "receptor-like protein kinase" (RLKs) are reported to have evolutionary diversity in different domains such as leucine-rich repeats, self-incompatibility domains, epidermal growth factor repeats and lectin domains therefore their presence in SNPs containing hubs makes probable candidates [230], "sucrose transporter" majorly support sink for sucrose transport in plants for growth and development [231].Overall the hubs containing HPF and LPF extracted from populations based on picroside concentrations in different tissues can be very crucial and ideal candidates for biomarkers analysis.

CONCLUSION

The present work consists of novel approaches of gene co-expression networks that can be beneficial for addressing complex systems level problems of large transcriptome datasets, particularly in medicinal plant species which are endemic in the Himalayan region with least genomic resources. The comparative transcriptome profiling using network analysis has envisioned a global approach that was not yet attempted in *Picrorhiza kurroa*. The study has highlighted key components playing indirect role in the specialized metabolites biosynthesis through highly interacting components or as a hub in the co-expression network. Furthermore, the study on acyltransferase specific networks have addressed catalytic validation of certain components through molecular docking. In addition, SNP analysis using diverse GBS data of 41 populations with different picroside concentrations identified SNPs in genes that are also mapping in functional co-expression networks. Overall, the approaches developed in the study can be beneficial in other plants species important for specialized metabolites biosynthesis.

Limitations of the Study

Overall, the study has proposed various promising future endeavors that can be achieved by validating through wet lab experiments. Although there are still many limitations that need to be addressed. Firstly, the diversity of the dataset is optimum and acceptable, but it lacks the quantitative point of view that can be extended in separate studies with having specific questions of interest. Secondly, the study has majorly focused on RNA-seq data that relies on functional annotation based on the sequence similarities therefore there are major chances of having constraints based on query coverage and sequence identity thresholds. Third, the current study is limited to addressing the loopholes in secondary metabolites biosynthetic pathways specifically for iridoid glycosides other functional modules can also be explored. Fourth, study is only limited to population specific SNP analysis advanced characterization of SNPs into synonymous and non-synonymous category can be explored.

FUTURE PROSPECTS

- Expanding co-expression analysis to more NGS transcriptomes from different accessions of *P. kurroa* and to map SNPs onto key modules and genes (associated with metabolites and biomass).
- the study can also be extended using approaches of integrative omics that can only possible when new multi-omics dataset such as genomics, metabolomics and transcriptomics get generated for opening new scope of study.
- In-depth Network based analysis of other GO related terms may play crucial role in various biological pathways which would require further exploration.

LIST OF PUBLICATIONS

International Journal Papers

1. Ashish Sharma, Dipto Bhattacharyya, Shilpa Sharma, Rajinder Singh Chauhan, "Transcriptome profiling reveal key hub genes in co-expression networks involved in Iridoid glycosides biosynthetic machinery in Picrorhiza kurroa", Genomics 2021, September, Pages 3381-3394, (Doi: <u>https://doi.org/10.1016/j.ygeno.2021.07.024</u>) (IF:5.7)

2. Ashish Sharma, Shilpa Sharma, Hemant Sood, Rajinder Singh Chauhan. "*Comparative coexpression networks pinpoint acyltransferases decorating structures of major iridoid glycosides in a medicinal herb, Picrorhiza kurroa*" <u>Plant Gene</u> 2022, September,31,100366 (Doi: https://doi.org/10.1016/j.plgene.2022.100366) (IF:2.5)

3. Ashish Sharma, Shilpa Sharma, Hemant Sood, Rajinder Singh Chauhan. "Mapping SNPs to functional modules/hubs in gene co-expression networks and populations varying for picrosides content in a medicinal herb, Picrorhiza kurroa." (Under submission)

Conferences attended.

1. Ashish Sharma and Rajinder S. Chauhan. 2020. Weighted co-expression network analysis captures key modules & hub genes associated with terpenoids biosynthetic machinery in a medicinal herb, Picrorhiza kurroa Royle ex. Benth. Presented in International Conference on Innovations in Biotechnology and Life Sciences at Delhi Technical University, Dec. 18-20, 2020.

2. Ashish Sharma, Shilpa Sharma, and Rajinder S. Chauhan. 2019. *Transcriptome-enabled co-expression networks capture biosynthetic machinery components of medicinal iridoid glycosides in a Himalayan herb, Picrorhiza kurroa*. Presented in International Conference on Genomics and Proteomics Pertaining to Biological Sciences at Aligarh Muslim University, Nov. 5-7, 2019.

REFERENCES

- M. Erb and D. J. Kliebenstein, "Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy", *Plant Physiol.*, vol. 184, no. 1, p. 52, Sep. 2020
- Z. Pang *et al.*, "Linking Plant Secondary Metabolites and Plant Microbiomes: A Review", *Front. Plant Sci.*, vol. 12, p. 621276, 2021
- [3] O. P. Gupta, S. G. Karkute, S. Banerjee, N. L. Meena, and A. Dahuja, "Contemporary understanding of miRNA-based regulation of secondary metabolites biosynthesis in plants", *Front. Plant Sci.*, vol. 8, p. 374, Mar. 2017
- [4] A. Bar-Even and D. Salah Tawfik, "Engineering specialized metabolic pathways—is there a room for enzyme improvements?", *Curr. Opin. Biotechnol.*, vol. 24, no. 2, pp. 310–319, Apr. 2013
- [5] L. B. Pickens, Y. Tang, and Y.-H. Chooi, "Metabolic Engineering for the Production of Natural Products", *Annu Rev Chem Biomol Eng.*, vol. 2, pp. 211–236, 2014
- [6] M. H. Medema and A. Osbourn, "Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways", *Nat. Prod. Rep.*, vol. 33, no. 8, pp. 951–962, Jul. 2016
- [7] A. Bräutigam and U. Gowik, "What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research", *Plant Biol.*, vol. 12, no. 6, pp. 831–841, Nov. 2010
- [8] V. K. Singh, A. K. Singh, S. Singh, and B. D. Singh, "Next-generation sequencing (NGS) tools and impact in plant breeding", *Adv. Plant Breed. Strateg. Breeding, Biotechnol. Mol. Tools*, vol. 1, pp. 563–612, Jan. 2016
- S. Van Dam, U. V. ~ Osa, A. Van Der Graaf, L. Franke, J. ~ Ao, and P. De Magalh,
 "Gene co-expression analysis for functional classification and gene-disease
 predictions", *Brief. Bioinform.*, vol. 19, no. 4, pp. 575–592, 2018
- [10] X. Niu *et al.*, "Weighted Gene Co-Expression Network Analysis Identifies Critical Genes in the Development of Heart Failure After Acute Myocardial Infarction", *Front. Genet.*, vol. 10, p. 1214, Nov. 2019

- [11] J. Gillis and P. Pavlidis, ""Guilt by Association" Is the Exception Rather Than the Rule in Gene Networks", *PLoS Comput. Biol.*, vol. 8, no. 3, p. 1002444, Mar. 2012
- [12] X. Rao and R. A. Dixon, "Co-expression networks for plant biology: Why and how", *Acta Biochim. Biophys. Sin. (Shanghai).*, vol. 51, no. 10, pp. 981–988, 2019
- [13] Y. Tai *et al.*, "Gene co-expression network analysis reveals coordinated regulation of three characteristic secondary biosynthetic pathways in tea plant (Camellia sinensis)", *BMC Genomics*, vol. 19, no. 1, pp. 1–13, 2018
- [14] H. Cao *et al.*, "Time-series transcriptomic analysis reveals novel gene modules that control theanine biosynthesis in tea plant (Camellia sinensis)", *PLoS One*, vol. 15, no. 9, p. e0238175, 2020
- [15] W. Sun *et al.*, "Weighted gene co-expression network analysis of the dioscin rich medicinal plant Dioscorea nipponica", *Front. Plant Sci.*, vol. 8, p. 789, 2017
- [16] J. Ha *et al.*, "Comprehensive RNA sequencing and co-expression network analysis to complete the biosynthetic pathway of coumestrol, a phytoestrogen", *Sci. Rep.*, vol. 9, p. 1934, 2019
- [17] D. C. J. Wong and J. T. Matus, "Constructing integrated networks for identifying new secondary metabolic pathway regulators in grapevine: Recent applications and future opportunities", *Front. Plant Sci.*, vol. 8, p. 505, 2017
- [18] Y. Tai *et al.*, "Analysis of terpenoid biosynthesis pathways in German chamomile (Matricaria recutita) and Roman chamomile (Chamaemelum nobile) based on coexpression networks", *Genomics*, vol. 112, no. 2, pp. 1055–1064, 2020
- [19] Z. Han *et al.*, "Identification of the gene network modules highly associated with the synthesis of phenolics compounds in barley by transcriptome and metabolome analysis", *Food Chem.*, p. 126862, 2020
- [20] L. Zou, G. Y. Zhong, B. Wu, Y. Yang, S. Li, and Z. Liang, "Effects of sunlight on anthocyanin accumulation and associated co-expression gene networks in developing grape berries", *Environ. Exp. Bot.*, vol. 166, p. 103811, Oct. 2019
- [21] A. Simonyi *et al.*, "The neuroprotective effects of apocynin", *Front. Biosci. Elit.*, vol. 4, no. 6, pp. 2183–93, 2012

- [22] M. Masood, "Picrorhiza kurroa: An ethnopharmacologically important plant species of Himalayan region", *Pure Appl. Biol.*, vol. 4, no. 3, pp. 407–417, 2015
- [23] J. P. Gaddipati, S. Madhavan, G. S. Sidhu, A. K. Singh, P. Seth, and R. K.
 Maheshwari, "Picroliv A natural product protects cells and regulates the gene expression during hypoxia/reoxygenation", *Mol. Cell. Biochem.*, vol. 194, no. 1–2, pp. 271–81, 1999
- [24] P. Verma, V. Basu, V. Gupta, G. Saxena, and L. Ur Rahman, "Pharmacology and Chemistry of a Potent Hepatoprotective Compound Picroliv Isolated from the Roots and Rhizomes of Picrorhiza kurroa Royle ex Benth. (Kutki)", *Curr. Pharm. Biotechnol.*, vol. 10, no. 6, pp. 641–9, 2009
- [25] Q. Jia, M. F. Hong, and D. Minter, "Pikuroside: A novel iridoid from Picrorhiza kurroa", J. Nat. Prod., vol. 62, no. 9, pp. 901–3, 1999
- [26] A. Sharma, D. Bhattacharyya, S. Sharma, and R. S. Chauhan, "Transcriptome profiling reveal key hub genes in co-expression networks involved in Iridoid glycosides biosynthetic machinery in Picrorhiza kurroa", *Genomics*, vol. 113, no. 5, pp. 3381– 3394, 2021
- [27] A. Pal Singh, "Kutkins: A Review of Chemistry and Pharmacology", *Ethnobot. Leafl.*, no. 1, p. 26, 2005
- [28] V. Kumar, A. Bansal, and R. S. Chauhan, "Modular design of picroside-II biosynthesis deciphered through NGS transcriptomes and metabolic intermediates analysis in naturally variant chemotypes of a medicinal herb, picrorhiza kurroa", *Front. Plant Sci.*, vol. 8, no. April, pp. 1–17, 2017
- [29] M. Hattori *et al.*, "Transformation of aucubin to new pyridine monoterpene alkaloids, aucubinines A and B, by human intestinal bacteria", *Phyther. Res.*, vol. 4, no. 2, pp. 66–70, 1990
- [30] Y. Wang *et al.*, "Ex vivo programmed macrophages ameliorate experimental chronic inflammatory renal disease", *Kidney Int.*, vol. 72, no. 3, pp. 290–9, 2007
- [31] Y. Sun, F. Feng, and X. Yu, "Pharmacokinetics of geniposide in Zhi-Zi-Hou-Pu decoction and in different combinations of its constituent herbs", *Phyther. Res.*, vol. 26,

no. 1, pp. 67–72, 2012

- [32] S. Pandit, K. Shitiz, H. Sood, and R. S. Chauhan, "Differential biosynthesis and accumulation of picrosides in an endangered medicinal herb Picrorhiza kurroa", J. *Plant Biochem. Biotechnol.*, vol. 22, no. 3, pp. 335–342, 2013
- [33] V. Kumar, K. Shitiz, R. S. Chauhan, H. Sood, and C. Tandon, "Tracking dynamics of enzyme activities and their gene expression in Picrorhiza kurroa with respect to picroside accumulation", *J. Plant Biochem. Biotechnol.*, vol. 25, no. 2, pp. 125–132, 2016
- [34] H. Sood and R. S. Chauhan, "Development of a low cost micropropagation technology for an endangered medicinal herb (Picrorhiza kurroa) of North-Western Himalayas", J. *Plant Sci.*, vol. 4, no. 2, pp. 21–31, 2009
- [35] K. Shitiz, N. Sharma, T. Pal, H. Sood, and R. S. Chauhan, "NGS Transcriptomes and Enzyme Inhibitors Unravel Complexity of Picrosides Biosynthesis in Picrorhiza kurroa Royle ex. Benth", *PLoS One*, vol. 10, no. 12, pp. 1–19, 2015
- [36] N. Sharma, V. Kumar, R. S. Chauhan, and H. Sood, "Modulation of Picroside-I Biosynthesis in Grown Elicited Shoots of Picrorhiza kurroa In Vitro", *J. Plant Growth Regul.*, vol. 35, no. 4, pp. 965–973, 2016
- [37] W. W. Bhat *et al.*, "Molecular Characterization of UGT94F2 and UGT86C4, Two Glycosyltransferases from Picrorhiza kurrooa: Comparative Structural Insight and Evaluation of Substrate Recognition", *PLoS One*, vol. 8, no. 9, 2013
- [38] I. Vashisht, T. Pal, H. Sood, and R. S. Chauhan, "Comparative transcriptome analysis in different tissues of a medicinal herb, Picrorhiza kurroa pinpoints transcription factors regulating picrosides biosynthesis", *Mol. Biol. Rep.*, vol. 43, no. 12, pp. 1395–1409, 2016
- [39] V. Kumar, "OMICS-Based Approaches for Elucidation of Picrosides Biosynthesis in Picrorhiza Kurroa", Omi. Approaches Plant Biotechnol., 2019
- [40] I. Vashisht, P. Mishra, T. Pal, S. Chanumolu, T. R. Singh, and R. S. Chauhan, "Mining NGS transcriptomes for miRNAs and dissecting their role in regulating growth, development, and secondary metabolites production in different organs of a medicinal

herb, Picrorhiza kurroa", Planta, vol. 241, no. 5, pp. 1255-1268, 2015

- [41] V. Kumar, R. S. Chauhan, and C. Tandon, "Biosynthesis and therapeutic implications of iridoid glycosides from Picrorhiza genus: the road ahead", *J. Plant Biochem. Biotechnol.*, vol. 26, no. 1, pp. 1–13, 2017
- [42] V. Kumar, H. Sood, M. Sharma, and R. S. Chauhan, "A proposed biosynthetic pathway of picrosides linked through the detection of biochemical intermediates in the endangered medicinal herb picrorhiza kurroa", *Phytochem. Anal.*, vol. 24, no. 6, pp. 598–602, 2013
- [43] V. Kumar, N. Sharma, H. Sood, and R. S. Chauhan, "Exogenous feeding of immediate precursors reveals synergistic effect on picroside-I biosynthesis in shoot cultures of Picrorhiza kurroa Royle ex Benth", *Sci. Rep.*, vol. 6, no. April, pp. 1–9, 2016
- [44] V. Kumar *et al.*, "An insight into conflux of metabolic traffic leading to picroside-I biosynthesis by tracking molecular time course changes in a medicinal herb, Picrorhiza kurroa", *Plant Cell. Tissue Organ Cult.*, vol. 123, no. 2, pp. 435–441, 2015
- [45] T. Bontpart, V. Cheynier, A. Ageorges, and N. Terrier, "BAHD or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds", *New Phytol.*, vol. 208, no. 3, pp. 695–707, 2015
- [46] C. Lehfeldt *et al.*, "Cloning of the SNG1 gene of arabidopsis reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism", *Plant Cell*, vol. 12, no. 8, pp. 1295–306, 2000
- [47] S. Irmisch *et al.*, "Discovery of UDP-glycosyltransferases and BAHD-acyltransferases involved in the biosynthesis of the antidiabetic plant metabolite montbretin A", *Plant Cell*, vol. 30, no. 8, pp. 1864–1886, 2018
- [48] M. Peng *et al.*, "Evolutionarily distinct BAHD N-acyltransferases are responsible for natural variation of aromatic amine conjugates in rice", *Plant Cell*, vol. 28, no. 7, pp. 1533–1550, 2016
- [49] A. Kumar *et al.*, "Isolation and HPLC assisted quantification of two iridoid glycoside compounds and molecular DNA fingerprinting in critically endangered medicinal Picrorhiza kurroa Royle ex Benth: implications for conservation", *Physiol. Mol. Biol.*

Plants, vol. 27, pp. 727-746, 2021

- [50] Y. Ganeshkumar, A. Ramarao, and C. Veeresham, "Picroside I and picroside II from tissue cultures of Picrorhiza kurroa", *Pharmacognosy Res.*, vol. 9, no. 1, pp. S53–S56, 2017
- [51] H. Sood and R. S. Chauhan, "Biosynthesis and accumulation of a medicinal compound, Picroside-I, in cultures of Picrorhiza kurroa Royle ex Benth", *Plant Cell. Tissue Organ Cult.*, vol. 100, no. 1, pp. 113–117, 2010
- [52] V. Kumar, H. Sood, and R. S. Chauhan, "Optimization of a preparative RP-HPLC method for isolation and purification of picrosides in Picrorhiza kurroa", *J. Plant Biochem. Biotechnol.*, vol. 25, pp. 208–214, 2016
- [53] P. Singh and R. K. Sharma, "Development of informative genic SSR markers for genetic diversity analysis of Picrorhiza kurroa", *J. Plant Biochem. Biotechnol.*, vol. 29, pp. 144–148, 2020
- [54] J. M. Rawat, B. Rawat, S. Mehrotra, A. Chandra, and S. Nautiyal, "ISSR and RAPD based evaluation of genetic fidelity and active ingredient analysis of regenerated plants of Picrorhiza kurroa", *Acta Physiol. Plant.*, vol. 35, pp. 1797–1805, 2013
- [55] K. Shitiz *et al.*, "Mining simple sequence repeats in Picrorhiza kurroa transcriptomes for assessing genetic diversity among accessions varying for picrosides contents", *Plant Genet. Resour. Characterisation Util.*, vol. 15, no. 1, pp. 79–88, 2017
- [56] R. van Treuren and T. J. L. van Hintum, "Comparison of anonymous and targeted molecular markers for the estimation of genetic diversity in ex situ conserved Lactuca", *Theor. Appl. Genet.*, vol. 119, no. 7, pp. 1265–79, 2009
- [57] M. Ghorbanpour, J. Hadian, S. Nikabadi, and A. Varma, Importance of Medicinal and Aromatic Plants in Human Life, in *Medicinal Plants and Environmental Challenges*, M. Ghorbanpour and A. Varma, Eds. Cham: Springer International Publishing, 2017, pp. 1–23.
- [58] C. A. Jadhav, D. N. Vikhe, and R. S. Jadhav, "Global and Domestic Market of Herbal Medicines: A Review", *Res. J. Sci. Tech*, vol. 12, no. 4, pp. 2349–2988
- [59] P. Balachandran and R. Govindarajan, "Ayurvedic drug discovery.", Expert Opin.

Drug Discov., vol. 2, no. 12, pp. 1631–52, 2007

- [60] S. Duangmano, S. Dakeng, W. Jiratchariyakul, A. Suksamrarn, D. R. Smith, and P. Patmasiriwat, "Antiproliferative effects of cucurbitacin B in breast cancer cells: Down-regulation of the c-Myc/hTERT/telomerase pathway and obstruction of the cell cycle", *Int. J. Mol. Sci.*, vol. 11, no. 12, pp. 5323–5338, 2010
- [61] S. Sanjay, S. H. Banu, and M. Chethankumar, "The study of potentiality of Picrorhiza kurroa root proteins to inhibit free radicals and α-Amylase enzyme", *Asian J. Pharm. Clin. Res.*, vol. 8, no. 2, pp. 220–5, 2015.
- [62] S. N. Shetty, S. Mengi, R. Vaidya, and A. D. B. Vaidya, "A study of standardized extracts of Picrorhiza kurroa Royle ex Benth in experimental nonalcoholic fatty liver disease", J. Ayurveda Integr. Med., vol. 1, no. 3, pp. 203–10, 2010
- [63] N. Sahoo and P. Manchikanti, "Herbal drug regulation and commercialization: An indian industry perspective", J. Altern. Complement. Med., vol. 19, no. 12, pp. 957– 963, Dec. 2013
- [64] R. A. Cacho, Y. Tang, and Y. H. Chooi, "Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi", *Front. Microbiol.*, vol. 5, no. DEC, p. 774, 2014
- [65] T. Dugé de Bernonville, N. Papon, M. Clastre, S. E. O'Connor, and V. Courdavault,
 "Identifying Missing Biosynthesis Enzymes of Plant Natural Products", *Trends Pharmacol. Sci.*, vol. 41, no. 3, pp. 142–146, Mar. 2020
- [66] M. Jain, "Next-generation sequencing technologies for gene expression profiling in plants", *Brief. Funct. Genomics*, vol. 11, no. 1, pp. 63–70, Jan. 2012
- [67] Y. Q. Ma X, Zhao H, Xu W and G. Z. and S. Z. Yan H, "Co-expression Gene Network Analysis and Functional Module Identification in Bamboo Growth and Development", *Front. Genet.*, vol. 9, p. 574, 2018
- [68] S. Mehta, A. K. Sharma, and R. K. Singh, "Advances in Ethnobotany, Synthetic Phytochemistry and Pharmacology of Endangered Herb Picrorhiza kurroa (Kutki): A Comprehensive Review (2010-2020)", *Mini-Reviews Med. Chem.*, vol. 21, no. 19, pp. 2976–2995, 2021

- [69] N. Sharma, R. S. Chauhan, and H. Sood, "Discerning picroside-I biosynthesis via molecular dissection of in vitro shoot regeneration in Picrorhiza kurroa", *Plant Cell Rep.*, vol. 35, no. 8, pp. 1601–1615, 2016
- [70] S. Pandit, K. Shitiz, H. Sood, P. K. Naik, and R. S. Chauhan, "Expression pattern of fifteen genes of non-mevalonate (MEP) and mevalonate (MVA) pathways in different tissues of endangered medicinal herb Picrorhiza kurroa with respect to picrosides content", *Mol. Biol. Rep.*, vol. 40, no. 2, pp. 1053–1063, 2013
- [71] D. R. Zerbino, "Using the Velvet de novo assembler for short-read sequencing technologies", *Curr Protoc Bioinforma.*, vol. chapter 11, no. unit 11.5, 2010
- [72] I. Vashisht, T. Pal, A. Bansal, and R. S. Chauhan, "Uncovering interconnections between kinases vis-à-vis physiological and biochemical processes contributing to picroside-I biosynthesis in a medicinal herb, Picrorhiza kurroa Royle ex. Benth", *Acta Physiol. Plant.*, vol. 40, p. 115, 2018
- [73] P. Gahlan *et al.*, "De novo sequencing and characterization of Picrorhiza kurrooa transcriptome at two temperatures showed major transcriptome adjustments", *BMC Genomics*, vol. 13, p. 126, 2012
- [74] A. Kharb and R. S. Chauhan, "Complexity of gene paralogues resolved in biosynthetic pathway of hepatoprotective iridoid glycosides in a medicinal herb, Picrorhiza kurroa through differential NGS transcriptomes", *Mol. Genet. Genomics*, vol. 296, no. 4, pp. 863–876, 2021
- [75] A. Kharb *et al.*, "Capturing acyltransferase(s) transforming final step in the biosynthesis of a major Iridoid Glycoside, (Picroside-II) in a Himalayan Medicinal Herb, Picrorhiza kurroa", *Mol. Biol. Rep.*, vol. 49, no. 6, pp. 5567–5576, 2022
- [76] R. Pandey, A. Sharma, H. Sood, and R. S. Chauhan, "ABC transporters mined through comparative transcriptomics associate with organ-specific accumulation of picrosides in a medicinal herb, Picrorhiza kurroa", *Protoplasma*, 2022
- [77] T. Sharma *et al.*, "The first draft genome of Picrorhiza kurrooa, an endangered medicinal herb from Himalayas", *Sci. Rep.*, vol. 11, p. 14944, 2021
- [78] T. Lübberstedt, Diagnostics in Plant Breeding, in *Diagnostics in Plant Breeding*, T.
Lübberstedt and R. K. Varshney, Eds. Dordrecht: Springer Netherlands, 2013, pp. 3–9.

- [79] O. A. Zabotina, Metabolite-Based Biomarkers for Plant Genetics and Breeding, in *Diagnostics in Plant Breeding*, T. Lübberstedt and R. K. Varshney, Eds. Dordrecht: Springer Netherlands, 2013, pp. 281–309.
- [80] W. Powell et al., "The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis", *Mol. Breed.*, vol. 2, no. 3, pp. 225–238, 1996
- [81] A. İpek, M. İpek, S. Ercişli, and N. A. Tangu, "Transcriptome-based SNP discovery by GBS and the construction of a genetic map for olive", *Funct. Integr. Genomics*, vol. 17, no. 5, pp. 493–501, 2017
- [82] J. He *et al.*, "Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding", *Front. Plant Sci.*, vol. 30, no. 5, p. 484, 2014
- [83] S. E. Levy and R. M. Myers, "Advancements in Next-Generation Sequencing", Annu. Rev. Genomics Hum. Genet., vol. 17, pp. 95–115, 2016
- [84] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, "Next-Generation Machine Learning for Biological Networks", *Cell*, vol. 173, no. 7, pp. 1581–1592, 2018
- [85] E. R. Mardis, "The impact of next-generation sequencing technology on genetics.", *Trends Genet.*, vol. 24, no. 3, pp. 133–141, Mar. 2008
- [86] T. Obayashi, Y. Aoki, S. Tadaka, Y. Kagaya, and K. Kinoshita, "ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index", *Plant Cell Physiol.*, vol. 59, no. 1, p. e3, 2018
- [87] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics", *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, Jan. 2009
- [88] U. Nagalakshmi, K. Waern, and M. Snyder, "RNA-Seq: a method for comprehensive transcriptome analysis", *Curr. Protoc. Mol. Biol.*, vol. Chapter 4, no. SUPPL. 89, 2010
- [89] Y. Chu and D. R. Corey, "RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation", *https://home.liebertpub.com/nat*, vol. 22, no. 4, pp. 271–274, Aug. 2012

- [90] G. F. Schneider and C. Dekker, "DNA sequencing with nanopores", *Nat. Biotechnol.*, vol. 30, no. 4, pp. 326–328, Apr. 2012
- [91] H. K. Yalamanchili, Y.-W. Wan, and Z. Liu, "Data analysis pipeline for RNA-seq experiments: From differential expression to cryptic splicing", *Curr Protoc Bioinforma.*, vol. 59:11.15.1, 2017
- [92] R. Schmieder and R. Edwards, "Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets", *PLoS One*, vol. 6, no. 3, p. e17288, 2011
- [93] Illumina, "Quality Scores for Next-Generation Sequencing: Assessing sequencing accuracy using Phred queality scoring", *Qual. scores next-generations Seq.*, pp. 1–2, 2011 [Online]. Available: https://www.illumina.com/documents/products/technotes/technote Q-Scores.pdf
- [94] FastQC. Jun. 2015. [Online]. Available: https://qubeshub.org/resources/fastqc
- [95] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data", *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, Aug. 2014
- [96] M. G. Grabherr *et al.*, "Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data", *Nat. Biotechnol.*, vol. 29, no. 7, p. 644, Jul. 2011
- [97] S. D. Jackman *et al.*, "ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter", *Genome Res.*, vol. 27, no. 5, pp. 768–777, May 2017
- [98] C. Camacho *et al.*, "BLAST+: Architecture and applications", *BMC Bioinformatics*, vol. 10, p. 421, 2009
- [99] S. F. Altschul *et al.*, "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs", *Nucleic Acids Res.*, vol. 25, no. 17, pp. 3389–402, 1997
- [100] V. H. Nguyen and D. Lavenier, "PLAST: Parallel local alignment search tool for database comparison", *BMC Bioinformatics*, vol. 10, p. 329, 2009
- [101] P. G. Koutrouli M, Karatzas E, Paez-Espino D, "A Guide to Conquer the Biological Network Era Using Graph Theory", *Front. Bioeng. Biotechnol.*, vol. 8, p. 34, 2020
- [102] S. van Dam, T. Craig, and J. P. de Magalhães, "GeneFriends: a human RNA-seq-based

gene and transcript co-expression database", *Nucleic Acids Res.*, vol. 43, no. D1, pp. D1124–D1132, Jan. 2015

- [103] A. E. Carpenter and D. M. Sabatini, "Systematic genome-wide screens of gene function", *Nat. Rev. Genet. 2004 51*, vol. 5, no. 1, pp. 11–22, Jan. 2004
- [104] R. Albert, H. Jeong, and A. L. Barabási, "Error and attack tolerance of complex networks", *Nat. 2000 4066794*, vol. 406, no. 6794, pp. 378–382, Jul. 2000
- [105] D. Silverbush and R. Sharan, "A systematic approach to orient the human protein– protein interaction network", *Nat. Commun. 2019 101*, vol. 10, no. 1, pp. 1–9, Jul. 2019
- [106] M. Costanzo et al., "A global genetic interaction network maps a wiring diagram of cellular function", Science (80-.)., vol. 353, no. 6306, p. aaf1420, Sep. 2016
- [107] B. He and K. Tan, "Understanding transcriptional regulatory networks using computational models", *Curr Opin Genet Dev*, vol. 37, pp. 101–108, 2016
- [108] E. U. Azeloglu and R. Iyengar, "Signaling Networks: Information Flow, Computation, and Decision Making", *Cold Spring Harb. Perspect. Biol.*, vol. 7, no. 4, 2015
- [109] L. Hakes, J. W. Pinney, D. L. Robertson, and S. C. Lovell, "Protein-protein interaction networks and biology - What's the connection?", *Nat. Biotechnol.*, vol. 26, no. 1, pp. 69–72, 2008
- [110] D. Szklarczyk *et al.*, "STRING v10: protein–protein interaction networks, integrated over the tree of life", *Nucleic Acids Res.*, vol. 43, no. D1, pp. D447–D452, Jan. 2015
- [111] D. Croft *et al.*, "The Reactome pathway knowledgebase", *Nucleic Acids Res.*, vol. 42, no. D1, pp. D472–D477, Jan. 2014
- [112] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "Data, information, knowledge and principle: back to metabolism in KEGG", *Nucleic Acids Res.*, vol. 42, no. D1, pp. D199–D205, Jan. 2014
- [113] D. Szklarczyk *et al.*, "The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored", *Nucleic Acids Res.*, vol. 39, pp. D561-8, 2011
- [114] K. Zuberi et al., "GeneMANIA Prediction Server 2013 Update", Nucleic Acids Res.,

vol. 41, no. W1, pp. W115–W122, Jul. 2013

- [115] T. Schmitt, C. Ogris, and E. L. L. Sonnhammer, "FunCoup 3.0: database of genomewide functional coupling networks", *Nucleic Acids Res.*, vol. 42, no. D1, pp. D380– D388, Jan. 2014
- [116] Y. Niu, D. Otasek, and I. Jurisica, "Evaluation of linguistic features useful in extraction of interactions from PubMed; Application to annotating known, high-throughput and predicted interactions in I2D", *Bioinformatics*, vol. 26, no. 1, pp. 111–119, Jan. 2010
- [117] A. Dräger and H. Planatscher, "Metabolic Networks", *Encycl. Syst. Biol.*, pp. 1249– 1251, 2013
- [118] M. R. Riaz, G. M. Preston, and A. Mithani, "MAPPS: A Web-Based Tool for Metabolic Pathway Prediction and Network Analysis in the Postgenomic Era", ACS Synth. Biol., vol. 9, no. 5, pp. 1069–1082, 2020
- [119] C. Wang *et al.*, "Structural and Biochemical Insights Into Two BAHD Acyltransferases (AtSHT and AtSDT) Involved in Phenolamide Biosynthesis", *Front. Plant Sci.*, vol. 11 610118, 2021
- [120] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation", *Nucleic Acids Res.*, vol. 44, no. D1, p. 4, 2016
- [121] G. Bebek, "Identifying Gene Interaction Networks", *Methods Mol. Biol.*, vol. 850, pp. 483–94, 2012
- [122] A. Baryshnikova, M. Costanzo, C. L. Myers, B. Andrews, and C. Boone, "Genetic Interaction Networks: Toward an Understanding of Heritability", *Annu. Rev. Genomics Hum. Genet.*, vol. 14, pp. 111–33, 2013
- [123] L. Van den Broeck, M. Gordon, D. Inzé, C. Williams, and R. Sozzani, "Gene Regulatory Network Inference: Connecting Plant Biology and Mathematical Modeling", *Front. Genet.*, vol. 11, p. 457, May 2020
- [124] J. Liu et al., "BinPacker: Packing-Based De Novo Transcriptome Assembly from RNA-seq Data", PLoS Comput. Biol., vol. 12, no. 2, pp. 1–15, 2016
- [125] Y. Peng, H. C. M. Leung, S. M. Yiu, and F. Y. L. Chin, IDBA A practical iterative

De Bruijn graph De Novo assembler, 2010.

- [126] E. Bushmanova, D. Antipov, A. Lapidus, and A. D. Prjibelski, "RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data", *Gigascience*, vol. 8, no. 9, p. giz100, 2019
- [127] B. Bushnell, BBMap: A Fast, Accurate, Splice-Aware Aligner, 2014.
- [128] D. G. Gilbert, Gene-omes built from mRNA-seq not genome DNA, in 7th Annual Arthropod Genomics Symposium, 2013, p. 5:1695.
- [129] J. Huerta-Cepas *et al.*, "EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses", *Nucleic Acids Res.*, vol. 8, no. 47, pp. D309–D314, 2019
- [130] Gene Ontology Consortium, "The Gene Ontology (GO) database and informatics resource", *Nucleic Acids Res.*, vol. 1, no. 32, pp. D258-61, 2004
- [131] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: New perspectives on genomes, pathways, diseases and drugs", *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353–D361, 2017
- [132] R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford, "Salmon provides fast and bias-aware quantification of transcript expression", *Nat. Methods*, vol. 14, pp. 417–419, 2017
- [133] H. Li *et al.*, "The Sequence Alignment/Map format and SAMtools", *Bioinformatics*, vol. 25, no. 16, pp. 2078–9, 2009
- [134] V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, and P. Geurts, "Inferring regulatory networks from expression data using tree-based methods", *PLoS One*, vol. 5, no. 9, p. e12776, 2010
- [135] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, "MEGA X: Molecular evolutionary genetics analysis across computing platforms", *Mol. Biol. Evol.*, vol. 35, no. 6, pp. 1547–1549, 2018
- [136] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading", J. Comput. Chem., vol. 31, no. 2, pp. 455–61, 2009

- [137] V. Kumar, V. Kumar, R. S. Chauhan, H. Sood, and C. Tandon, "Cost effective quantification of picrosides in Picrorhiza kurroa by employing response surface methodology using HPLC-UV", J. Plant Biochem. Biotechnol., vol. 24, no. 4, pp. 376– 384, 2015
- [138] J. M. Catchen, A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait, "Stacks: Building and genotyping loci de novo from short-read sequences", *G3 Genes, Genomes, Genet.*, vol. 1, no. 3, pp. 171–82, 2011
- [139] J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, "Stacks: An analysis tool set for population genomics", *Mol. Ecol.*, vol. 22, no. 11, pp. 3124–40, 2013
- [140] H. Singh, P. Gahlan, and S. Kumar, "Cloning and expression analysis of ten genes associated with picrosides biosynthesis in Picrorhiza kurrooa", *Gene*, vol. 515, no. 2, pp. 320–8, 2013
- [141] P. Singh *et al.*, "Spatial transcriptome analysis provides insights of key gene(s) involved in steroidal saponin biosynthesis in medicinally important herb Trillium govanianum", *Sci. Rep.*, vol. 7, no. October 2016, pp. 1–12, 2017
- [142] A. J. Nagano *et al.*, "Deciphering and prediction of transcriptome dynamics under fluctuating field conditions", *Cell*, vol. 51, no. 6, pp. 1358–1369, 2012
- [143] X. Zhang, M. Gou, and C. J. Liu, "Arabidopsis kelch repeat F-Box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonialyase", *Plant Cell*, vol. 25, no. 12, pp. 4994–5010, 2013
- [144] C. Weitzel and H. T. Simonsen, "Cytochrome P450-enzymes involved in the biosynthesis of mono- and sesquiterpenes", *Phytochem. Rev.*, vol. 14, pp. 7–24, 2015
- [145] T. L. Speicher, P. Z. Li, and I. S. Wallace, "Phosphoregulation of the plant cellulose synthase complex and cellulose synthase-like proteins", *Plants*, vol. 7, no. 3, p. 52, 2018
- [146] Y. Xiao, X. He, Y. Ojeda-Lassalle, C. Poovaiah, and H. D. Coleman, "Expression of a hyperthermophilic endoglucanase in hybrid poplar modifies the plant cell wall and enhances digestibility", *Biotechnol. Biofuels*, vol. 11, p. 225, 2018

- [147] P. N. M. V. N. U. and P. V. B. Emilio Gutierrez-Beltran, Pernilla H. Elander, Kerstin Dalman, Jose Luis Crespo, "Tudor staphylococcal nuclease acts as a docking platform for stress granule components in Arabidopsis thaliana", *bioRxiv*, vol. 37, pp. 19–23, 2020.
- [148] K. Leśniewicz, E. Poreba, M. Smolarkiewicz, N. Wolff, S. Stanisławski, and P.
 Wojtaszek, "Plant plasma membrane-bound staphylococcal-like DNases as a novel class of eukaryotic nucleases", *BMC Plant Biol.*, vol. 26, no. 12, p. 195, 2012
- [149] Y. He *et al.*, "The arabidopsis pleiotropic drug resistance transporters PEN3 and PDR12 mediate camalexin secretion for resistance to botrytis cinerea", *Plant Cell*, vol. 31, no. 9, pp. 2206–2222, 2019
- [150] K. Shiratake, M. Notaguchi, H. Makino, Y. Sawai, and L. Borghi, "Petunia PLEIOTROPIC DRUG RESISTANCE 1 Is a Strigolactone Short-Distance Transporter with Long-Distance Outcomes", *Plant Cell Physiol.*, vol. 60, no. 8, pp. 1722–1733, 2019
- [151] A. Pandey, M. Sharma, and G. K. Pandey, "Emerging roles of strigolactones in plant responses to stress and development", *Front. Plant Sci.*, vol. 5, no. 7, p. 434, 2016
- [152] J. A. López-Ráez et al., "Does abscisic acid affect strigolactone biosynthesis?", New Phytol., vol. 187, no. 2, pp. 343–354, 2010
- [153] D. H. Huizinga *et al.*, "Farnesylcysteine lyase is involved in negative regulation of abscisic acid signaling in arabidopsis", *Mol. Plant*, vol. 3, no. 1, pp. 143–55, 2010
- [154] D. N. Crowell, D. H. Huizinga, A. K. Deem, C. Trobaugh, R. Denton, and S. E. Sen, "Arabidopsis thaliana plants possess a specific farnesylcysteine lyase that is involved in detoxification and recycling of farnesylcysteine", *Plant J.*, vol. 50, no. 5, pp. 839– 47, 2007
- [155] X. Wang *et al.*, "Molecular Cloning, Characterisation, and Heterologous Expression of Farnesyl Diphosphate Synthase from Sanghuangporus baumii", *Mol. Biotechnol.*, vol. 62, no. 2, pp. 132–141, 2020
- [156] D. Richaud, C. Stange, A. Gadaleta, P. Colasuonno, R. Parada, and A. R. Schwember, "Identification of Lycopene epsilon cyclase (LCYE) gene mutants to potentially

increase β-carotene content in durum wheat (Triticum turgidum L.ssp. Durum) through TILLING", *PLoS One*, vol. 13, no. 12, p. e0208948, 2018

- [157] Y. Yang, F. Ge, Y. Sun, D. Liu, and C. Chen, "Strengthening triterpene saponins biosynthesis by over-expression of farnesyl pyrophosphate synthase gene and RNA interference of cycloartenol synthase gene in Panax notoginseng cells", *Molecules*, vol. 22, no. 4, p. 581, 2017
- [158] A. Sampathkumar *et al.*, "Primary wall cellulose synthase regulates shoot apical meristem mechanics and growth", *Dev.*, vol. 146, no. 10, 2019
- [159] S. E. Kim *et al.*, "Transgenic sweetpotato plants overexpressing tocopherol cyclase display enhanced α-tocopherol content and abiotic stress tolerance", *Plant Physiol. Biochem.*, vol. 144, pp. 436–44, 2019
- [160] K. Lu *et al.*, "The Arabidopsis kinase-associated protein phosphatase KAPP, interacting with protein kinases SnRK2.2/2.3/2.6, negatively regulates abscisic acid signaling", *Plant Mol. Biol.*, vol. 102, pp. 199–212, 2020
- [161] L. N. Ding *et al.*, "Advances in plant GDSL lipases: from sequences to functional mechanisms", *Acta Physiol. Plant.*, vol. 41, p. 151, 2019
- [162] M. Ruppert, J. Woll, A. Giritch, E. Genady, X. Ma, and J. Stöckigt, "Functional expression of an ajmaline pathway-specific esterase from Rauvolfia in a novel plantvirus expression system", *Planta*, vol. 222, no. 5, pp. 888–98, 2005
- [163] Q. Zhao, G. Wang, J. Ji, C. Jin, W. Wu, and J. Zhao, "Over-expression of Arabidopsis thaliana β-carotene hydroxylase (chyB) gene enhances drought tolerance in transgenic tobacco", J. Plant Biochem. Biotechnol., vol. 23, pp. 190–198, 2014
- [164] J. Binenbaum, R. Weinstain, and E. Shani, "Gibberellin Localization and Transport in Plants", *Trends Plant Sci.*, vol. 23, p. 5, 2018
- [165] H. Tomiyama *et al.*, "Inhibition of Abscisic Acid 8'-Hydroxylase Affects Dehydration TolerancTomiyama, H., Sato, M., Opio, P., Saito, T., Ohkawa, K., Ohara, H., Todoroki, Y., & Kondo, S. (2020). Inhibition of Abscisic Acid 8'-Hydroxylase Affects Dehydration Tolerance and Root", *J. Plant Growth Regul.*, vol. 39, pp. 1577–1586, 2020

- [166] Z. Wang *et al.*, "Characterization and functional analysis of phytoene synthase gene family in tobacco", *BMC Plant Biol.*, vol. 21, p. 32, 2021
- [167] N. Suzuki, G. Miller, J. Morales, V. Shulaev, M. A. Torres, and R. Mittler,
 "Respiratory burst oxidases: The engines of ROS signaling", *Curr. Opin. Plant Biol.*, vol. 14, no. 6, pp. 691–9, 2011
- [168] S. Hazra, D. Bhattacharyya, and S. Chattopadhyay, "Methyl jasmonate regulates podophyllotoxin accumulation in podophyllum hexandrum by altering the ROSresponsive podophyllotoxin pathway gene expression additionally through the down regulation of few interfering miRNAs", *Front. Plant Sci.*, vol. 14, no. 8, p. 164, 2017
- [169] E. L. *et al.*, "Sulphate as a xylem-borne chemical signal precedes the expression of ABA biosynthetic genes in maize roots", *J. Exp. Bot.*, vol. 61, no. 12, pp. 3395–405, 2010.
- [170] M. Galli et al., "The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family", Nat. Commun., vol. 9, p. 4526, 2018
- [171] J. J. Benschop *et al.*, "Contrasting interactions between ethylene and abscisic acid in Rumex species differing in submergence tolerance", *Plant J.*, vol. 44, no. 5, pp. 756– 68, 2005
- [172] T. Li *et al.*, "The jasmonate-activated transcription factor MdMYC2 regulates ETHYLENE RESPONSE FACTOR and ethylene biosynthetic genes to promote ethylene biosynthesis during apple fruit ripening", *Plant Cell*, vol. 29, no. 6, pp. 1316– 1334, 2017
- [173] "Zhang Z, Xu J, Sheng Z, Sui Y, Palli SR (2011) Steroid receptor co-activator is required for juvenile hor- mone signal transduction through a bHLH-PAS transcription factor, methoprene tolerant. J Biol Chem 286: 8437–8447. doi: 10.1074/jbc.M110.191684 PMID".
- [174] J. Kim and H. Y. Kim, "Molecular characterization of a bHLH transcription factor involved in Arabidopsis abscisic acid-mediated response", *Biochim. Biophys. Acta*, vol. 1759, no. 3–4, pp. 191–194, Mar. 2006
- [175] F. Li et al., "Regulation of cotton (Gossypium hirsutum) drought responses by

mitogen-activated protein (MAP) kinase cascade-mediated phosphorylation of GhWRKY59", *New Phytol.*, vol. 215, no. 4, pp. 1462–1475, 2017

- [176] S. Liu, B. Kracher, J. Ziegler, R. P. Birkenbihl, and I. E. Somssich, "Negative regulation of ABA Signaling By WRKY33 is critical for Arabidopsis immunity towards Botrytis cinerea 2100", *Elife*, vol. 15, no. 4, p. e07295, 2015
- [177] X. Wang *et al.*, "Differentially expressed bZIP transcription factors confer multitolerances in Gossypium hirsutum L.", *Int. J. Biol. Macromol.*, vol. 1, no. 146, pp. 569–578, 2020
- [178] A. Banerjee and A. Roychoudhury, "Abscisic-acid-dependent basic leucine zipper (bZIP) transcription factors in plant abiotic stress", *Protoplasma*, vol. 254, no. 1, pp. 3– 16, 2017
- [179] D. L. Luo, L. J. Ba, W. Shan, J. F. Kuang, W. J. Lu, and J. Y. Chen, "Involvement of WRKY Transcription Factors in Abscisic-Acid-Induced Cold Tolerance of Banana Fruit", J. Agric. Food Chem., vol. 65, no. 18, pp. 3627–3635, 2017
- [180] J. P. An, X. W. Zhang, C. X. You, S. Q. Bi, X. F. Wang, and Y. J. Hao, "MdWRKY40 promotes wounding-induced anthocyanin biosynthesis in association with MdMYB1 and undergoes MdBT2-mediated degradation", *New Phytol.*, vol. 224, no. 1, pp. 390– 395, 2019
- [181] R. Ahmad *et al.*, "GOLDEN2-LIKE Transcription Factors Regulate WRKY40 Expression in Response to Abscisic Acid", *Plant Physiol.*, vol. 179, no. 4, pp. 1844– 1860, Apr. 2019
- [182] M. W. Yaish *et al.*, "The APETALA-2-Like Transcription Factor OsAP2-39 Controls Key Interactions between Abscisic Acid and Gibberellin in Rice", *PLOS Genet.*, vol. 6, no. 9, p. e1001098, Sep. 2010
- [183] F. Lippold *et al.*, "Fatty acid phytyl ester synthesis in chloroplasts of Arabidopsis", *Plant Cell*, vol. 24, no. 5, pp. 2001–2014, 2012
- [184] R. D. Wehler, Phytyl ester synthases from Arabidopsis thaliana and Solanum lycopersicum, 2018.
- [185] Q. Zhang, T. V. Linnemann, L. Schreiber, and D. Bartels, "The role of transketolase

and octulose in the resurrection plant Craterostigma plantagineum", *J. Exp. Bot.*, vol. 67, no. 11, pp. 3551–9, 2016

- [186] P. Bolaños-Villegas, "The Role of Structural Maintenance of Chromosomes Complexes in Meiosis and Genome Maintenance: Translating Biomedical and Model Plant Research Into Crop Breeding Opportunities", *Front. Plant Sci.*, vol. 31, no. 12, p. 659558, 2021
- [187] A. Cagliari, R. Margis, F. Dos Santos Maraschin, A. C. Turchetto-Zolet, G. Loss, and M. Margis-Pinheiro, "Biosynthesis of triacylglycerols (TAGs) in plants and algae", *Int. J. Plant Biol.*, vol. 2, no. 1, p. e10, 2011
- [188] P. Aghaie and S. A. H. Tafreshi, "Central role of 70-kDa heat shock protein in adaptation of plants to drought stress", *Cell Stress Chaperones*, vol. 25, no. 6, pp. 1071–1081, 2020
- [189] A. Figueiredo, F. Monteiro, and M. Sebastiana, "Subtilisin-like proteases in plantpathogen recognition and immune priming: A perspective", *Front. Plant Sci.*, vol. 5, p. 739, 2014
- [190] J. Kuciński et al., "Evolutionary History and Activity of RNase H1-Like Proteins in Arabidopsis thaliana", Plant Cell Physiol., vol. 61, no. 6, pp. 1107–1119, 2020
- [191] X. L. Sun *et al.*, "GsSRK, a G-type lectin S-receptor-like serine/threonine protein kinase, is a positive regulator of plant tolerance to salt stress", *J. Plant Physiol.*, vol. 170, no. 5, pp. 505–15, 2013
- [192] Y. Olvera-Carrillo, J. L. Reyes, and A. A. Covarrubias, "Late embryogenesis abundant proteins: Versatile players in the plant adaptation to water limiting environments", *Plant Signal. Behav.*, vol. 6, no. 4, pp. 586–589, 2011
- [193] A. V. Morant *et al.*, "β-Glucosidases as detonators of plant chemical defense", *Phytochemistry*, vol. 69, no. 9, pp. 1795–813, 2008
- [194] H. Unno *et al.*, "Atomic structure of plant glutamine synthetase: A key enzyme for plant productivity", *J. Biol. Chem.*, vol. 281, no. 39, pp. 29287–96, 2006
- [195] W. Lu *et al.*, "Identification and characterization of fructose 1,6-bisphosphate aldolase genes in Arabidopsis reveal a gene family with diverse responses to abiotic stresses",

Gene, vol. 503, no. 1, pp. 65–74, 2012

- [196] K. Yazaki, "ABC transporters involved in the transport of plant secondary metabolites", *FEBS Lett.*, vol. 580, no. 4, pp. 1183–91, 2006
- [197] J. Zhang, J. Zhang, and R. J. Peters, Why are momilactones always associated with biosynthetic gene clusters in plants?, *Proceedings of the National Academy of Sciences* of the United States of America. 2020.
- [198] A. Piasecka, N. Jedrzejczak-Rey, and P. Bednarek, "Secondary metabolites in plant innate immunity: Conserved function of divergent chemicals", *New Phytol.*, vol. 206, no. 3, pp. 948–964, 2015
- [199] W. Liu *et al.*, "The ubiquitin conjugating enzyme: An important ubiquitin transfer platform in ubiquitin-proteasome system", *Int. J. Mol. Sci.*, vol. 21, no. 8, p. 2894, 2020
- [200] A. Vatén *et al.*, "Callose Biosynthesis Regulates Symplastic Trafficking during Root Development", *Dev. Cell*, vol. 21, no. 6, pp. 1144–55, 2011
- [201] P. Su *et al.*, "Probing the single key amino acid responsible for the novel catalytic function of ent-kaurene oxidase supported by NADPH-cytochrome P450 reductases in Tripterygium wilfordii", *Front. Plant Sci.*, vol. 8, p. 1756, 2017
- [202] L. Wang *et al.*, "Structure and mechanism of human diacylglycerol O-acyltransferase 1", *Nature*, vol. 581, no. 7808, pp. 329–332, 2020
- [203] M. L. Sullivan, "Beyond brown: Polyphenol oxidases as enzymes of plant specialized metabolism", *Front. Plant Sci.*, vol. 5, p. 783, 2015
- [204] A. Tisi, R. Angelini, and A. Cona, "Wound healing in plants: Cooperation of copper amine oxidase and flavin-containing polyamine oxidase", *Plant Signal. Behav.*, vol. 3, no. 3, pp. 204–6, 2008
- [205] T. D. Nguyen and T. T. T. Dang, "Cytochrome P450 Enzymes as Key Drivers of Alkaloid Chemical Diversification in Plants", *Front. Plant Sci.*, vol. 12, p. 682181, 2021
- [206] L. Chen, Q. Q. Wang, L. Zhou, F. Ren, D. Di Li, and X. B. Li, "Arabidopsis CBLinteracting protein kinase (CIPK6) is involved in plant response to salt/osmotic stress

and ABA", Mol. Biol. Rep., vol. 40, no. 8, pp. 4759-67, 2013

- [207] N. Bouché, A. Fait, M. Zik, and H. Fromm, "The root-specific glutamate decarboxylase (GAD1) is essential for sustaining GABA levels in Arabidopsis", *Plant Mol. Biol.*, vol. 55, no. 3, pp. 315–25, 2004
- [208] L. Hartmann *et al.*, "Crosstalk between two bZIP signaling pathways orchestrates saltinduced metabolic reprogramming in arabidopsis roots", *Plant Cell*, vol. 27, no. 8, pp. 2244–60, 2015
- [209] Y. Yu et al., "Regulation Mechanisms of Plant Basic Leucine Zippers to Various Abiotic Stresses", Front. Plant Sci., vol. 11, p. 1258, 2020
- [210] B. M. Binder *et al.*, "The Arabidopsis EIN3 binding F-box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling", *Plant Cell*, vol. 19, no. 2, pp. 509–23, 2007
- [211] Y. X. Huang, Y. G. Yin, A. Sanuki, N. Fukuda, H. Ezura, and C. Matsukura,
 "Phosphoenolpyruvate carboxykinase (PEPCK) deficiency affects the germination, growth and fruit sugar content in tomato (Solanum lycopersicum L.)", *Plant Physiol. Biochem.*, vol. 96, pp. 417–25, 2015
- [212] M. Kato, T. Aoyama, and M. Maeshima, "The Ca2+-binding protein PCaP2 located on the plasma membrane is involved in root hair development as a possible signal transducer", *Plant J.*, vol. 74, no. 4, pp. 690–700, 2013
- [213] M. W. Ganal, T. Altmann, and M. S. Röder, "SNP identification in crop plants", Curr. Opin. Plant Biol., vol. 12, no. 2, pp. 211–217, 2009
- [214] J. Mammadov, R. Aggarwal, R. Buyyarapu, and S. Kumpatla, "SNP Markers and Their Impact on Plant Breeding", *Int. J. Plant Genomics*, vol. 2012, p. 11, 2012
- [215] E. Yol, M. Basak, S. Kızıl, S. J. Lucas, and B. Uzun, "A High-Density SNP Genetic Map Construction Using ddRAD-Seq and Mapping of Capsule Shattering Trait in Sesame", *Front. Plant Sci.*, vol. 12, p. 917, Jun. 2021
- [216] J. He, X. Zhao, A. Laroche, Z. X. Lu, H. K. Liu, and Z. Li, Genotyping-by-sequencing (GBS), An ultimate marker-assisted selection (MAS) tool to accelerate plant breeding, *Frontiers in Plant Science*. 2014.

- [217] J. Zhao, L. C. Davis, and R. Verpoorte, Elicitor signal transduction leading to production of plant secondary metabolites, *Biotechnology Advances*. 2005.
- [218] L. X. Zhang, Q. S. Guo, Q. S. Chang, Z. B. Zhu, L. Liu, and Y. H. Chen, "Chloroplast ultrastructure, photosynthesis and accumulation of secondary metabolites in Glechoma longituba in response to irradiance", *Photosynthetica*, 2015
- [219] R. Lyu *et al.*, Reprogramming plant specialized metabolism by manipulating protein kinases, *aBIOTECH*. 2021.
- [220] W. J. Zhang, S. Hanisch, M. Kwaaitaal, C. Pedersen, and H. Thordal-Christensen, "A component of the Sec61 ER protein transporting pore is required for plant susceptibility to powdery mildew", *Front. Plant Sci.*, 2013
- [221] T. M. Yergaliyev *et al.*, "The involvement of ROS producing aldehyde oxidase in plant response to Tombusvirus infection", *Plant Physiol. Biochem.*, 2016
- [222] P. L. Rodriguez, Protein phosphatase 2C (PP2C) function in higher plants, *Plant Molecular Biology*. 1998.
- [223] T. Xue et al., "Genome-wide and expression analysis of protein phosphatase 2C in rice and Arabidopsis", BMC Genomics, 2008
- [224] J. H. Fowler, J. Narváez-Vásquez, D. N. Aromdee, V. Pautot, F. M. Holzer, and L. L. Walling, "Leucine Aminopeptidase Regulates Defense and Wound Signaling in Tomato Downstream of Jasmonic Acid", *Plant Cell*, vol. 21, no. 4, p. 1239, 2009
- [225] A. Ahmad, Y. Dong, and X. Cao, "Characterization of the PRMT gene family in rice reveals conservation of arginine methylation", *PLoS One*, 2011
- [226] J. Y. Zhao et al., "The Ankyrin-Repeat Gene GmANK114 Confers Drought and Salt Tolerance in Arabidopsis and Soybean", Front. Plant Sci., 2020
- [227] G. Hu, N. Yalpani, S. P. Briggs, and G. S. Johal, "A porphyrin pathway impairment is responsible for the phenotype of a dominant disease lesion mimic mutant of maize", *Plant Cell*, 1998
- [228] J. Monaghan and X. Li, "The HEAT Repeat Protein ILITYHIA is Required for Plant Immunity", *Plant Cell Physiol.*, vol. 51, no. 5, pp. 742–753, May 2010

- [229] J. Gilkerson, J. M. Perez-Ruiz, J. Chory, and J. Callis, "The plastid-localized pfkB-type carbohydrate kinases FRUCTOKINASE-LIKE 1 and 2 are essential for growth and development of Arabidopsis thaliana", *BMC Plant Biol.*, vol. 12, no. 1, pp. 1–17, Jul. 2012
- [230] J. M. Cock, V. Vanoosthuyse, and T. Gaude, "Receptor kinase signalling in plants and animals: Distinct molecular systems with mechanistic similarities", *Curr. Opin. Cell Biol.*, vol. 14, no. 2, pp. 230–236, Apr. 2002
- [231] N. A. Eckardt, "The function of SUT2/SUC3 sucrose transporters: The debate continues", *Plant Cell*, vol. 15, no. 6, pp. 1259–1262, Jun. 2003