Oleochemical Changes During Seed Development in Oil yielding Legume tree, *Pongamia pinnata* L. Pierre

Thesis submitted to the University of Hyderabad for the award of

Doctor of Philosophy

By

K Tamna Singha

15LPPH07

Supervisor: Prof. Attipalli R Reddy

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad 500046
Telangana (India)

August, 2023

UNIVERSITY OF HYDERABAD

(A Central University established in 1974 by an Act of the Parliament)

HYDERABAD-500046, INDIA

CERTIFICATE

This is to certify that **Mrs. K Tamna Singha** has carried out the research work embodied in the present thesis under the supervision and guidance of **Prof. Attipalli R Reddy** for a full period prescribed under the Ph.D ordinances of this University. We recommend her thesis entitled "**Oleochemical Changes During Seed Development in Oil yielding Legume tree, *Pongamia pinnata* L. Pierre**" for the award of Doctor of Philosophy of the University.

Prof. Attipalli R Reddy

(Supervisor)

Prof. Subramanyam Rajagopal

(Supervisor)

Prof. S. RAJAGOPAL
Dept. of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

UNIVERSITY OF HYDERABAD

(A Central University established in 1974 by an Act of the Parliament)

HYDERABAD-500046, INDIA

DECLARATION

I, **Mrs. K Tamna Singha**, hereby declare that this thesis entitled "**Oleochemical Changes During Seed Development in Oil yielding Legume tree, *Pongamia pinnata* L. Pierre**" submitted by me under the guidance and supervision of **Prof. Attipalli R Reddy** is a bonafide research work. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. A report on plagiarism statistics from the University Librarian is enclosed.

K Tamna Singha

K Tamna Singha

Prof. Attipalli R Reddy

(Supervisor)

Prof. Subramanyam Rajagopal

(Supervisor)

Prof. S. RAJAGOPAL
Dept. of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

UNIVERSITY OF HYDERABAD

(A Central University established in 1974 by an Act of the Parliament)

HYDERABAD-500046, INDIA

CERTIFICATE

This is to certify that the thesis entitled "**Oleochemical Changes During Seed Development in Oil yielding Legume tree, *Pongamia pinnata* L. Pierre**" is a record of bonafide work done by Mrs K Tamna Singha, a research scholar for PhD program under the Department of Plant Science, School of Life Sciences, University of Hyderabad under my supervision and guidance. This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other University or Institution for award of any degree or diploma.

Parts of the thesis have been:

A. Published in the following publications:

1. **Industrial Crops and Products (2019) 140: 111621.**

B. Presented in the following conference:

1. Poster presentation in 8th International Conference on "Photosynthesis and Hydrogen Energy Research for Sustainability" – 2017. Evidence of seed photosynthesis and its role on storage product accumulation in the developing green seeds of *Pongamia pinnata* (L.) Pierre, a potential biofuel tree species.

Further, the student has passed the following courses towards the fulfilment of the coursework requirement for Ph.D. degree.

Sl. No.	Course Code	Name	Credits	Pass/Fail
1.	AS-801	Analytical Techniques	4	**Pass**
2.	AS-802	Research Ethics, Data Analysis and Biostatistics	3	**Pass**
3.	AS-803	Lab Work & Seminar	5	**Pass**

Supervisor

Prof. Attipalli R Reddy

Head

Dept. of Plant Science

अध्यक्ष / HEAD
वनस्पति विज्ञान विभाग / Dept. of Plant Sciences
जैविक विज्ञान संकाय / School of Life Sciences
हैदराबाद विश्वविद्यालय / University of Hyderabad
हैदराबाद / Hyderabad-500 046, भारत / INDIA

Dean

School of Life Sciences

अध्यक्ष / Dean
जीव विज्ञान संकाय / School of Life Sciences
हैदराबाद विश्वविद्यालय / University of Hyderabad
हैदराबाद / Hyderabad-500 046.

UNIVERSITY OF HYDERABAD

(A Central University established in 1974 by an Act of the Parliament)

HYDERABAD-500046, INDIA

PLAGIARISM FREE CERTIFICATE

This is to certify that the similarity index of this thesis, as checked by the Library of the University of Hyderabad, is 36%. Out of this, 31% similarity has been identified from the candidate, K Tamna Singha's own publication which forms the major part of the thesis.

The details of student's publication:

1. **Singha KT, Sreeharsha RV, et al., 2019. Dynamics of metabolites and key regulatory proteins in the developing seeds of *Pongamia pinnata*, a potential biofuel tree species. *Ind Crops Prod*. 140: 111621.**

This publication, constituting a part of the present thesis, have been deposited in various public repositories. All the publications related to this thesis have been appended at the end of the thesis. Only 5% similarity was identified from other sources in the present thesis, which is according to prescribed regulations of the University. Hence, the present thesis may be considered as plagiarism-free.

Prof. Attipalli R Reddy

(Supervisor)

Prof. Subramanyam Rajagopal

(Supervisor)

Prof. S. RAJAGOPAL
Dept. of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

ACKNOWLEDGEMENTS

I am honored and it is a great privilege to express my sincere gratitude to my supervisor **Prof. Attipalli Ramachandra Reddy**, for making me a part of his esteemed research team and by giving me the opportunity to work under his supervision. His guidance and valuable advices had been a blessing for me at every stage of completing the thesis. His suggestions enlighten me to carry out my research work with more enthusiasm and working as a PhD student in his laboratory is one of my greatest privileges I ever achieved.

It is my great privilege to extend my sincere gratitude to my co-supervisor **Prof. Rajagopal Subramanyam,** for his highly valuable advices during the final stages of completing my thesis. My heartfelt thanks for his generosity and guidance which immensely helped me in completing my PhD journey.

I extend my sincere gratitude and thanks to my doctoral committee members Prof. A S Raghavendra and Prof. R P Sharma for their valuable suggestions and constant advices which immensely guided to carry forward the research work.

I would like to thank the present and former heads of the department of Plant Sciences, Prof. Rajagopal Subramaniam, Prof. G. Padmaja and Prof. CH Venkata Ramana for all the departmental facility being provided.

I would also like to thank the present and former deans School of Life Sciences, Prof. P Reddanna, Prof. MNV Prasad, Prof. S Dayananda and Prof. N Shiva Kumar for providing the experimental facility present in the school.

I take this opportunity to extend my heartfelt thanks to Mr. Sreenivas Ghatty, founder of Tree Oil India Limited (TOIL), Zaheerabad, and his team for giving me the opportunity to perform field study including the collection of the samples.

My heartfelt thanks to lab mates Dr. Sreeharsha, Dr. Chaitanya, Dr. Sumit, Dr. Debashree, Dr. Suresh, Dr. Shalini, Dr. Madana, Dr. Sita Rami Reddy, and Divya for maintaining healthy and most cheerful environment in and around the lab.

I thank DBT and BBL for my fellowship and financial aid.
I thank DST, DBT, CSIR and BASF for providing the funds to our laboratory.

My love and heartfelt thanks to all my friends for their endless encouragement and support. My enduring journey of PhD would not have been indulging without them.

Last but not the least, it is my privilege to thank dear mama and dear papa for always being there and supporting me throughout the crests and troughs of my life. Finally, I thank the almighty for all the blessings and teachings showered upon me.

****

This thesis is dedicated to

my daughter, my dear Sija.

LIST OF FIGURES

Details		Page No.
Figure 1.1:	*Difference in structure of dicot and monocot seeds. Image courtesy (Catusse et al., 2012).*	6
Figure 1.2:	*The detail diagrammatic representation of developing embryos in soybean seeds. Stages shown are 1. Globular stage; 2. Heart stage; 3. Cotyledonary stage; 4. Early maturation stage. Embryo proper and suspensor are enclosed by the endosperm, endothelium and epidermis respectively. Outer and inner integument are also classified as shown which provides a protective coating surrounding the developing seeds. Adopted from Goldberg Lab 2015.*	7
Figure 1.3:	*Schematic representation of the fate of sucrose towards sink tissues and the coordinated subcellular network of seed development. When sucrose reaches the developing seeds, it has been successfully metabolized with the help of invertases inside and outside of the sink tissues. Sucrose synthase also take part in breaking of sucrose to hexoses. The hexoses are metabolized with the help of respective hexokinases present inside the developing seeds where they are channelized through different metabolisms in order to initiate the storage product biosynthesis. Stein and Granot 2018.*	13
Figure 1.4:	*Molecular control of storage products mainly the lipids in developing seeds. The precursors for de novo fatty acid synthesis in maturing embryos are derived from sucrose through the glycolytic pathway and/or the OPPP. Fatty acids produced in the plastids are then exported towards the cytosol in the form of acyl CoAs and used to form triacylglycerols which ultimately get stored inside the oil bodies. The amino acids required for the synthesis of storage proteins during the maturation process are either directly imported from the maternal tissues or synthesized/modified in the embryonic tissues. Storage proteins are ultimately stored in specific vacuoles. Solid arrows represent positive transcriptional regulations. The key controlling elements of seed oil biosynthesis such as LEC 1/2, WRI1, FUS 3, ABI 3 are highlighted in the purple boxes. Adapted and modified from Santos-Mendoza et al., 2008, Jiang et al., 2012.*	15
Figure 1.5:	*Overview of major reactions involved in fatty acid and triacylglycerol synthesis. (A) Plastid fatty acid synthesis; (B) Fatty acid transport from plastidial membrane and (C) TAG synthesis. Major events occurring inside the plastids including the production of NADPH and ATP taking part in fatty acid synthase pathway.*	20

Details		Page No.
	Fatty acids are exported with the thioesterases (fat A/B) towards endoplasmic reticulum where the acyl editing and TAG accumulation takes place with the help of DGAT and PDAT. Inspired and modified from Bates et al., 2013.	
Figure 1.6:	*The omics approaches to unfold metabolic regulation. Large-scale studies in the field of omics have been successfully exploring the differences in gene expression, protein and metabolite abundance and modification of post-translational protein, and providing a different level of understanding the cellular processes. Modified from (Gahlot et al., 2013).*	21
Figure 1.7:	*Types of biofuel sources and the by-products which can be used in different automobile industries for replacement of petro-diesel. Adopted and inspired from SHELL Global 2020.*	24
Figure 1.8:	*Potential non-edible oil resources as biodiesel feedstock commonly found across India and around the globe (Kumar and Sharma 2011).*	27
Figure 1.9:	*Published data from our previous publications (a) Percentage of oil content during four developmental stages of P. pinnata. (b) Oil-red staining of seed endosperm showing oil accumulation patterns in P. pinnata seed. (c, d) Four developmental stages of P. pinnata pod and seed used in quantification studies of lipid biosynthetic genes. (Sreeharsha et al., 2016).*	30
Figure 1.10:	*Framing the objectives: understanding the regulation of Seed development in P. pinnata is crucial.*	31
Figure 2.1:	*P. pinnata plantations and germplasms located in TOIL, Zaheerabad, Telangana, India.*	35
Figure 2.2:	*Average weather data. March 2016 - February 2017and March 2017 - February 2018.*	35
Figure 2.3*:*	*High yielding P. pinnata accessions and their characteristics of oil content with Pod yield.*	42
Figure 2.4:	*Time scale from flowering to seed development in P. pinnata. Showing the period from flowering to pod formation till maturation.*	42
Figure 2.5:	*The experimental Layout.*	43
Figure 2.6:	*Schematic procedure followed for the metabolite profiling.*	49
Figure 2.7:	*Schematic procedure followed for the 2-D Proteomic analysis.*	51

Details		Page No.
Figure 3.1:	*Stomatal anatomy of the ventral surface of P. pinnata leaves: (A) SEM image wide view (Scale: 100 µm) (B) SEM image narrow view (Scale:10 µm) (C) Acridine orange-stained stomate observed under confocal microscope (Scale:10 µm).*	56
Figure 3.2:	*Foliar content of total carbohydrate, starch and soluble sugars. Values are mean ± SD (n=3). C – Carbohydrates, S – Sucrose, SS – Soluble sugars.*	57
Figure 3.3:	*The cross sections of randomly collected P. pinnata pods at different stages of development (very young to a matured stage); "a"- Funicle (attachment of the seed with the pod), "b" – Seed coat, "c" – Cotyledons, "d" – Inner layer of Pod wall, "e" – Outer layer of Pod wall. A – D; from a very young to a mature pod. (Scale – 2mm).*	63
Figure 3.4:	*The separated cotyledons from seeds collected at the interval of 30 days from the first appearance of pods (A). Morphological measurements of developing seeds (B). Selected stages for further analysis (C).*	64
Figure 3.5:	*Position of seed inside the pod and cotyledons separated from seed coat at each developmental stage (A). Morphological and biochemical analysis of developing P. pinnata seeds. Changes in seed dry weight, moisture content, % cotyledon and seed coat throughout development (B). Values are mean ± S.D, (n = 40); the scale is presented by white bar with the length of 0.3 cm.*	65
Figure 3.6:	*Changes in the chlorophyll a, b, total chlorophyll and chlorophyll a/b (A). Expression patterns of key regulatory genes represented in fold change related to photosynthesis (B). Fold change >2 is considered as upregulation. Values are mean ± SD (n=3). Data are given as means ± SD; (n=3). *, ** represents significant differences within the stages analyzed by one–way ANOVA, P<0.05; P<0.01 respectively.*	67
Figure 3.7:	*Pattern of increasing storage products (lipid, protein, total carbohydrate and starch) with development. Data are given as means ± SD; (n=3). *, ** represents significant differences within the stages analyzed by one-way ANOVA, P<0.05; P<0.01 respectively.*	68
Figure 3.8:	*Cotyledonary cross section under SEM from S1 to S4; SG – storage granules, CW – cell wall, Scale (20–50 µ) (A). Cotyledonary cross section under TEM from S1 to S4; PSV – protein storage vacuoles, OB – oil bodies, Scale: 0.2–2 µ (B).*	70
Figure 3.9:	*The oil bodies, protein bodies, and starch granules of P. pinnata cotyledonary sections observed under TEM, Scale: 0.2–2 µ.*	71

Details		Page No.
Figure 3.10:	*Nile red staining of seed sections observed under confocal microscope, oil bodies are marked by white arrow indicates oil body accumulated regions, Scale: (5µ).*	72
Figure 3.11:	*HPLC peaks of the major soluble sugars of P. pinnata seeds (cotyledons and seed coat separated) from S1 to S4. "F" – Fructose, "G" – Glucose, "S" – Sucrose.*	75
Figure 3.12:	*Quantification of the soluble sugars of developing P. pinnata seeds (cotyledon and seed coat separated). Values are represented in mg g^{-1} FW of the sample analyzed. Data are given as means ± SD; (n=3).*	76
Figure 3.13:	*The HPLC peaks of sugar observed after the invertase assay at different pH. Calculation was based on the fructose content after 5 min of incubation for all the assay sample. Acidic invertases pH 4 (A); neutral invertases pH 7, alkaline invertases pH 7.5 (B).*	78
Figure 3.14:	*The invertase activities (nanomole h^{-1} g^{-1} fresh weight of the seeds) based on the different pH in the developing P. pinnata seeds. Values are mean ± SD (n=3).*	79
Figure 3.15:	*Changes in the SS and SPS activity from S1 to S4. Their activities are represented as nanomole h^{-1} g^{-1} fresh weight, where left and right axis represents SS and SPS respectively. Values are mean ± SD (n=3).*	81
Figure 3.16:	*Changes in the fructokinase and glucokinase activity from S1 to S4. Their activities are represented as nanomole h^{-1} g^{-1} fresh weight. Values are mean ± SD (n=3).*	82
Figure 3.17:	*Changes in the expression of transcripts belonging to the various enzymes of sugar metabolism in the developing P. pinnata seeds. Values are represented in log 2–fold change of gene expression (n=3).*	83
Figure 3.18:	*Pie charts representing the metabolites based on the biochemical characteristics (A), biological functions (B).*	86
Figure 3.19:	*Heat map of metabolite concentrations (µg g–1 FW) at S1, S2, S3 and S4. Analysis was carried out in Metaboanalyst 4.0.*	87
Figure 3.20:	*The PCA biplot analysis of all the metabolites isolated from S1 to S4.*	88
Figure 3.21:	*Percentage of fatty acids accumulated in the TAGs at different developing stages (A). Total lipid profiling through TLC (B). Values are mean ± SD (n=3). *, ** represents significance difference ($P < 0.05$, $P < 0.01$) between the fatty acid content with respect to 120 DAA, analyzed by t–test.*	90

Details		Page No.
Figure 3.22:	*GCMS peaks of the FAMEs analyzed. "a" – Palmitic acid; 16:0; "b" – Stearic acid; 18:0; "c" – Oleic acid, 18:1; "d" – Linoleic acid, 18:2.*	91
Figure 3.23:	*2D SDS gel picture of all the developing stages (S1 to S4).*	94
Figure 3.24:	*The PCA biplot analysis of all the gel spots observed with the help of Image G (S1 to S4).*	95
Figure 3.25:	*Identified spots with the reference Gel. Selected regions with spot numbers are given for all the developing stages (S1 to S4).*	96
Figure 3.26:	*Pie charts representing the identified proteins based on the similarly of protein accessions of the reference organisms (A). Biological functions (B).*	97
Figure 3.27:	*Changes in the expression of key regulatory proteins from S1 to S4.*	98
Figure 3.28:	*Stage specific identified proteins observed at S1 and S4.*	99
Figure 4.1:	*Diagrammatic representation of developing P. pinnata seeds. SC – seed coat, C –cotyledons.*	107
Figure 4.2:	*Schematic representation of metabolism of sucrose into hexoses (glucose and fructose) and further into different metabolic events.*	113
Figure 4.3:	*Our proposed model for sugar-protein-lipid changes during seed development in P. pinnata.*	114
Figure 4.4:	*Diagrammatic representation of metabolism of sucrose into hexoses (glucose and fructose) and further accumulation of major storage products in the form of lipids and proteins.*	115
Figure 4.5:	*Classification of metabolites based with their respective pathways and metabolism. All the detected and identified metabolites are clubbed into their respective known pathways and other related cellular metabolism.*	117
Figure 4.6:	*Schematic presentation on the regulatory events during oil biosynthesis in P. pinnata seeds (S1 to S4). Stage wise changes in the metabolite concentrations belonging to various pathways in their subcellular locations, which later channelized towards the fatty acid biosynthesis and ultimately the TAG accumulations*	120
Figure 4.7:	*Stage wise expression profile of crucial Proteins controlling key regulations of overall seed development. Developing seeds collected from 120 to 300 DAA (A). S1 to S4 has been categorized into three main phases.*	123

List of Tables

Details		Page No.
Table 1.1:	*Overview of a decade of research (2012 – 2022) on P. pinnata in the relevant focused areas of interest with features and outcomes.*	3
Table 1.2:	*Uses and seed characteristics of few important legumes.*	5
Table 1.3:	*List of prominent non–feed 2nd generation biofuel stock with the lipid and fatty acid compositions. PA– palmitc acid; SA– stearic acid; OA– oleic acid; LA– linoleic acid; LiA– linolenic acid; RC–ricinoleic acid.*	26
Table 3.1:	*Leaf gas exchange characteristics along with stomatal aperture and stomatal density of P. pinnata leaves. Values are mean ± SD (n=3).*	56
Table 3.2:	*Pigment content analysis in the leaves P. pinnata. Values are mean ± SD (n=3).*	56
Table 3.3:	*List of free amino acids and their content detected in the foliar parts of P. pinnata leaves. Values are mean ± SD (n=3).*	58
Table 3.4:	*List of the identified proteins with nano–LCMS along with their functional ontology.*	60,61,62
Table 3.5:	*Stage wise changes in the size and the count per cell of different types of storage granules observed in P. pinnata cotyledonary sections (S1 to S4). The sizes of are in micron (µ) and values are represented as ranges observed in triplicates.*	73
Table 3.6:	*Details of the identified spots with MALDI MSMS.*	100,101
Table 4.1:	*Current information available on the seed development in P. pinnata*	108

ABBREVIATIONS

2-dimentional	**2D**
Abscisic acid	**ABA**
Abscisic acid insensitive 3	**ABI3**
Acridine Orange	**AO**
Acyl carrier protein	**ACP**
Adenosine triphosphate	**ATP**
Days after flowering	**DAF**
Deoxy-ribonucleic acid	**DNA**
Diacylglycerol	**DAG**
Double distilled water	**DDW**
Dry weight	**DW**
Endoplasmic reticulum	**ER**
Fatty acid methyl esters	**FAMEs**
Fatty acids	**FAs**
Fatty acyl-ACP thioesterase A	**FATA**
Fatty acyl-ACP thioesterase B	**FATB**
Free fatty acid	**FFA**
Fresh weight	**FW**
Fructokinase	**FRK**
Fructose	**F**
Fusion	**FUS3**
Gas chromatography	**GC**
Glucose	**G**
Glyceraldehyde-3-P	**GlyAld-3P**
Glycerol-3-phosphate	**G-3-P**
Hexokinase	**HKX**
High pressure liquid chromatography	**HPLC**
Invertase	**INV**
Infra-Red-Gas-Analyzer	**IRGA**
Isopentenyl pyrophosphate	**IPP**
Keto-acyl synthase 2	**KASII**
Label free quantification	**LFQ**
Linoleic acid	**C18:2**
Linolenic acid	**C18:3**
Lyso-phosphatidyl-choline	**Lyso-PC**
Mass spectrometry	**MS**
Matrix Assisted Laser Desorption/Ionization	**MALDI**
Micro ribonucleic acid	**miRNA**
Monoacylglycerol	**MAG**

Oil body	**OB**
Oleic acid	**C18:1**
Palmitic acid	**C16:0**
Phosphatidylcholine	**PC**
Phosphatidylcholine diacylglycerol acyltransferase	**PADT**
Phosphatidylcholine diacylglycerol choline-phospho-transferase	**PDCT**
Phosphatidylcholine synthase	**PCS**
Phosphatidylethanolamine	**PE**
Phosphatidylinositol	**PI**
Phospholipids	**PL**
Photosystem I	**PS1**
Photosystem II	**PSII**
Polyacrylamide gel electrophoresis	**PAGE**
Principle component analysis	**PCA**
Protein bodies	**PB**
Protein storage vacuole	**PSV**
Real-time polymerase chain reaction	**RT-PCR**
Ribonucleic acid	**RNA**
Sodium dodecyl sulphate	**SDS**
Standard deviation	**SD**
Starch granules	**SG**
Stearic acid	**C18:0**
Stearoyl ACP desaturase	**SAD**
Sucrose phosphate synthase	**SPS**
Sucrose synthase	**SS**
Sucrose	**S**
Thin layer chromatography	**TLC**
Tree oil India limited	**TOIL**
Triacylglycerols	**TAGs**

Table of Contents

CONTENTS **PAGES**

CHAPTER 1

INTRODUCTION

1.1	**A short note on highly potential biofuel legume tree *P. pinnata***	**1–3**
1.2	**Typical view on seed development of legumes – A brief History**	**4–9**
1.3	**The storage components of developing legume seeds**	**9–11**
1.4	**Carbon source for the sink tissues**	**11–14**
1.5	**The controlling elements of seed oil biosynthesis**	**14–19**
1.6	**The role of metabolomics and proteomics to understand the regulatory pathways and its association with storage products**	**19–22**
1.7	**Importance of biofuel production from prominent high oil yielding seeds as biofuels in India**	**22–28**
1.8	**Non-feed tree legume *P. pinnata* as a potential biofuel feedstock**	**28–32**

CHAPTER 2

MATERIALS AND METHODS

2.1	**Maintenance of plants and the field conditions**	**34–35**
2.2	**Stomatal characteristics**	**36**
2.3	**Measurement of photosynthesis rate in *P. pinnata*.**	**37**
2.4	**Biochemical analysis of *P. pinnata* leaves**	**37–38**
2.5	**Quantification of free amino acids using HPLC**	**39**
2.6	**Label free quantification (LFQ) of foliar proteins**	**40–41**
2.7	**Collection of seed samples, storage conditions and experimental planning**	**41–43**
2.8	**Measurement of seed dimensions, fresh and dry weights**	**44**
2.9	**Estimation of seed pigment**	**44**
2.10	**Estimation of seed storage compounds**	**44–45**
2.11	**Microscopic studies of *P. pinnata* seeds**	**45**
2.12	**Sugar extraction and estimation through HPLC**	**45–46**

Table of Contents

CONTENTS	PAGES
2.13 Enzyme extraction	**46**
2.14 Invertase assay	**46**
2.15 Sucrose synthase and sucrose–phosphate synthase assay	**47**
2.16 Fructokinase and hexokinase assay	**48**
2.17 Metabolite profiling	**48–49**
2.18 Thin layer chromatography for lipid profiles	**50**
2.19 2–D Protein profiling and identification using MALDI TOF MS/MS analysis	**51–52**
2.20 Gene expression analysis by real time PCR	**52**
2.21 Statistical analysis	**53**
CHAPTER 3	
RESULTS	
3.1 Stomatal characteristics and the photosynthetic gas exchange parameters	**55–56**
3.2 Foliar biochemistry	**56–58**
3.3 Identification of leaf proteins using nano–LC–MSMS	**58–62**
3.4 *P. pinnata* pods and morpho–physiology of seeds	**63–66**
3.5 Pigment content and major storage compounds	**66–69**
3.6 Anatomy of the cotyledonary sections and identification of storage granules	**69–73**
3.7 Major soluble sugars of developing seeds	**74–76**
3.8 The stage wise changes in the concentrations of invertases at different pH	**77–79**
3.9 Activities of sucrose synthase, sucrose phosphate synthase and hexokinases in *P. pinnata* seeds	**80–83**
3.10 Complete metabolome analysis of developmental stages of *P. pinnata* seed	**84–88**

Table of Contents

CONTENTS		PAGES
3.10	Total lipid and fatty acid profile	89–91
3.11	2D–proteomic profile of developing *P. pinnata* seed	92–101
CHAPTER 4		
DISCUSSION		
4.1	The vegetative growth physiology and photosynthetic characteristics of *P. pinnata*	103–104
4.2	An overview on changes in the characteristic of developing *P. pinnata* seeds	105–110
4.3	Fate of sucrose to lipids and other storage compounds while picking the favorite hexose	110–115
4.4	Dynamic changes in primary metabolites with respect to different pathways inside the developing seeds.	116–118
4.5	Quality of *P. pinnata* seed oil at maturation and the end of the long spanned developmental phase.	119–120
4.6	Few proteins are identified which help in undermining the pathways related to storage production in *P. pinnata*.	121–123
CHAPTER 5		
SUMMARY AND CONCLUSIONS		125–129
CHAPTER 5		
LITERATURE CITED / REFERENCES		131–156
PUBLICATIONS		

CHAPTER 1

INTRODUCTION

1.1 A short note on highly potential biofuel legume tree *P. pinnata*.

Tree legume, *P. pinnata* has been named as an “untapped resource” to “versatile”, “multipurpose”, “magical *Pongamia pinnata*” etc by the various available research till date (Scott et al., 2008; Sangwan et al., 2010; Patwardhan 2021). However, the traditional uses have been well known for a long time and utilized in many civilizations from ages (Sangwan et al., 2010). If we look through a decade of research on *P. pinnata*, there is a rapid progress on multiple areas which include new possibilities with future interest and advances (Table 1.1). Due to its multifaceted nature, various studies are available mostly in the field of biofuel to medicine and pharmacology, salinity and phytoremediation also (Marriboina et al., 2017; Degani et al., 2022; Rajarajan et al., 2022). All these features make *P. pinnata* a suitable for utilizing harsh barren lands while also benefiting the production of future energy resources. However, being a wild tree and distributed across various locations around the globe, the establishment of proper germplasm data is required for *P. pinnata*. Studies have been carried out using various genetic markers such as RAPD, SSR, AFLP and other molecular markers to frame a well–defined genetic identity (Biswas et al., 2013). The growing rapid interest with this tree legume is mainly because of the potentiality of its seeds for usage as biofuel which is characterized with the emergence of revolutionary advances in renewable energy research (Scott et al., 2008). It is also because of the appropriate composition of *P. pinnata* seed oil to be a biodiesel feedstock which captures and demands a sheer knowledge to focus. Detailing the mechanism of seed development and unravelling the pathways behind those regulating the production of seed oil biosynthesis would be a milestone in the current available research data.

Focused areas	Features/Outcomes		References
Sequencing and cloning	• Chloroplast and mitochondrial genome sequencing • Circadian clock gene identification • Whole transcriptome sequencing	• Draft genome sequencing • Seed transcript analysis during development • Cloning and characterization of desaturases	*Ramesh et al., 2014; Winarto et al., 2015; Aadi Moolam et al., 2016; Huang et al., 2016 and 2018; Sreeharsha et al., 2016 and 2022*
Genetic markers	• SSR • RADP • Chloroplast and mitochondrial markers • Circadian clock genes	• AFLP • EST SSR • ISSR • TE–AFLP	*Jiang et al., 2012; Biswas 2013; Rajaranjan et al., 2022;*
In–vitro regeneration	• Shoot bud induction/re–generation • Cotyledonary organogenesis	• Micropropagation and genome analysis	*Kesari et al., 2012; Nagar et al., 2015; Singh et al., 2016; Tan et al., 2018*
As therapeutic and pharmaceutical applications	• Natural therapeutics for neurodegenerative diseases • Wound healing • Leave extract with gold and silver nanoparticles	• Cardioprotective activity from leaf extract • Anti–inflammatory activity from seed pods • Antipyretic uses	*Shilpi et al., 2012; Li et al., 2015*
Biofuel	• Comparison of corrosive behavior • Biosorption of zinc ions • Long term storage and oxidation stability	• Cl engine performance • Thermal degradation studies • Enzymatic transesterification	*Khayoon et al., 2012; Ortiz–Martínez et al., 2016; Jain 2019; Kumar and Pal et al., 2021*
Stress Tolerance	• High Salinity tolerant • Sodium ion sequestration and root membrane permeability	• Salinity induced gene expression studies • Phytoremediation of arsenic and nickel • Cultivation barren and marshy land	*Arpiwi et al., 2013; Kumar et al., 2017; Yu X et al., 2021; Shoaib et al., 2021*

Table 1.1: Overview of a decade of research (2012 – 2022) on P. pinnata in the relevant focused areas of interest with features and outcomes.

Available knowledge on the underlying mechanism of legume seed development also needs to look through properly in order to carry forward with the required objectives in the current study.

1.2 Typical view on seed development of legumes – A brief History.

Legumes are widely distributed and domesticated since millennia across all of the diverse soil and climatic conditions. Apart from the nitrogen fixation, most of them are known for the source of nutritional importance, medicinal properties, fodders and other important sources which is targeted for various uses in agronomy as well as other industries (Table 1.2). Diversity of legumes is among the largest of its kind which consist of various species, sub–species including different germplasms or accessions. They include small herbs, shrubs to large trees (Graham et al., 2003; Tharanathan et al., 2003). With the wide scope of uses, legume seeds are diverse, mostly dicotyledonous and share a characterized process of embryogenesis and seed development (Weber et al., 2005; Catusse et al., 2012). Embryogenesis is a tightly organized and coordinated process to create the unique morphological structure which carries many features and characteristics of a given species (de Vries and Weijers 2017). The life cycle of sporophytic angiosperm cannot be completed without a controlled regulation of the complex embryogenesis which starts from the formation of zygote to a completely developed seed (Goldberg et al., 1994). Pattern formation and cyto–differentiation are two main events controlling the process of embryogenesis in both monocots and dicots (Figure 1.1). The stages of embryo development in legumes usually starts from the post fertilization followed by formation of globular to heart shaped structures to cell expansion and maturation (Goldberg et al., 1994) (Figure 1.2).

Legume species	Uses (Food/ Fodder/ Timber/Others)	Seed characteristics	References
Herbs/shrubs			
Pea/ *Pisum sativum L.*	• Food	• Small; Round	*Cousin 1997*
Chickpea/ *Cicer* spp.	• Food/Fodder	• Small/Medium/ Large; Round	*Singh et al., 2004*
Faba bean/ *Vicia faba L.*	• Food/Fodder	• Medium/Large; Oval	*De Cillis et al., 2019*
Cowpea/ *Vigna unguiculata*	• Food/Fodder	• Small; Oval/ Kidney	*Gomes et al., 2021*
Lentil/ *Lens* spp.	• Food	• Small; Round	*Tullu et al., 2001*
Pigeon Pea/ *Cajanus cajan (L.)*	• Food	• Small/Medium; Round/Lens	*Sameer Kumar et al., 2017*
Peanut/ *Arachis hypogaea L.*	• Food	• Small/Medium; Oval	*Akhtar et al., 2014*
Mungbean/ *Vigna radiata*	• Food	• Small; Oblong/Oval	*Herath et al., 2018*
Horsegram/ *Macrotyloma uniflorum*	• Food	• Small; Round	*Chauhan et al. 2009*
Trees			
Quickstick/ *Gliricidia sepium*	• Fodder/Timber	• Medium; Round	*Elevitch and Francis 2006*
"Khair"/ *Acacias*	• Food/Fodder	• Medium; Oval	*Sivakumar et al., 2013*
Golden shower tree/ *Cassia fistula*	• Fodder/Medicine	• Medium; Round	*Rodrigues–Junior et al., 2020*
Mesquites/ *Prosopis spp.*	• Food/Fodder/Timber	• Medium/Large; Oval	*Patnaik et al., 2017*
Silk plants/ *Albizia spp.*	• Forage/Timber/ Medicine	• Small/Medium; Oval	*Khurana and Singh 2000*
Tree bean/ *Parkia roxburghi*	• Food/Forage	• Medium/Large; Round	*Singha et al., 2021*
Rosewood/ *Dalbergia sissoo*	• Forage/Timber	• Medium; Oblong	*Vakshasya et al., 1992*
Derris spp.	• Medicine/Others	• Medium/Large; Oval	*Fernando 2012*
Pongam tree/ *Pongamia pinnata*	• Medicine/Fodder/ Others	• Medium/Large; Oval	*Scott et al., 2008*

Table 1.2: *Uses and seed characteristics of few important legumes.*

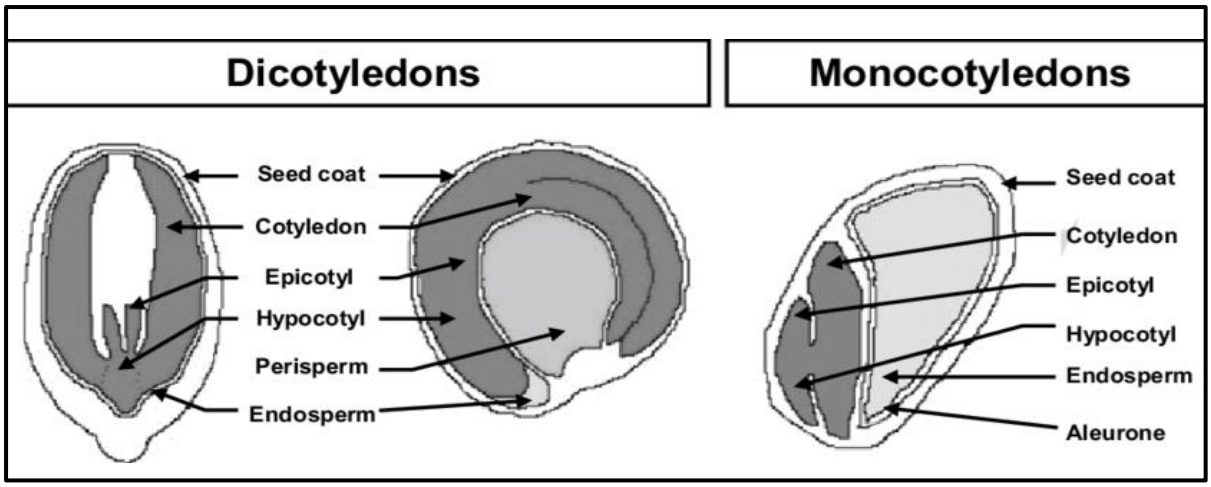


Figure 1.1: *Difference in structure of dicot and monocot seeds. Image courtesy (Catusse* et al., *2012).*

During early stages itself, differentiation of the seed compartments is programmed to form seed coat, embryo and endosperms (Murray 1987; Weber et al., 1998). In legumes, endosperm is absorbed by the developing embryos which are transformed to storage rich cotyledons (Figure 1.2) (Goldberg et al., 1994). Immediately after fertilization, the divided zygote will give rise to a tiny apical cell which develops into two different tissues called embryo proper and the suspensor (a large basal cell). The suspensor differentiates to give support and nourishment to the embryo proper but later on degenerates itself with due course of embryogenesis (Figure 1.2) (Goldberg et al., 1994; Le Brandon et al., 2007). The well maintained and nourished embryo proper later on differentiates and represents the new sporophytic generation consisting of meristematic tissues responsible for generating root/shoot systems (Figure 1.2) (Goldberg et al., 1994; Laux et al., 2004). Seed formation completes with the combined efforts from maternal and filial parts.

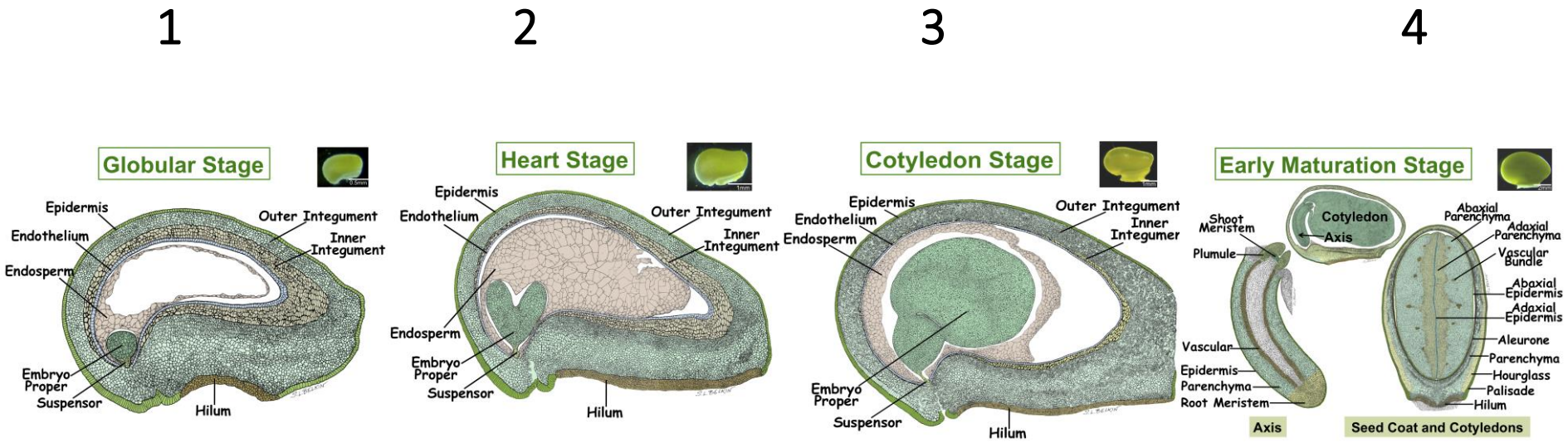


Figure 1.2: *The detail diagrammatic representation of developing embryos in soybean seeds. Stages shown are 1. Globular stage; 2. Heart stage; 3. Cotyledonary stage; 4. Early maturation stage. Embryo proper and suspensor are enclosed by the endosperm, endothelium and epidermis respectively. Outer and inner integument are also classified as shown which provides a protective coating surrounding the developing seeds. Adopted from Goldberg Lab 2015.*

In legumes, the developing cotyledonary tissues are the main storage reserves rich in protein, starch or lipids, enclosed by well–formed and protective seed coat, which is maternal in origin (Weber et al., 1998). The wholly enclosed seed with the cotyledons and seed coat inside the developing pod are attached to the funicle of pod wall which again attached with the main branches with the help of peduncle. Due to their versatile nature and great economic importance, legumes are exploited for advanced research and databases. Many legume species of human interest have yielded a well–known repository of all the datasets such as information on genes, transcripts, protein etc (Le et al., 2007). However, it is a recent trend to focus on the minute events leading to the regulation of storage biosynthesis with special interest on lipids (Song et al., 2017; Shang et al., 2020). The prior knowledge on legume seed morphology, physiology and genetic regulations are studied in various model species which are well exploited to an extent to carry forward an advance approach of metabolic engineering (Verdier and Thompson 2008; Thompson et al., 2009; Boelt et al., 2015). In order to achieve the goal, emerging technologies such as genomics and transcriptomics have been used to study the overall process of seed development in model soybean plant where a stage wise expression of genes from the very beginning of embryogenesis to senescence were analyzed (Le et al., 2007). According to the study, the genetic regulation of protein and lipid biosynthesis starts from the early maturation and continues till the seed loses the pigment when they reached late maturation phase (Le et al., 2007). Identification of those expressed proteins through various approaches of proteomics is currently taking up the pace along with the other “omics” to understand the above processes (Parreira et al., 2016). Additionally, the mining of metabolites through the available methods of metabolomics and assembling those spatially during the entire process

of seed development is also an ongoing approach (Wu et al., 2021). The biosynthesis of these storage products is an integral part of the maturation process, also it is essential for propagation towards their next generation (Weber et al., 2005). Therefore, it is necessary to unravel the events occurring in each phase of embryogenesis till maturation, which lead to the accumulation of proteins, lipids and other storage products.

1.3 The storage components of developing legume seeds.

Process of embryogenesis and seed development is accomplished by seed filling where the storage products are made and distributed across the developing cotyledons (Golombek et al., 2001; Borisjuk et al., 2003; Gallardo et al., 2008). There is a plethora of available literature on storage products of legume seeds wherein carbohydrates, proteins and lipids are accumulated as “starch granules”, “protein bodies/storage vacuoles” and “oil bodies” respectively (Gallardo et al., 2008). The process of accumulation of these storage products also requires coordinated molecular process, involving subcellular targeting, packaging by Golgi body giving the respective shapes (Hills 2004). Recent advances in technologies such as transmission electron microscope (TEM), fluorescence dyes etc. have made the visualization of the storage granules and studies on their biogenesis is much rapid and productive (Melkus et al., 2009; Zhao et al., 2016). There are also spatial and temporal control points inside the cotyledons which govern storage and lipid accumulation (Borisjuk et al., 2005). Amount of accumulated reserve materials and formation of these granules in the seeds decide their uses and benefits for different prospects. Following are the details of storage products observed across different legumes species:

❖ **Protein bodies:** The most common proteinous legumes are widely used in production of food grains and other nutritional supplements (Duranti 2006). Depending on the types of amino acids synthesized and accumulated in the storage protein bodies of matured seeds, it becomes rich in nitrogen, sulphur or other essential nutrients. Moreover, it has also been found that the special proteins such as legumin, vicillin and globulins may act in assembling and formation of protein bodies inside the legume seeds (Shewry et al., 1995; Tang et al., 2011; Le Signor et al., 2017). They vary in sizes ranging from 0.5 µm to 25 µm in diameter in grain legumes to other tree legumes (Shevkani et al., 2019).

❖ **Starch granules:** Apart from cereals, legumes are also very rich source of carbohydrates, which is nutritionally important and very common in most of the grain species (Wang et al., 2003). The chemistry of legume seed carbohydrate has amylose and amylopectin as their main component to form starch granules in developing seeds (Gallant et al., 1992). The mechanism of starch granule formation starts with the import from source to amyloplast, (Yu et al., 2021). Studies on starch accumulation in most of the grain legumes showed variation in size and structure. The shape resembles oval/round which varies from 6.0–60.0 µm in most of the species (Wani et al., 2016). Grams, peas and faba beans are some of most commonly known legumes rich in carbohydrates as their storage products.

❖ **Oil bodies:** Oil yielding legumes are the recent areas of key research because of demand, availability and efficiency (Bates et al., 2013). Ground nut, Soybean and *P. pinnata* seed oil are most exploited oil legumes in terms of food as well fuel uses. Hence, research on oil biosynthesis is emerging in legumes where efforts are also being made to understand their role in biofuel production and other industrial applications (Song et al., 2017). The formation

of oil bodies present in the seeds of legumes and other crop species begins with the biosynthesis of fatty acids from the fatty acid synthase (FAS) pathway with the action of a much popular enzyme accase (acetyl–coA–carboxylase) (Tong and Harwood 2006). The fatty acids synthesized may again undergo rounds of unsaturation with the help of "desaturases" which then follows with the formation of acyl chains with glycerides leading to the final production of Tri–acyl–glycerides (TAGs) (Hills et al., 2004; Bates et al., 2013). The TAGs will constitute to form a micellar structure surrounded by special low molecular weight proteins commonly known as "oleosins". The mechanism of interaction between the TAGs and the associated proteins such as oleosins are still yet to be captured properly in seed oil body formation (Chen et al., 2018). As far as the morphology of oil bodies in lipid rich plant sources are of concern, they are well characterized and have been successfully studied in oil yielding legumes where the sizes are round and ranges from 40nm to 390nm (Singha et al., 2019; Shang et al., 2020).

1.4 Carbon source for the sink tissues.

Determining the source of carbon skeleton in the storage products of legumes seeds is an area of concern in order to unfold the product–precursor relationships in oil body or other storage body formation. The associated maternal tissues and vascular elements with the developing cotyledons provide the full supply of carbon sources in the form of sucrose, and also others such as peptides, soluble carbohydrates or oligosaccharides etc. (Weber et al., 2005; Aguirre et al., 2018) (Figure 1.3). The loading and unloading of the nutrients are coordinately controlled by signaling elements and contribute to formation of the storage products. For the transportation of the sucrose to the filial tissues, there are special transporters performing the coordinated functions to supply the contents to the developing embryos.

One of the reported factors controlling the supply of nutrients in the sink tissues are SUT (Sucrose Transporters) (Figure 1.3). For the transportation of the sucrose to the filial tissues, there are special transporters performing the coordinated functions to supply the contents to the developing embryos. The reported factors controlling the supply of nutrients in the sink tissues are SUT (Sucrose Transporters), MST (Monosaccharide Transporters) and SWEET (Sugar Will Eventually be Exported Transporters) gene families in legumes and other *Fabaceae* members (Doidy et al., 2019). It has been observed that sucrose is the most abundant sugar in legume sink tissues (Obendorf and Górecki 2012). After the successful transport of sucrose to the developing fruits, there are catabolizing enzymes which take turns to form hexoses. Two important enzymes observed in legumes are the invertases and the sucrose synthase (Weber et al., 2005; Wang et al., 2013). Invertase and sucrose synthase cleave sucrose to glucose + fructose and fructose + UDP–glucose respectively (Moriguchi et al., 1992; Weber et al., 1995). Moreover, the invertases of seeds are classified mainly into two types 1) acid invertases; 2) neutral/alkaline invertases. Acid invertase is active at pH $\leq$ 4 and generally localized in the vacuoles (soluble) or cell wall (insoluble) whereas neutral/alkaline invertases are active at pH between 7 to 7.5 (Fotopoulos et al., 2005). Sucrose synthase is another enzyme which mediates a reversible conversion of sucrose to hexoses, where the conversion of hexoses to sucrose can be again achieved with the action of sucrose phosphate synthase (Moriguchi et al., 1992; Weber et al., 1997). This controlled regulation of sucrose metabolism is crucial for maintaining cell division, cell expansion and storage biosynthesis and the overall process of seed development (Figure 1.3; Granot et al., 2014; Stein and Granot 2018).

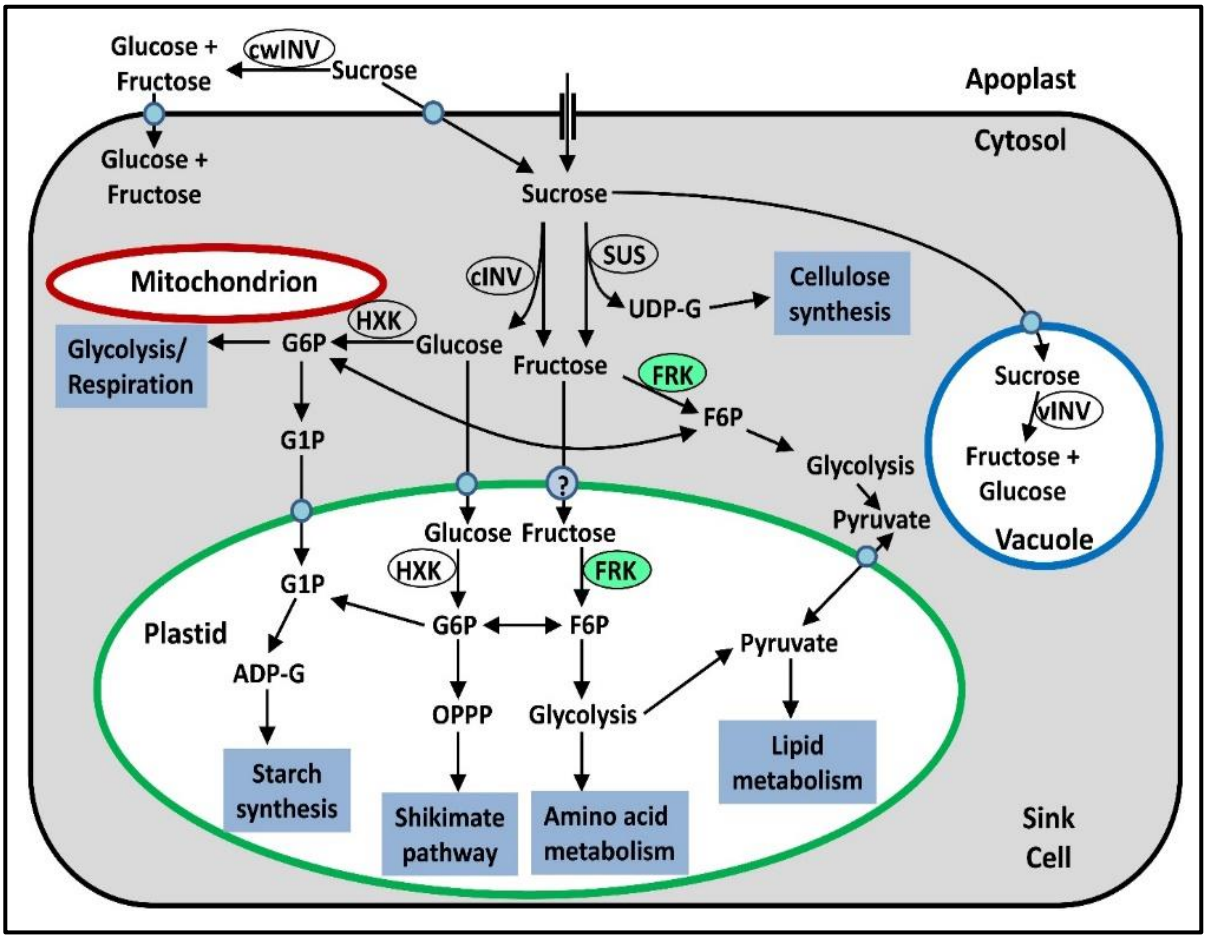


Figure 1.3: *Schematic representation of the fate of sucrose towards sink tissues and the coordinated subcellular network of seed development. When sucrose reaches the developing seeds, it has been successfully metabolized with the help of invertases inside and outside of the sink tissues. Sucrose synthase also take part in breaking of sucrose to hexoses. The hexoses are metabolized with the help of respective hexokinases present inside the developing seeds where they are channelized through different metabolisms in order to initiate the storage product biosynthesis. Stein and Granot 2018.*

Based on the plant species, the preference of the hexose sugar varies, where further breakdown by the hexokinases is carried forward to the other metabolic processes (Granot et al., 2014). Glucose is cleaved preferably by HXK to form glucose–6– phosphate where fructose is the target of FRK to form fructose–6–phosphate (Renz and Stitt 1993). The control of these enzymes is the expression of their gene families which take turns during the overall process of plant development (Granot et al., 2014; Aguilera–Alvarado et al., 2019). These metabolizing enzymes of sucrose and other sugars in developing seeds not only control the storage biosynthesis but also the molecular regulation of programmed events maintaining the developmental phases (Weber et al., 2005; Zhang et al., 2018).

1.5 The controlling elements of seed oil biosynthesis.

In general, the seed development bears a complex regulatory network of metabolic and developmental machinery which include signaling elements associated with accumulation of storage compounds (Baud et al., 2008; Atabani et al., 2013; Sreeharsha et al., 2016). Further, the seed storage-compound synthesis overlays the developmental progression of embryogenesis and is to an extent governed by the metabolite and hormonal signals inside the embryo (Borisjuk et al., 2003). The proportional distribution of seed storage products into carbohydrates, oils and proteins will depend on the influx of metabolic pathways during the development and can vary between species (Schiltz et al., 2005; Ekman et al., 2008; Chaitanya et al., 2015). The two major steps of TAG biosynthesis are de novo fatty acid biosynthesis in the plastids and TAG assembly in the endoplasmic reticulum. There are several controlling elements yet to be known and studied thoroughly, few of the key elements controlling the oil biosynthesis pathway are listed here:

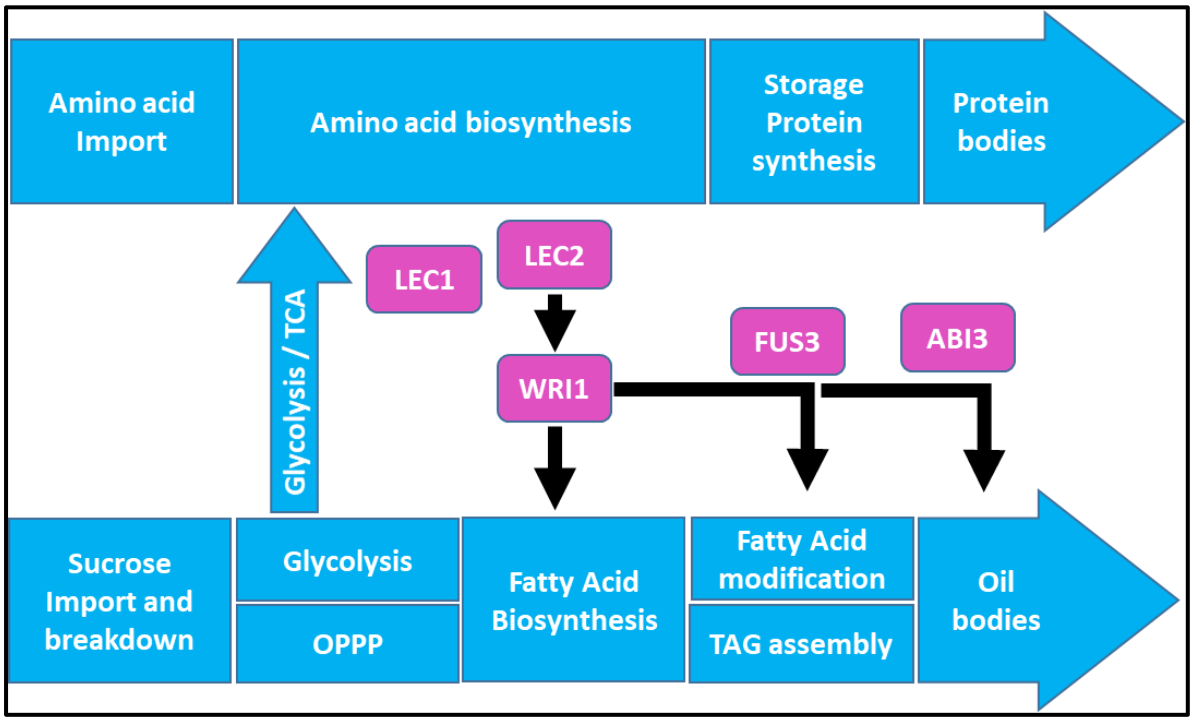


Figure 1.4: *Molecular control of storage products mainly the lipids in developing seeds. The precursors for de novo fatty acid synthesis in maturing embryos are derived from sucrose through the glycolytic pathway and/or the OPPP. Fatty acids produced in the plastids are then exported towards the cytosol in the form of acyl CoAs and used to form triacylglycerols which ultimately get stored inside the oil bodies. The amino acids required for the synthesis of storage proteins during the maturation process are either directly imported from the maternal tissues or synthesized/modified in the embryonic tissues. Storage proteins are ultimately stored in specific vacuoles. Solid arrows represent positive transcriptional regulations. The key controlling elements of seed oil biosynthesis such as LEC ½, WRI1, FUS 3, ABI 3 are highlighted in the purple boxes. Adapted and modified from Santos-Mendoza et al., 2008, Jiang et al., 2012.*

❖ **Sucrose:** Apart from being the main sugar source, sucrose is also involved in various developmental processes (Wobus and Weber 1999; Yoon et al., 2021). For oil development, carbon is delivered to fatty acid synthesis in seed plastids via glycolysis with hexose and triose as the predominant carbohydrates entering the plastid. Previous reports signified the role of sucrose import into embryo and its metabolism in the cytosol and plastids in the formation of starch and oil (Luthra et al., 1991; Eastmond and Rawsthorne, 2000). It is well established in several model legumes that the dynamic changes of hexose/sucrose ratios play a key role in the commencement of storage product synthesis as the seed development progressed. On the other hand, the primary metabolism inside the seed which includes major pathways of glycolysis and TCA link the biosynthesis of carbohydrates, amino acids and fatty acids which also provides energy balance inside the developing seeds (Rolletschek et al., 2005; Schwender et al., 2015).

❖ **Phytohormones:** The crosstalk between plant hormones such as auxin, cytokinin, ABA and ethylene is an essential criterion to maintain the developing processes such as the cell division, differentiation, maturation and senescence etc (McAtee et al., 2013; Sanz et al., 2015). The plant hormone ABA is known to have a significant role in regulating the accumulation of storage products for the developing embryo in many plants (Finkelstein et al., 2002; Ali et al., 2022). The regulatory functions of ABA in maintaining the flowering to seed development and maturation to germination involve a sequential regulation of the hormonal signals (Ali et al., 2022). The combined role of IAA and ABA's role in the production of unsaturated fatty acids in microalgae have unfolded new insights for understanding the phytohormone activities (Lin et al., 2022). However, the regulation in higher plants is still on the way to uncover and illustrated.

❖ **Transcription factors:** Transcription factors are the starting switch for controlling specific genetic programs. The transcriptional regulation of seed oil biosynthesis has unfolded the role of WRINKLED1 (WRI1), which is widely known across the high lipid rich crops (Wang et al., 2007; Kong et al., 2020). Other known factors till date include a transcription factor of the miRNAs, APETELA EAFY COTYLEDON1 (LEC1), LEAFY COTYLEDON2 (LEC2), FUSCA3 (FUS3), ABSCISIC ACID3 (ABI3), ABSCISIC ACID5 (ABI5) and SPATULA are known to regulate multiple aspects of seed development (Wang et al., 2007; Liu et al., 2017; Kumar et al., 2020). WRI1 is also studied to be a control switch for APETALA2/ethylene responsive element–binding protein (AP2/EREBP) family that plays important role during plant seed oil accumulation (Chen et al., 2018) (Figure 1.4). Other families of transcription factors such as NAC, MYB, DOF, GATA and HD–ZIP have been reported to be actively involved in many processes of seed development which improves the oil accumulation (Rajavel et al., 2021). Reciprocally, the expression of MYB89 represses the seed oil accumulation by inhibiting *wri1* along with other genes (Le et al., 2007). With the evolving technologies of sequencing and molecular characterizations, the orchestration of transcription factors can be well understood in seed developmental process.

❖ **Genetic regulation:** Understanding the legume seed development using genomics has unfolded the stage wise regulation of seed development starting from the early embryogenesis to senescence (Le Brandon et al., 2007; Hudson and Hudson 2021). However, a focus on increasing the seed oil production in various crops are the key concern of many researches till date (Wang et al., 2007; Eskandari et al., 2013; Hudson and Hudson 2021).

Among them, finding the QTLs for lipid metabolism, breeding approaches between high yielding varieties, cross breeding and transgenic methods are some of the commonly known approaches (Hobbs et al., 2004; Weselake et al., 2009). However, modern methods such as the whole genome sequencing, transcriptome sequencing and expression studies enabled the targeting of genes to better understand the mechanism of lipid accumulation in high oil yielding species. Enhancing de–novo biosynthesis of fatty acids by overexpression of genes such as *accase* through transgenics can improve the yield of oil production. Moreover, focus has been also shifted towards the understanding the process of unsaturation of fatty acid and acyl–lipid metabolism of TAGs (Bates et al., 2013). The characterization of desaturases from novel high oil yielding plant species can meet the need of industrial uses of cloning and its products. Another focus is the controlled expression of thioesterases (*fatA*, *fatB*) which cleaves the required fatty acid for the acyl pool and to understand their regulation in model species. Again, the expression of Diacylglycerol O–acyltransferase (*dgat*) and Lysophosphatidic acid acyltransferase (*lpaat*) which work for TAGs pathway and oil body formation are also of great importance (Ohlrogge and Browse 1995; Snyder et al., 2009; Gacek et al., 2017; Miklaszewska et al., 2021) (Figure 1.5). Storage lipids such as oil bodies from the fatty acids with glycerol backbones require a structural modification with the help of small molecular weight proteins (Bates et al., 2013). Those proteins are formed most commonly by genes encoding oleosins (*ole*), followed by caleosins (*clo*) and steroleosin which are the integral part of the oil bodies in seed plants (Chen et al., 2018).

❖ **Seed Photosynthesis:** The photoautotrophic nature of seeds indicates an active chloroplast with the capacity of photosynthesis (Ruuska et al., 2004). Specially in oilseeds such as *Brassica, Arabidopsis* and *Glycine max*, light plays an important role in regulating the biosynthesis of enzymes during the seed development, supporting the active role of seed photosynthesis (Wang et al., 2016; Nwafor et al., 2022). The dynamic shift in the expression patterns of photosystem related proteins along with FA biosynthesis enzymes during the seed development of *Arabidopsis* and *Glycine max* support the active role of seed photosynthesis. Hence, studying the pigments and associated protein expression patterns in large seeds with long developmental cycles will provide more insights into the role of seed photosystems in synthesizing storage compounds. It was reported in certain crop plants that the seed carries out photosynthesis during embryogenesis which provides energy for FA biosynthesis and also helps in re–fixing the respiratory CO_2 release (Schwender et al., 2004). Though the absence of seed photosynthesis will not affect the overall FA biosynthesis, it will be considered as an important factor in promoting carbon storage, energy flux and lipid biosynthesis in oilseeds (Figure 1.5).

1.6 The role of metabolomics and proteomics to understand the regulatory pathways and its association with storage products.

All the biological processes are regulated by various factors, where metabolic regulation of seed oil biosynthesis and storage products is also of the crucial ones. On the other hand, the primary metabolism inside the seed includes major pathways of glycolysis and TCA link the biosynthesis of carbohydrates, amino acids and fatty acids which also provide energy balance inside the developing seeds (Rolletschek et al., 2005; Schwender et al., 2015).

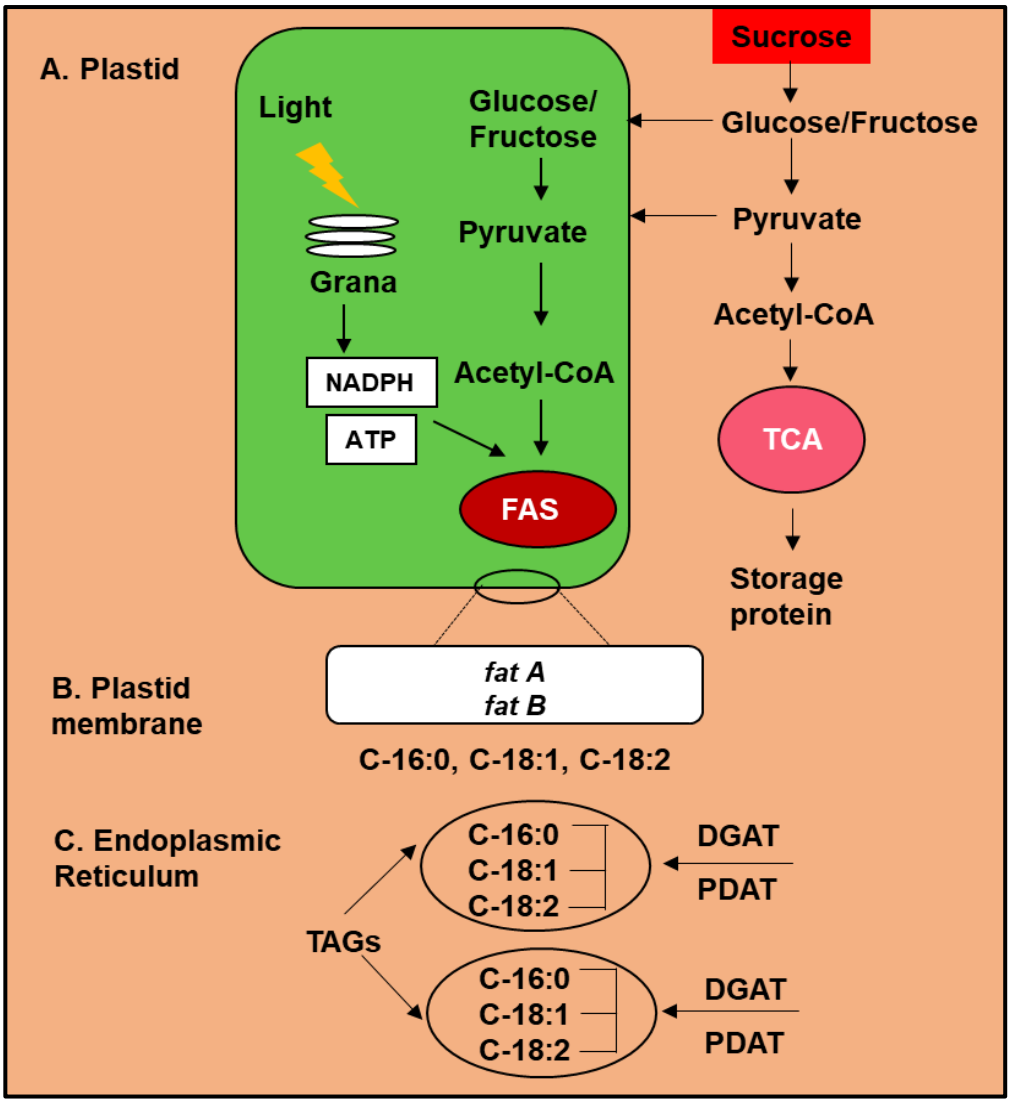


Figure 1.5: *Overview of major reactions involved in fatty acid and triacylglycerol synthesis. (A) Plastid fatty acid synthesis; (B) Fatty acid transport from plasdial membrane and (C) TAG synthesis. Major events occurring inside the plastids including the production of NADPH and ATP taking part in fatty acid synthase pathway. Fatty acids are exported with the thioesterases (fat A/B) towards endoplasmic reticulum where the acyl editing and TAG accumulation takes place with the help of DGAT and PDAT. Inspired and modified from Bates* et al., *2013.*

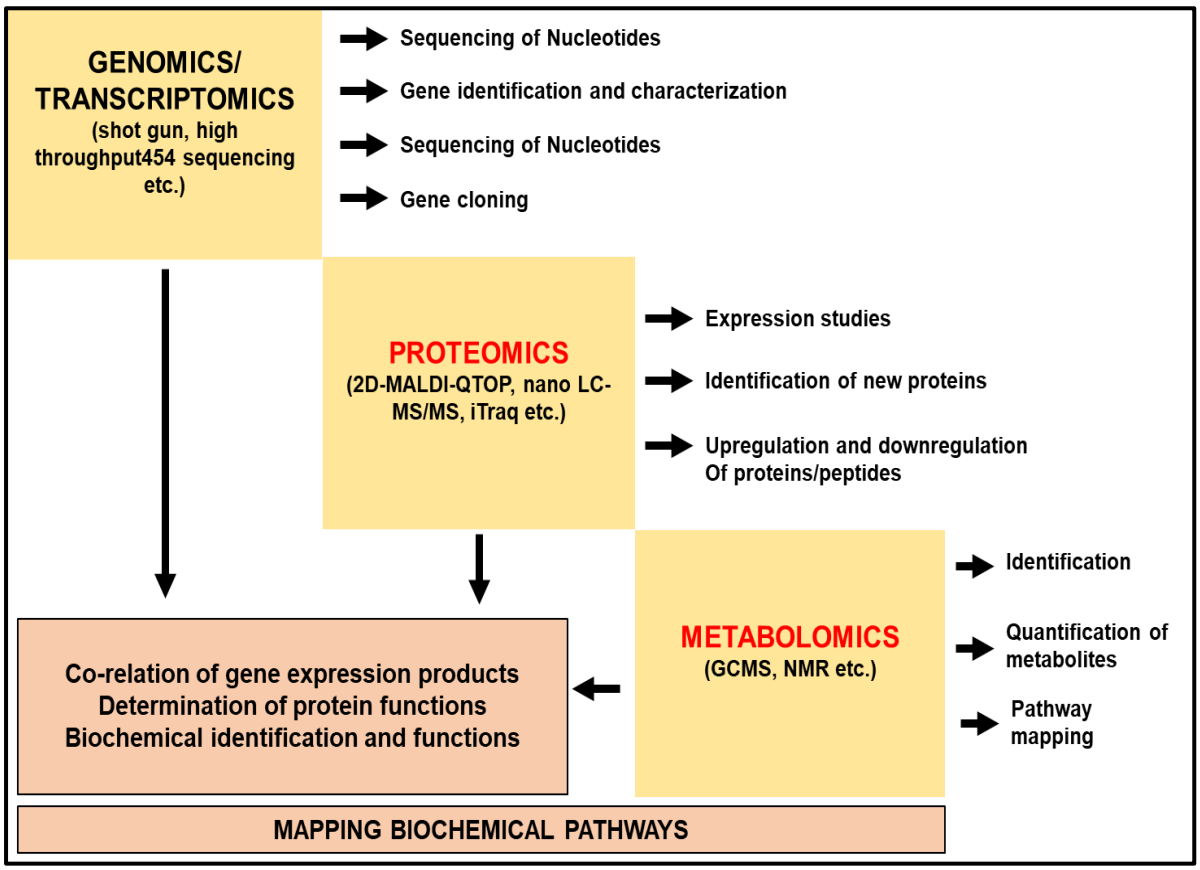


Figure 1.6: *The omics approaches to unfold metabolic regulation. Large–scale studies in the field of omics have been successfully exploring the differences in gene expression, protein and metabolite abundance and modification of post–translational protein, and providing a different level of understanding the cellular processes. Modified from (Gahlaut et al., 2013).*

The omics strategy of understanding the primary metabolism of seed oil research has gone beyond the genomics and transcriptomics, where metabolomics and proteomics are emerging techniques to contribute to the overall pathway engineering for yield improvement (Figure 1.6) (Teh et al., 2017; Wang et al., 2019). It can show the expression and abundance of metabolites and proteins periodically based on the known pathways. For example, a proteomic approach to understand the seed filling metabolism in model legume *Medicago truncaluta* revealed expression and accumulation of enzymes of differentially regulated in developmental and other metabolic pathways (Gallardo et al., 2003). These integrated approaches have importance in targeting the pathways related to seed oil biosynthesis and related metabolisms. Primary pathways leading to oil accumulation after the breakdown of sucrose and hexoses that start from glycolysis, pentose phosphate pathway to TCA and fatty acyl coA can be analyzed based on metabolite concentrations (Tohge et al., 2015). Similarly, proteomics also gives insights into the oil body proteins associated with the TAGs accumulation (Shao et al., 2019). Hence, a precise knowledge on regulations of these crucial metabolites will provide a broader understanding into the process of lengthy seed development associated with oil composition (Wishart et al., 2011). Advancing instrumentation along with the rapid data analysis platforms also make these analyses possible for mapping the metabolic pathways, finding biomarkers, target and untargeted analysis. (Monteiro et al., 2013; Schrimpe–Rutledge et al., 2016).

1.7 Importance of biofuel production from prominent high oil yielding seeds as biofuels in India.

Energy crisis is imminent and search for the efficient source is a challenging goal till now. The continuous increase in world's population, rapid industrialization and urbanization creates an evident situation to tackle those. The demand for sustainable fuel alternatives can be seen as emerging, popular and their implementations are profound in many regions around the globe (Rodionova et al., 2017). Question is, what are those and where does the biofuel/biodiesel stay in terms of their efficacy and suitability in overall prospects? The available sources are classified mainly in 4 generations According to European Academies Science Advisory Council (EASAC). They are: (1) cereals and oil seeds (1^{st} generation), (2) the non–edible plant source and waste materials (2^{nd} generation), (3) Algae, yeast, sea weed etc, (3^{rd} generation), (4) Genetically engineered and enhanced species (4^{th} generation) (Singh et al., 2018). Interestingly, efforts are being made to blend the bio–products with current petro–diesel products which can be suitable for different automobiles based on the compatibility with the respective engines (Demirbas 2008; Shell global 2020) (Figure 1.7). Our work and the following chapters are on the 2^{nd} generation non–edible seed oils from a tree legume species *Pongamia pinnata*. The non–edible oil or other plant parts escapes the food vs fuel emergency but lack of the area for cultivation is a trailing constraint for their proper implementation (Sims et al., 2010). A broader utility of those species is required to meet the degree of efficacy and sustainability. Therefore, other uses such as abiotic stress tolerance, high CO_2 sequestration, insect/pest resistance and allelopathy are also areas to exploit and unravel in those species (Tilman et al., 2006; Abugre et al., 2010; Marriboina et al., 2017).

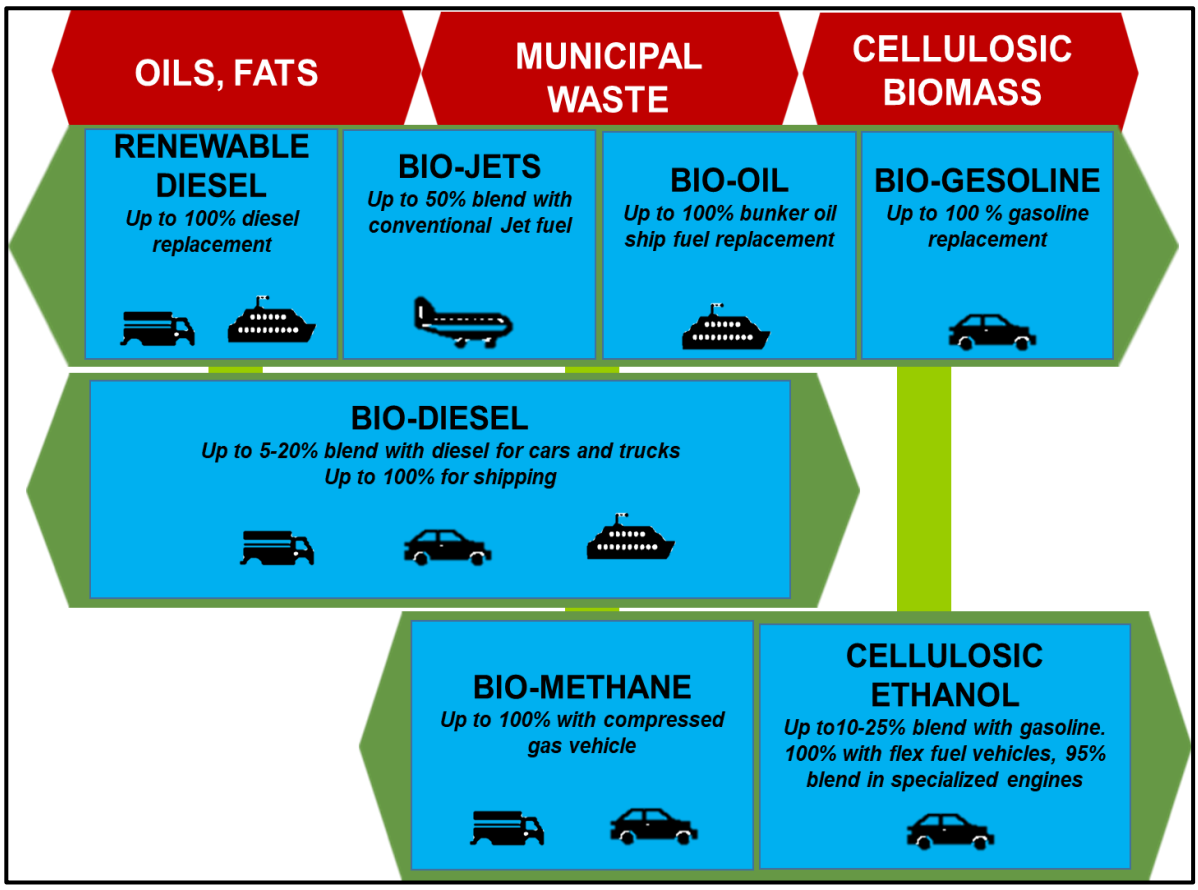


Figure 1.7: *Types of biofuel sources and the by–products which can be used in different automobile industries for replacement of petro–diesel. Adopted and inspired from SHELL Global 2020.*

Above all, the quantity and the quality of oil produced also are the important factors, where presence of saturated and unsaturated fatty acids such as SFA, MUFA, PUFA and others decides the suitability to be a good biofuel source (Tyson et al., 2006; Puhan et al., 2010). It has been observed that high MUFA for example oleic acid (C18:1) in total oil content makes it more suitable as a biofuel source (Cao et al., 2014). Focused research is now on to understand the biological regulations (genetics, biochemistry, physiology etc) of fatty acid production and unsaturation in those species. Seed oil industries have now welded these criteria among the sources and giving efforts on their improvement. Several non–edible, high oil yielding plant species are exploited for understanding the metabolic and molecular regulation of lipid metabolism (Kumar and Sharma 2011). Some of well characterized non–edible tree species grown in many regions of the country are listed along with the seed fatty acid compositions in Table 1.3. Among these second–generation biofuel sources, the highest reported oil content is observed in *Riccinus*, *Simarouba* and *Calophyllum* species where the oil content can reach up to 70 % (Table 1.3). Another important criterion is the fatty acid content and the quality of the seed oil produced which varies widely among those tree species. Therefore, the efficiency of the biofuel also depends majorly on the fatty acid content and also to various other factors. Over time it has been well tested and observed that high content of MUFAs (mono–unsaturated fatty acids) such as oleic acid (C18:1) has increased the efficacy of the biofuel which can be utilized by various automobile industries. Therefore, the profile and percentage of specific fatty acids also plays an essential role in the selection of species as a prominent biofuel source. The composition of fatty acids varies in most of the known and reported biofuel candidates across the country.

Biofuel species	Total lipid content in the seed dry mass	Fatty acid profile and percentage (Round up values)						References
		PA C16:0 (%)	SA C18:0 (%)	OA C18:1 (%)	LA C18:2 (%)	LiA C18:2 (%)	Others (%)	
Madhuca indica **(Mahua Oil)**	**20–30 %**	**25**	**23**	**37**	**14**	**—**	**1.5**	**Kapilan and Reddy 2008**
Jatropha curcas **(Jatropha oil)**	**30–40 %**	**15**	**7.0**	**45**	**33**	**0.2**	**1.0**	**Chaitanya et al., 2015**
Pongamia pinnata **(Pongam/karanj oil)**	**30–40%**	**10**	**7.0**	**51–65**	**17–18**	**4.0**	**—**	**Bala et al., 2011; Singha et al., 2019**
Ricinus communis **(Castor oil)**	**48–60%**	**10**	**10**	**60**	**10**	**7.0**	**89 (RC)**	**Sbihi et al., 2018**
Simarouba glauca	**30–50%**	**19**	**14**	**63**	**3.0**	**—**	**1.0**	**Nayak et al., 2017**
Calophyllum inophyllum **(Polanga oil)**	**50–70%**	**14**	**15**	**38**	**28**	**—**	**5.0**	**Nayak and Mishra 2017**
Camelina sativa **(Linseed oil)**	**30–50%**	**5.0**	**2.0**	**19**	**18**	**55**	**—**	**Nayak et al., 2017**

Table 1.3: *List of prominent non–feed 2nd generation biofuel stock with the lipid and fatty acid compositions. PA– palmitc acid; SA– stearic acid; OA– oleic acid; LA– linoleic acid; LiA– linolenic acid; RC– ricinoleic acid.*

Figure 1.8: *Potential non–edible oil resources as biodiesel feedstock commonly found across India and around the globe (Kumar and Sharma 2011).*

The oleic acid content in *Pongamia*, *Riccinus* and *Calophyllum* reaches up to 60 % or more (Table 1.3). Many species are already considered as potential biofuel seed oils which are spread across the country (Figure 1.8). Their contributions as alternate energy resources are well being investigated while they are also studied with other various added prospects of research (Sharma et al., 2020).

1.8 Non–feed tree legume *P. pinnata* as a potential biofuel feedstock.

The cultivation of *P. pinnata* is suitable in diverse tropical and subtropical environments. The oil content of *P. pinnata* seeds ranges from 35 to 40 % of seed dry weight and 55 – 65 % of the total lipid has been reported as oleic acid which is the ideal fatty acid for biodiesel production (Sreeharsha et al., 2016; Xiong et al., 2018). Due to the accumulation of high levels of poly unsaturated fatty acids in the *P. pinnata* seed oil, it is considered as a potential biofuel feedstock with an optimized efficiency to use as a biodiesel for diesel engines (Singh et al., 2018; Jain et al., 2019). Moreover, not only the *P. pinnata* oil, but also the seed wastes after oil extraction have recently been used to produce bio-ethanol (Muktham et al., 2016). The biochemical characteristics and protein profiling during seed development has been investigated in *P. pinnata* (Kesari and Rangan, 2011; Pavithra et al., 2014). The proximal chemical composition of mature *P. pinnata* seed along with fatty acid composition were also reported earlier (Bala et al., 2011; Sharma et al., 2011; Pavithra et al., 2012). However, in depth analysis of metabolite and proteomic profiles throughout the lengthy developmental period of *P. pinnata* seed have not yet been reported. It is crucial to study the metabolome and proteome of the developing seed to understand the complex association between various metabolites and FA biosynthesis to oil body formation.

The genetic control of *P. pinnata* oil biosynthesis through microRNA sequencing has been reported recently (Jin et al., 2019; Sreeharsha et al., 2022). We have established the in–depth molecular analysis of flowering to seed oil biosynthesis pathway through transcriptome profiling and unraveled the transcriptomic control over these processes (Sreeharsha et al., 2016) (Figure 1.9). We have also analyzed the seed morphology and the oil content in four stages of seed development (Figure 1.9). However, the changes from sugars to oil formation during the ontogeny of *P. pinnata* seed have not been understood till now. Therefore, the current study initiates a detailed study on vegetative growth morphology and physiology by measuring the photosynthetic efficiency during seed development as well as associated mechanisms of oil biosynthesis which were not understood. It is still an unanswered question that, during the prolong phase of seed development how these regulatory mechanisms control the overall maintenance of the seed development. The current information of seed development and oil accumulation of *P. pinnata* is limited in many aspects, such as determination of sources of carbon skeleton for storage products and oil bodies of seeds. The main objective of the current study is to track the passage of carbon sources channelizing towards the seed oil accumulation. In this study, a few crucial enzymes during oil biosynthesis have been also detected and tracking of metabolic pathways leading to storage accumulation is also accomplished. Moreover, the identification of key regulatory proteins in different developing stages of *P. pinnata* seed has given the additional information on further control of the FA accumulation. Following chapters will also reveal unusual photosynthetic behavior of the green developing seeds which have been identified to help in understanding the carbon influx to FA biosynthesis. Our results also provide information on anatomical characteristics of seeds during oil body accumulation.

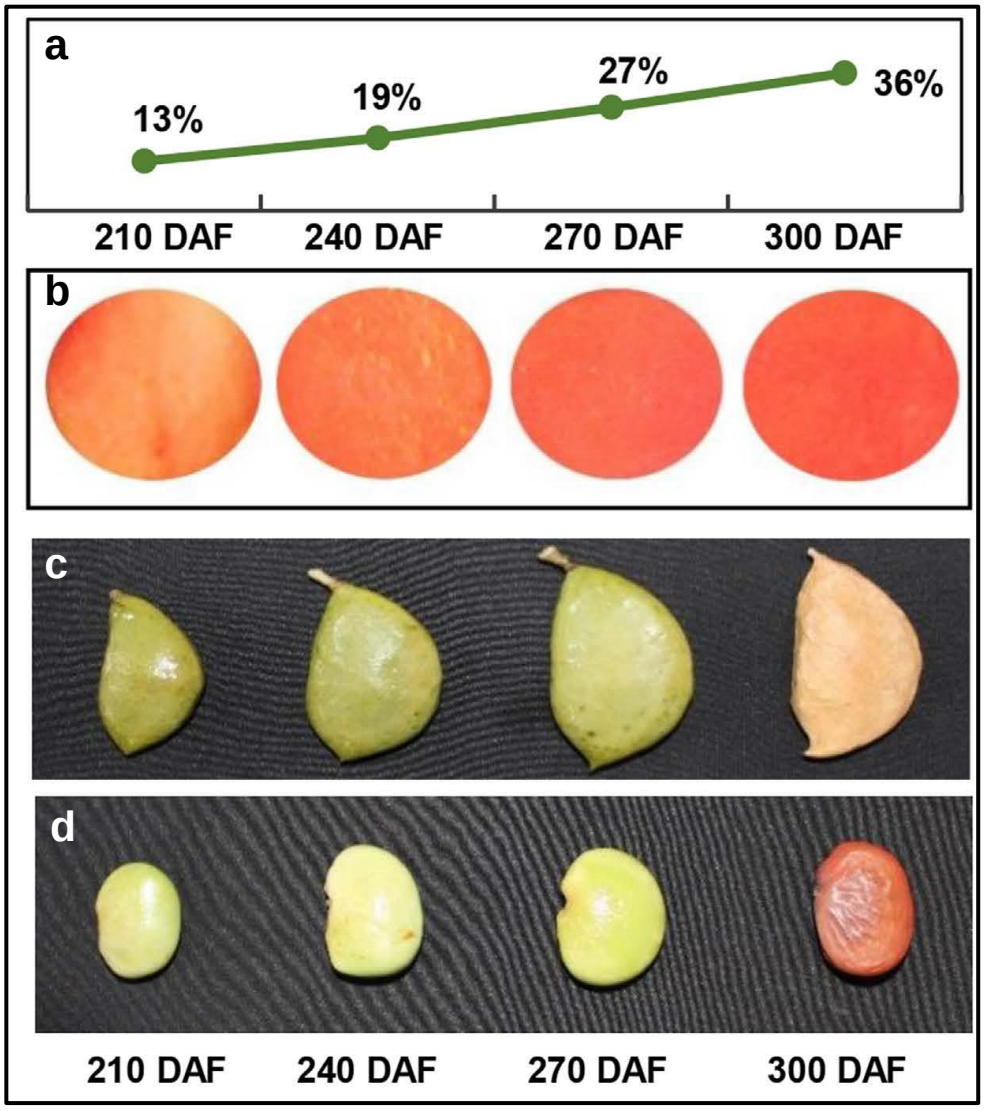


Figure 1.9: *Published data from our previous publications (a) Percentage of oil content during four developmental stages of P. pinnata. (b) Oil-red staining of seed endosperm showing oil accumulation patterns in P. pinnata seed. (c, d) Four developmental stages of P. pinnata pod and seed used in quantification studies of lipid biosynthetic genes. (Sreeharsha et al., 2016).*

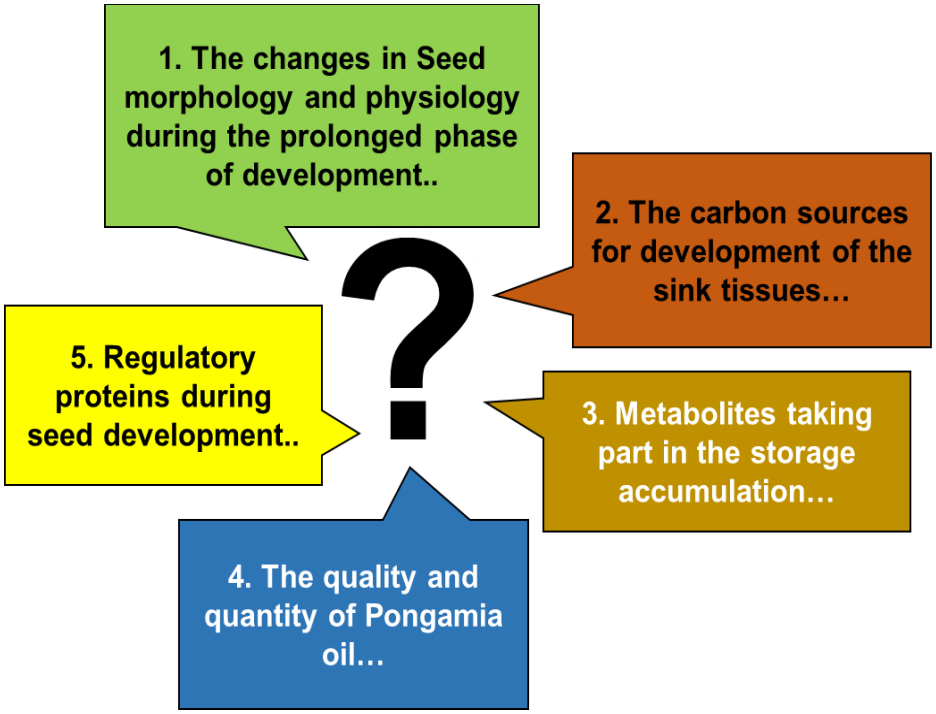


Figure 1.10: *Framing the objectives: understanding the regulation of Seed development in P. pinnata is crucial.*

The outcome of the present findings can answer few queries regarding the control and regulation of seed development in the tree legume species *P. pinnata*. Our results will deepen the knowledge on metabolic regulation of *P. pinnata* seed development and oil accumulation which is important for further metabolic engineering of this potential biofuel feedstock for different industrial and agricultural applications. Therefore, based on the above-mentioned points (Figure 1.10), the objectives have been framed.

The objectives are:

- To study the foliar characteristics ahead of the onset of seed development.
- To elucidate the morphological and physiological changes in seeds during different stages of development.
- To analyze the role of sucrose and hexoses and their counter balance in regulating seed oil biosynthesis.
- Metabolite profiling and analysis in the developing seeds of *P. pinnata*.
- Identification of stage wise regulatory proteins using 2D proteomics.

CHAPTER 2

MATERIALS AND METHODS

2.1 Maintenance of plants and the field conditions.

All the experiments were carried out using plant material from a well–maintained plantations of *P. pinnata* has been established at Tree Oils India Limited (TOIL) Zaheerabad, Medak district, Telangana state, India (latitude 17°36′; longitude 77°31′E; 622 m MSL). Under the natural sunlight trees were planted with uniform 2 m spacing between and 4 m within the rows (Figure 2.1). Plants were maintained with regular watering in summer months, however, there was alternate watering and no watering during monsoon and winter months respectively. A traditional mixture of cow–dung with vermicompost was provided as the farmyard manure for the entire growth seasons. The weather data during our experimentation is shown in Figure 2.2. Ten-year-old *P. pinnata* plantation, established in the experimental farm of Tree Oils India Limited (TOIL) Zaheerabad, Medak district, Telangana state, India (latitude 17°36′; longitude 77°31′E; 622 m MSL) was selected for the present study. The study site has a tropical, hot-steppe agroclimate with the summer months between March to May having maximum temperature up to ~42 ℃ and an average temperature of 22 – 23 ℃ in the winter months of September to February, monsoon starts during June to October with average rainfall ranging from 700 to 1500 mm (Figure 2.2). Plants were maintained under natural photoperiod with uniform 2 m spacing between as well as within the rows. The recorded range of light intensity in the region on a normal sunny day ranged between 1,200–2,000 µmol/m^2/s during 10.00 – 14.00 h solar time. The plants were regularly watered during hot summer months with alternate watering during winter months and no watering during monsoon season. Nitrogen was provided by applying cow dung mixed with vermicompost as farmyard manure at the rate of 12 kg/year twice in equal splits during growth. Trees had a circular canopy with 15–20 tertiary branches.

Figure 2.1*: P. pinnata plantations and germplasms located in TOIL, Zaheerabad, Telangana, India.*

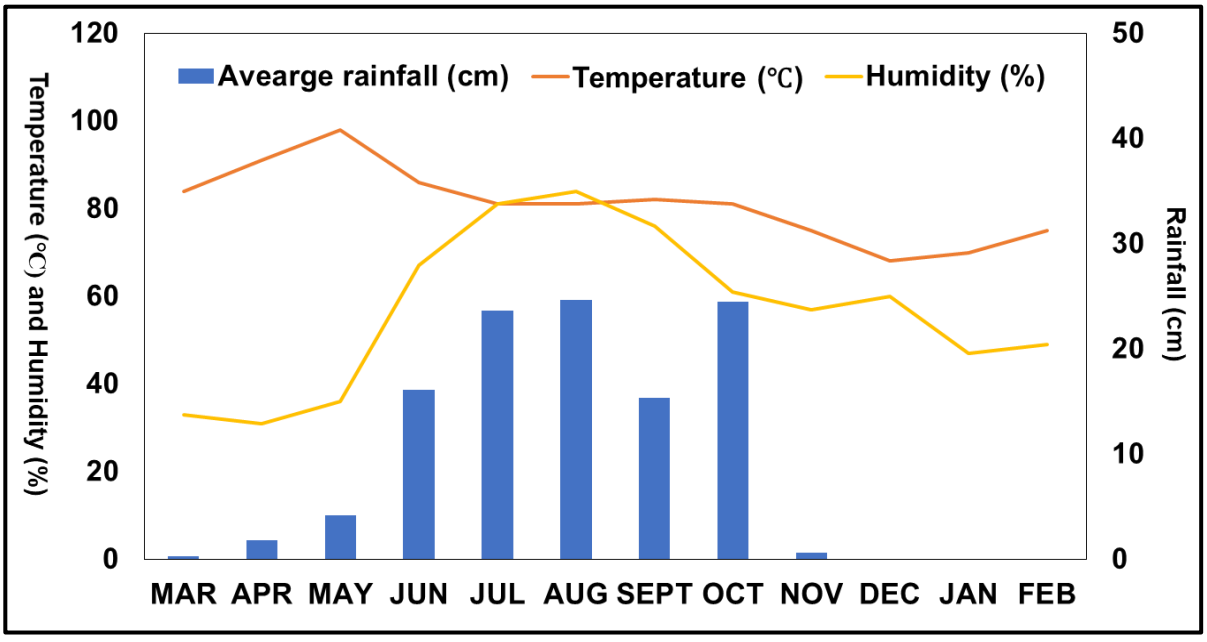


Figure 2.2: *Average weather data. March 2016 – February 2017and March 2017 – February 2018.*

2.2 Stomatal characteristics.

Fresh leaf samples of *P. pinnata* were collected during the morning period (9AM to 12PM). They were washed and fixed in formaldehyde (2.5 %) for 5 to 6 h followed by dehydration in serial concentrations of ethanol (10, 20, 30, 40, 50, 60, and 70 %). Dehydrated samples are then placed on the surface metal stubs (Copper) and the mounted samples were coated with gold in a sputter coater (FC–1100, Japan). Samples were observed under SEM (EDAX; Philips XL 30 ESEM–FEI Instruments, Hillsboro, OR). Stomatal and cell counts were averaged for 10 tree accessions, with three technical replicates. The stomatal density was calculated as number of stomata per square millimeter (mm^{-2}) of the leaf surface area and the stomatal aperture was calculated manually using the image software associated with the instrument.

For fluorescent staining the method described by Tanaka et al (2006) was followed with modifications where the lower epidermal peel of *P. pinnata* leaves were dipped into 10 mM Acridine Orange (Sigma, Merk) in MES buffer of pH 6.5 for about 5 to 10 min at room temperature in dark (covered with aluminum foil). The peels were carefully removed followed by washing 4–5 times with MES (pH 6.5) buffer. After mounting the stained peels on slides with a cover slip, it was observed under confocal microscope (LIECA TCS SP2 AOBS, Heidelberg, Germany) under excitation of 488 nm and emission range of 530 to 650 nm using an Argon ion laser (10 % power) as the excitation source.

2.3 Measurement of photosynthesis rate in *P. pinnata.*

Photosynthetic efficiency of the mature *P. pinnata* trees available at TOIL was determined by measuring the leaf gas exchange parameters using an infrared gas exchange system (LI–6400XT, LI–COR Inc., Lincoln, USA). Photosynthetic readings were taken on fully mature leaves during the morning session with adequate sunlight (7.00–11.00 hrs.). The experimental conditions maintained are: saturating photosynthetically active radiation (PAR) of 1,800 µmol $m^{-2}s^{-1}$ supplied by a LED light source of the gas exchange system which is attached to a leaf chamber. The air temperature of the chamber was at about 25 ◦C and the relative humidity approximately at 50–60 %. Gas exchange parameters such as light saturated net photosynthetic rate (A_{sat}), stomatal conductance (g_s), and transpiration rate (E) were recorded. Later, the leaf water use efficiency (WUEi) was determined manually by A_{sat}/E.

2.4 Biochemical analysis of *P. pinnata* leaves.

Chlorophyll content in the leaf samples was estimated according to the method reported by Hiscox and Israelstam (1979). Fresh tissues (100 mg) were ground and refluxed for 1 h in 10 mL of DMSO, extracted liquid was collected and absorbance was measured at 645 and 663 nm using a UV-Visible spectrophotometer (Eppendorf, Germany) and chlorophyll contents were calculated according to Arnon (1949) using the formula: chlorophyll a (g/L) = $0.0127 \times A_{663} - 0.00269 \times A_{645}$; chlorophyll b (g/L) = $0.0229 \times A_{645} - 0.00468 \times A_{663}$; the results were expressed as mg/g FW (where A_{663}, A_{645} represents the absorbance measured at 663 and 645 nm respectively).

Total carbohydrates and starch content in the leaves were estimated according to Anthrone method as described by Hedge and Hofreiter (1962), where 100 mg of fresh seed tissue was refluxed with 5 mL of 2.5 N HCl followed by neutralizing with NaOH. The mixture was centrifuged for 10 min at 6,000×g. An aliquot of supernatant (1 mL) was taken and added with 4 mL of Anthrone reagent. After heating at 80–90 ℃ for 8 min in a water bath, the solution was allowed to cool rapidly and the absorbance was measured at 630 nm. Spectrometric readings were quantified using glucose standards and represented in mg/g FW of the seed.

For starch estimation, 100 mg of fresh tissue was first ground with hot 80% ethanol. The supernatant of this mixture was removed after centrifugation. The residue was washed repeatedly with 80% hot ethanol till the washings did not give green color with Anthrone reagent. The dried residue was mixed with 5 mL of water and 6.5 mL of 52% perchloric acid. The mixture was kept at 0℃ for 20 min and centrifuged at 8,000×g for 5 min, supernatant was collected and the extraction was repeated twice. The collected supernatant (1 mL) was mixed with 4 mL of Anthrone reagent and starch was estimated by the method used in estimation of carbohydrates. Similarly, an aliquot of the supernatant was mixed with 1.1 % hydrochloric acid followed by 30 min boiling in a water bath. Solution was diluted with 10 mL of DDW and the soluble sugars were estimated using Anthrone reagent as described earlier for starch and total carbohydrate estimation.

2.5 Quantification of free amino acids using HPLC.

Extraction of free amino acids was done according to Bieleski and Turner (1966) with modifications, 100mg of leaf sample was ground with 600 µL of chilled solution containing H_2O:$CHCl_3$:MeOH (3:5:12 v/v) followed by centrifugation for 2 min at 4 ℃. Supernatant was collected and mixed with another 300 µL and 450 µL of chilled $CHCl_3$ and H_2O respectively. After centrifugation the upper phase was collected and filtered using a 0.45 µM pore size membrane filter. For quantification and amino acid profile analysis using High performance liquid chromatography HPLC, the filtrate was pre column derivatized with o-phthalaldehyde, OPA reagent (27 mg OPA with 90 mg Sodium tetraborate, N–Acetyl–L–Cystiene (66 mg), 4.5 mL HPLC grade water, 0.5 mL MeOH) (Sigma, USA) and FMOC reagent (Sigma, USA) (50 mg of FMOC dissolved in 20 % ACN). An aliquot of sample (10 µL) with 5 µL each of OPA and FMOC reagent was mixed and total 20 µL was injected into the C18 packed column (Spinchotech, 4.5 mm x 150 mm, 5 µm). The HPLC system comprises of dual pump with a PDA detector (Shimadzu, prominence LC–20 AD HPLC system). The injected sample was separated by binary gradient of sol A – 75 mm phosphate buffer, pH – 6.3: MeOH:ACN (76.5:20:3.5) and sol B – MeOH:ACN:H_2O (45:45:10) at 1.2 mL / min flow rate. Step gradient was set from 0 % of B from up-to 18.1 min, 57 % of B from18.1 to 18.6, 100 % of B from 18.6 to 22.3 min, 100 % of A from 22.3 to 26 mins. All the amino acids were detected at 338 nm and proline was detected at 262 nm with a band width of 20. Quantification was achieved with standard mixture of amino acids (Sigma, USA) using area normalization method (Kühnreich and Holzgrabe 2016).

2.6 Label free quantification (LFQ) of foliar proteins.

Fully developed *P. pinnata* leaves were collected in liquid N_2 and 500 mg of the cryo–preserved leaf sample was then crushed using 2 mL of extraction buffer (25 mM Tris HCl, 15 mM $MgCl_2$, 15 mM EGTA, 75 mM NaCl, 2 mM 1,4–DTT, 0.1% nonidet P–40, 1 mM NaF, 1 mM PMSF). The extract was centrifuged at 10,000 rpm for 15 min at 4 °C and the supernatant was collected. For the precipitation of proteins, the supernatant was aliquoted into tubes and 4 volumes of ice cold 0.1 M ammonium bicarbonate solution (in methanol) was added. The mixer was then kept for precipitation of proteins at − 20 °C overnight. The proteins were precipitated by centrifuging at 10,000 rpm at 0 °C. The pellet was retained and washed subsequently for 3 times with chilled methanol and 2 times with chilled acetone. The collected and pooled protein sample was then taken for the LC quantification in triplicates as technical repetitions in order to ensure a valid outcome. The pooled sample was treated with 100 mM DTT for 1 h at 95 °C followed by 250 mM IDA at room temperature for 45 min at dark. After the incubation, the protein samples were digested with trypsin and again incubated for overnight at 37 °C. The digested peptides were extracted in 0.1 % formic acid, incubated at 37 °C for 45 min followed by centrifugation at 10,000 g. The supernatant was collected, vacuum dried and dissolved in 20 µL of 0.1 % formic acid. Separation with the help of liquid chromatography (LC) was performed on ACQUITY UPLC system (Waters, UK) with 75 µm×150 mm×1.7 µm BEH C–8 column (Waters, UK). Sample (10 µL) was injected to the column and the separated peptides were directed to Q–TOF instrument (Waters Synapt G2) for the identification of MS/MS. Program was set to a gradient elution for 60 min using buffer A: 0.1 % formic acid in water and buffer B: 0.1 % formic acid in acetonitrile.

For the identification of proteins, all of the data acquired were further processed using the Protein Lynx Global Server (PLGS) software 3.0.2. For the functional identification, the gene ontology for the proteins was accomplished with the help database from Uniprot.

2.7 Collection of seed samples, storage conditions and experimental methodology.

After the analysis of foliar characteristics of different high yielding *P. pinnata* accessions available at TOIL (Figure 2.3), randomly tagging and the development of pod and seed was monitored regularly. Trees with 4 years of age develops pods which carries 1 to 4 seeds composing a tiny embryo with large cotyledons that are enclosed by a seed coat (Figure 2.4). Flowering initiated in the month of March with visible inflorescence which continued until April (Figure 2.4). Natural pollination resulted in formation of pods with first visible appearance recorded at 80–90 days after the flowering started. Seed development and maturation was completed by nearly 300 days (~10 months) from the first visible appearance of flowers (Figure 2.4). The appearance of flower buds in each tertiary branches was considered as 0 days after flowering (DAF). Individual flowers were tagged randomly and the development of pod and seed was monitored regularly in the tagged trees. The tagged pods were collected at 120, 180, 240 and 300 DAF to study the seed morphology and stored at –80 ℃ for further biochemical and molecular analyses (Figure 2.5). The pod contained very tiny seeds with negligible dry mass during first initial stages of development (30 to 100 DAF). Therefore, the experiments were carried out when the seeds reached 120 DAF till maturation. Seeds inside 120, 180, 240 and 300 DAF old pods were considered as stage1 (S1), stage2 (S2), stage3 (S3) and stage4 (S4) respectively (Figure 2.5). The study was performed for two successive flowering seasons of *P. pinnata* (for two consecutive years, March 2016 to February 2017 and March 2017 to February 2018) (Figure 2.2).

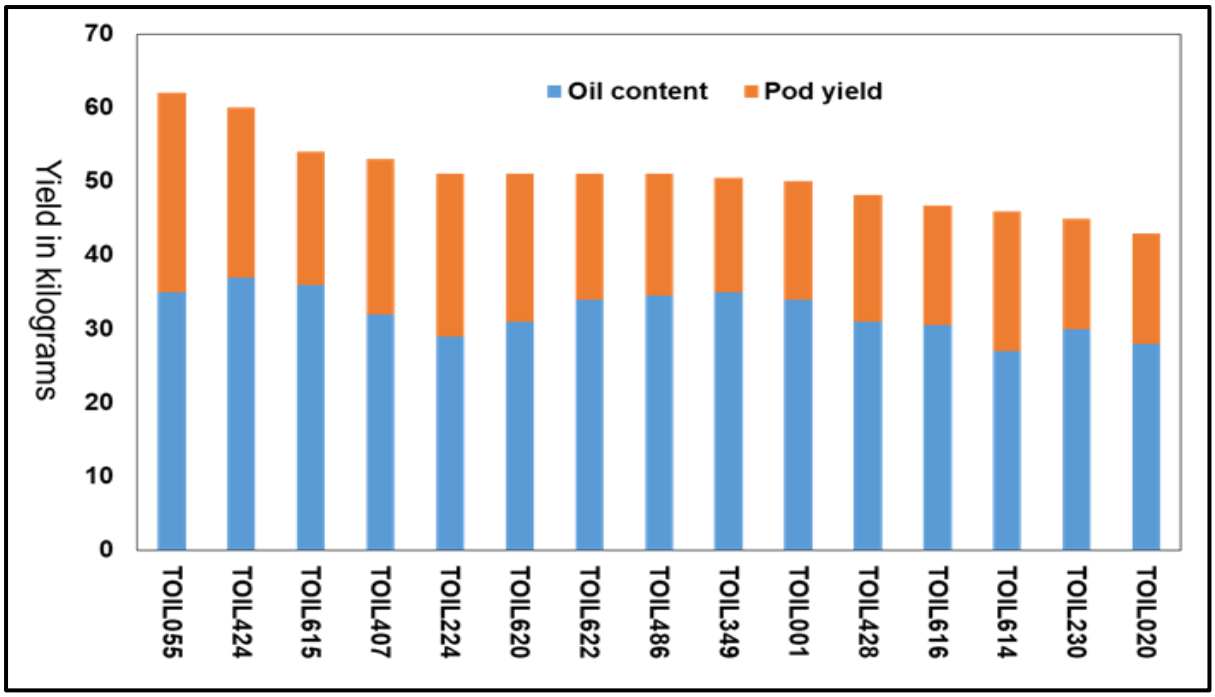


Figure 2.3*: High yielding P. pinnata accessions and their characteristics of oil content with Pod yield.*

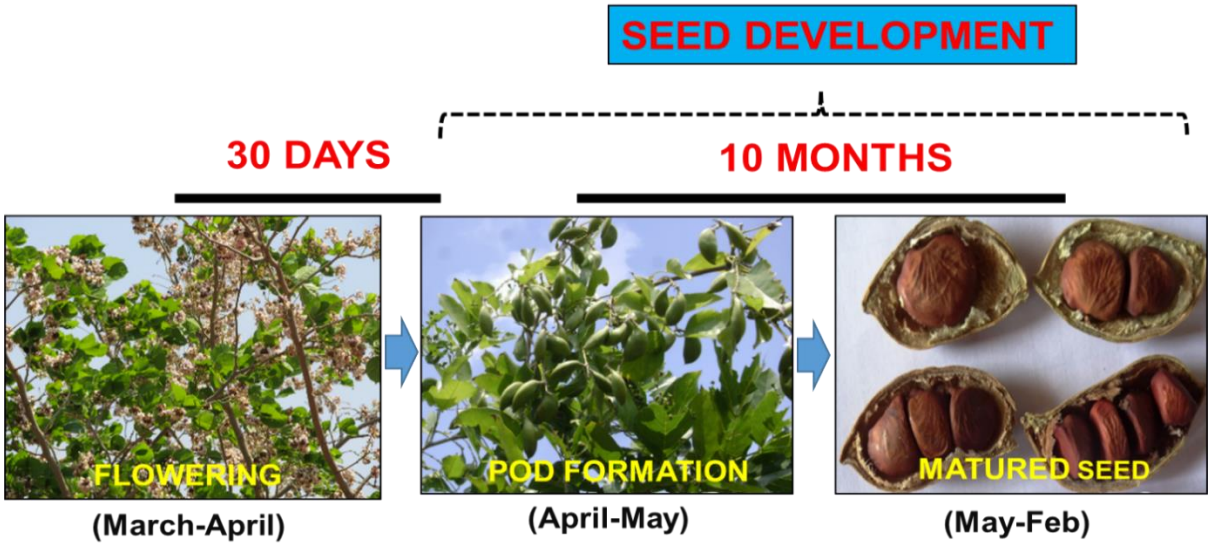


Figure 2.4: *Time scale from flowering to seed development in P. pinnata. Showing the period from flowering to pod formation till maturation.*

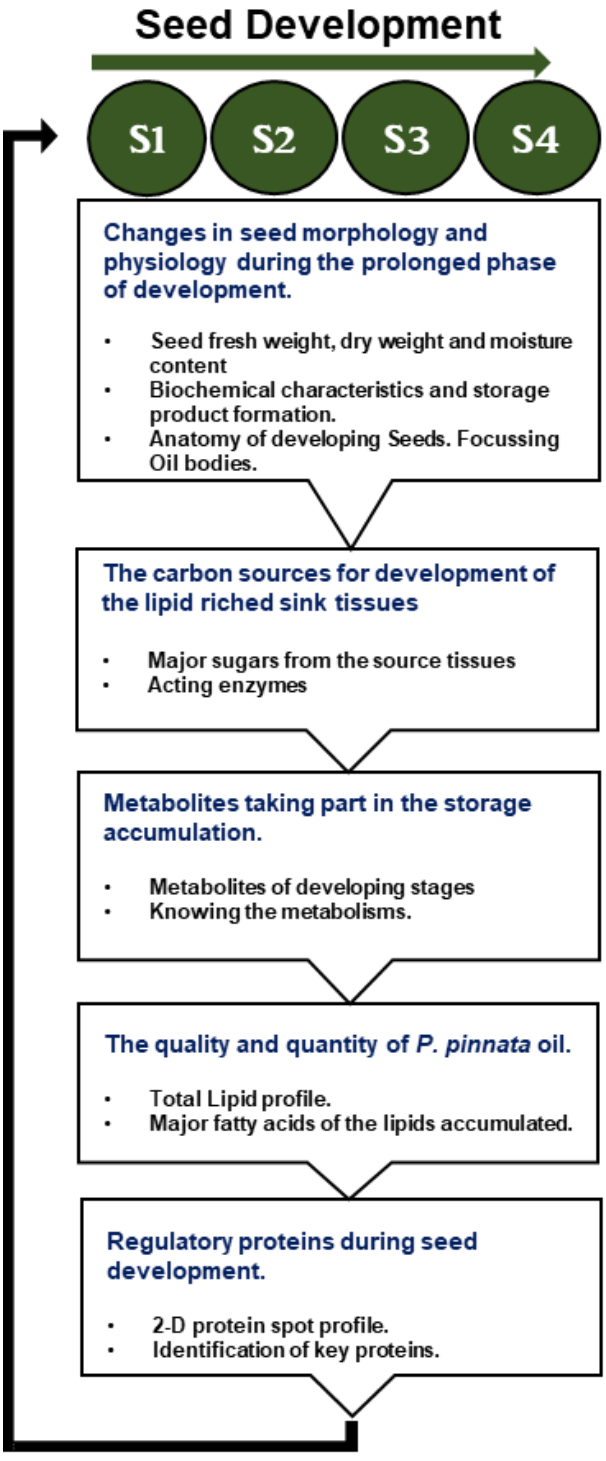


Figure 2.5: *The experimental Layout.*

2.8 Measurement of seed dimensions, fresh and dry weights

Fresh seeds were initially weighed and dried for 12 h at 65–70°C in a ventilated hot air oven. The weight of seed coat and cotyledon were measured gravimetrically and percentage was calculated using the formula: (weight of cotyledon / weight of total seed) ×100; (weight of seed coat / weight of total seed) ×100, respectively. The seed dry mass was recorded and seed moisture content (%) was calculated according to the formula: (FW–DW/FW) ×100.

2.9 Estimation of seed pigments

Chlorophyll content in the seed samples was estimated according to the method explained earlier for the foliar tissues.

2.10 Estimation of seed storage compounds

The freshly collected seeds were used for estimating the major storage compounds. The total carbohydrates along with starch content were estimated by Anthrone method described earlier for the leaf tissues.

For total protein quantification, fresh tissue (100 mg) was ground with liquid nitrogen and mixed with 2 mL of protein extraction buffer consisting of 25 mM Tris-HCl (pH 7.2), 15 mM $MgCl_2$, 15 mM EGTA, 75 mM NaCl, 2 mM DTT, 1 mM NaF, 1 mM PMSF with 0.1% Nonidet, followed by centrifugation at 12,000×g for 5 min at 4 °C. Supernatant was collected and protein was estimated according to Bradford et al. (1976). Quantification of total protein was achieved by using BSA as the protein standard and represented as mg/g FW.

Total lipids were extracted by using the method described by Bligh and Dyer (1959) with slight modifications. Fresh seeds were ground with $CHCl_3$:MeOH:H_2O (2:2:1.8) v/v and the mixture was centrifuged at 10,000×g for 15 min. The lower $CHCl_3$ layer was collected and the extraction was repeated three times.

The chloroform was separated by a Rotary evaporator (Heidolph, Germany) and total lipid was estimated gravimetrically and represented as g oils/100 g seed. All of the chemical reagents used for biochemical analysis were purchased from Sigma Aldrich, USA.

2.11 Microscopic studies of *P. pinnata* seeds

Fresh seeds were made into thin sections and fixed in 2.5 % glutaraldehyde for 8–12 h followed by re-fixing again in 0.2 % osmium tetra oxide for 2–3 h. After washing with 0.1 % PBS and series of ethanol dehydration, thin sections were embedded in epoxy resin (Araldite 502). Ultrathin sections were made using ultra microtome (LEICA EM UC6, Germany) and observed under Scanning electron microscope (TESCAN S8000, Czech Republic). For TEM, ultrathin sections were fixed in grids and stained with uranyl acetate and lead citrate which was observed later in TEM (FEI Model, Tecnai G2S Twin, Spain) (200 kV). Nile red staining was performed by following the method reported by Greenspan et al., (1985) with slight modifications. Prefixed thin sections of the fresh tissues were infiltrated with the working concentration of 2 µg/mL Nile red solution in 0.1 % HEPES buffer for 1 to 2 h under dark. After washing for 4–5 times with HEPES buffer, the sections were visualized under confocal microscope (LIECA TCS SP2 AOBS, Heidelberg, Germany) under excitation of 488 nm and emission range of 530 to 650 nm. All images were obtained with 10 % laser power.

2.12 Sugar extraction and estimation through HPLC

Both the seed coat and the cotyledons from fresh seeds were separated and ground with chilled MeOH (80 %). The supernatant was collected and dried using speed vacuum spinner for 1 Hour. The dried remains were again mixed with HPLC grade MilliQ water and filtered further using 0.45 µ nitrocellulose membrane filter (Merck millipore, Sigma).

The filtrate (20 µL) was then injected into the HPLC system with PDA detector (Shimadzu, prominence LC–20 AD HPLC system). The running solvent consists of 70 % ACN and the sugars were eluted within 10 min run time using an isocratic gradient of flow rate1 mL min^{-1}. The quantification of sugars is based on sugar standards analyzed individually or in mixture. Concentrations of 1–100 µM of sugars (fructose, glucose and sucrose) were used for standard curve and internal spikings were also analyzed for further confirmation.

2.13 Enzyme Extraction

The freshly collected *P. pinnata* pods were immediately put into the liquid nitrogen. The collected seeds are washed thoroughly with MQ water inside the cold room (2–5 ℃). We have followed the procedure reported by (Hill et al., 2003) for the extraction of enzymes, with some modifications in the buffer concentrations. The seed coats and cotyledons from 5–10 seeds were pulled up separately and ground with 1 mL of buffer A consisting of 50 mm HEPES–KOH (pH 7.5), 5 mM KCl, 2 mM $MgCl_2$, 1 mM EDTA, 5 mM DTT, and 10 mg BSA and 5 mg of PVPP with the help of a chilled mortar and pestle. Finely ground mixture was allowed to settle for 2 min and the supernatant was collected carefully with the help of a pipette. The resultant supernatant was used for estimation of all the enzymatic activities.

2.14 Invertase Assay

Alkaline invertase, 50 µL of supernatant extract was added to 50 µL of assay buffer consisting of 50 mm HEPES–NaOH, pH 7.5 and 25 µL of 0.5 M Sucrose. The reaction mixture was kept for 5 min incubation and later kept to boiling to stop the reaction. Reaction at zero–time of incubation was taken as the control samples and it had been boiled immediately. The activity was calculated by subtracting the fructose content of control from the incubated reactions.

Acid invertase assay followed the same method as described above but with different assay buffer consisting of 300 mm sodium acetate (pH 4). Moreover, the reaction was stopped by the addition of 50 µL of 50 g L^{-1} $ZnSO_4$. The mixture was then kept in the ice for 5 min and kept for 5 more min for boiling. The zinc content in the solution was then precipitated by addition of 200 µL of 100 mm K_2CO_3 followed by 5 min centrifugation at 8,000g and the supernatant was collected for estimation of fructose. For neutral invertase assay, the supernatant (50 µL) was mixed with assay buffer containing 50 mM HEPES, pH 7.0. To start the reaction sucrose (25 µL) was added and kept for 5 min followed by heat treatment for 3–5 min. The supernatant was collected and fructose content measured as mentioned earlier.

2.15 Assay of sucrose synthase and sucrose-phosphate-synthase

The breakdown of sucrose by SUS was analyzed by adding 50 µL of the fresh enzyme extract to assay solution containing 1mM UDP, 1 mM PyroPhosphate, 1 mM NAD^+, 2 units of phosphoglucomutase, and 2 units of the Glucose 6–P dehydrogenase as described by Miron and Schaffer 1991 & Rende et al., 2017. The assay was started by adding 25 µL of sucrose to the mixture and kept for incubation for 30 min at 37˚C. The reaction was stopped by adding 1ml of alkaline copper tartrate reagent and fructose released was estimated.

The activity of SPS was determined by using assay buffer containing 50 mM HEPES–NaOH (pH 7.5), 15 mM $MgCl_2$, 25 mM Fructose–6–Phosphate, 25 mM Glcose–6–Phosphate, 25 mM UDP–Glucose, and 50 µL of the extract. Mixtures were incubated for 30 min at 37 °C, and incubation was terminated with the addition of 70 µL of 30 % KOH. In control reactions KOH was added at 0 min. Sucrose was estimated using HPLC.

2.16 Fructokinase and hexokinase assay

Breakdown of hexoses for the formation of Glc6P or Fru6P coupled to $NADP^+$ production was followed for this assay with different concentrations of substrates (glucose or fructose). The reaction mixture was prepared with 100 µL of extract and 100 µL of assay buffer containing 50 mM glycyl–Gly–KOH–pH 8.0, 4 mM $MgCl_2$, 1 mM NAD, 1 mM ATP, 1 unit of Glc–6–phosphate dehydrogenase (from *Leuconostoc mesenteroides,* Sigma), and 3 units of phosphor–glucose–isomerase (Sigma, USA). Glucose or fructose with concentrations 1–100 mM was used for the reaction to start in order to determine the activity of HEXK or FRK respectively. After 5 min of incubation, the absorbance was measured at 340 nm.

2.17 Metabolite profiling

P. pinnata seeds at different developing stages were ground into fine powder in liquid nitrogen. 100 mg of the seed powder was homogenized with 1.4 mL precooled methanol by vortexing for 10s and 60 µL of ribitol (0.2 mg/mL) was added to the mixture as internal standard followed by 10s of vortexing. This solution was ultra-sonicated for 10 min, followed by centrifugation at 11000×g for 10 min. The supernatant was transferred to a new tube and mixed with 750 µL of precooled chloroform and 1.5 mL of precooled water. The mixture was then centrifuged at 2200×g for 15 min. 150 µL of extraction solution from upper phase was dried under vacuum and stored at –80 ℃ until derivatization. The extract was methoxyaminated, silylated and the dried extract was dissolved in 20 µL of methoxy amine hydrochloride pyridine solution (40 mg/mL) which was incubated at 30 ℃ with vigorous shaking for 90 min. Then 80 µL of N-methyl-N-(trimethylsilyl) trifluoro-acetamide-solution was added to the sample followed by 30 min incubation at 37 ℃ with vigorous shaking. The derivatized sample was centrifuged at 20,000×g for 8 min.

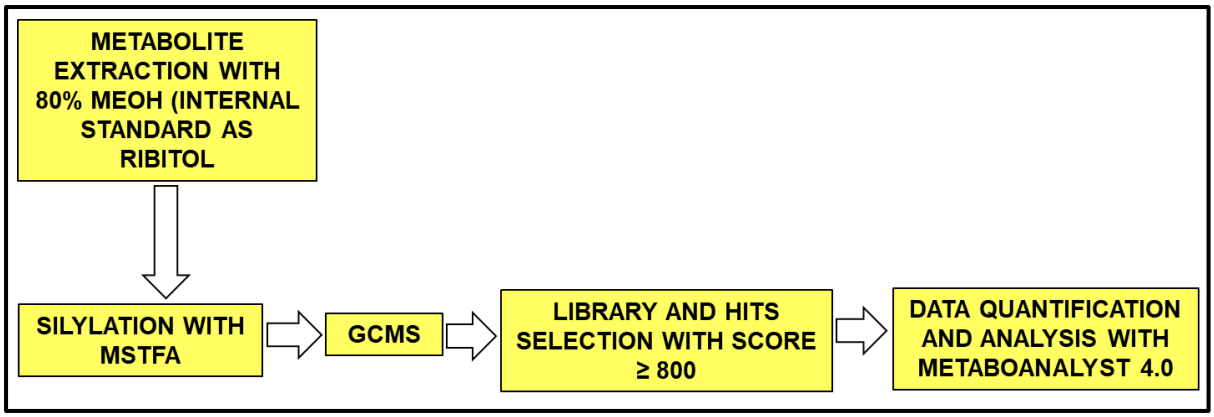


Figure 2.6: *Schematic procedure followed for the metabolite profiling.*

The supernatant was then transferred to vials for measurement. Samples were measured with gas chromatography coupled with LECO Pegasus R 4D GC GC–TOF spectrometry (GC-TOF-MS) (Agilent 6890, USA). Each sample was injected under both split less and split 25 times mode for better quantification of candidates with a wide capacity range. Candidates were manually annotated by comparing their retention times (RTs) and mass spectra to those of standards in GMD database (Kopka et al., 2005) with a minimum match factor of 700. The peak areas of the same metabolite with different derived groups were merged and normalized to internal standard of ribitol. The concentration of the metabolites was calculated with respect to the known concentration of ribitol and represented as mg/g FW. The differential analysis at developing stages were analyzed using MetaboAnalyst Version 4.0 [Software] (available from http://www.metaboanalyst. ca/MetaboAnalyst/faces /home. xhtml) (Figure 2.6).

2.18 Thin layer chromatography and lipid profile of TAGs

Total lipid (5 mg) extracted from all the samples were dissolved in chloroform and spotted on the TLC Silica gel 60 F254 (Merck, Germany). The separation was carried out with hexane and ethyl acetate (9:1 v/v) as the mobile phase and silica gel bonded with aluminum sheet as the stationary phase. The TLC plates were air dried and exposed to iodine vapor. Iodine-stained spots were compared with the standards of phospholipids, DAGs and TAGs separated in TLC plates with the same mobile phase.

The TAGs from the multiple TLC plates were scrapped carefully and extracted using chloroform. Further, chloroform was separated from the extracted TAGs with the help of rotary evaporator (Heidolph, Germany). The fatty acid profiling and quantification was carried by following the methods reported by Coetzee et al., (2008) and Sun et al., (2017) with slight modifications. FAMEs were prepared by refluxing 50 mg of the TAGs extracted with 5 % H_2SO_4 in methanol (w/v) for 6 h on a hot plate. Prepared esters were analyzed through gas chromatography (GC-TOF-MS) (Agilent 6890, USA) with DB225 column (inner diameter = 0.25 mm, length = 37 m, thickness = 0.25 µm; Agilent, USA). The injector and flame ionization detectors were set at 250 °C and 270 °C, respectively. The oven temperature was set at 160 °C for 2 min and then increased to 230 °C at a rate of 5 °C/min. Nitrogen was used as the carrier gas at a flow rate of 1 mL/min. Fatty acids were confirmed by comparing with standard FAMEs mix (C14 – C22, Supelco, Sigma Aldrich, USA) analyzed at different concentrations. Quantification of fatty acids was carried out using area normalization method with percentage area of each peak corresponding to the identified fatty acid.

2.19 2–D Protein profiling and identification using MALDI TOF MS/MS analysis

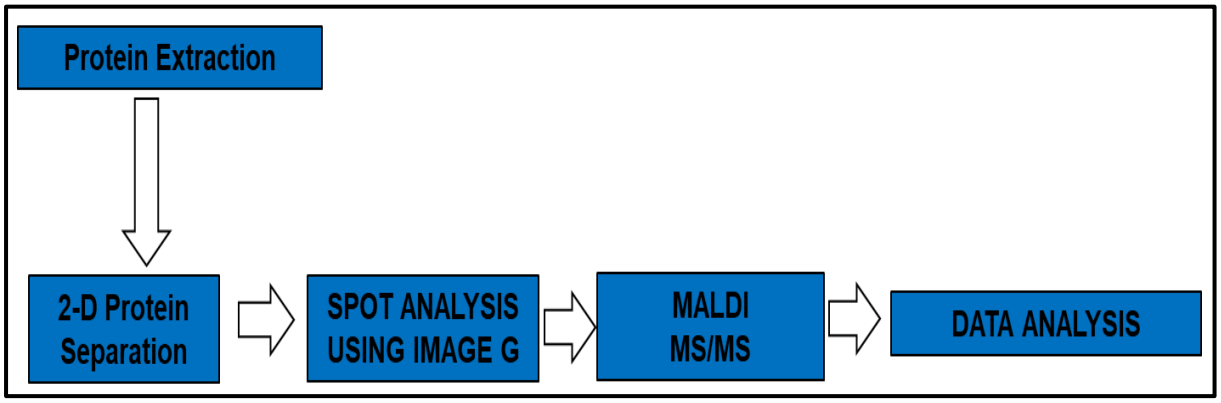


Figure 2.7: *Schematic procedure followed for the 2–D Proteomic analysis.*

Gel based 2D proteome was carried out according to the method described by Sengupta et al. (2011) (Figure 2.7). Protein extracted earlier (200 µL) was precipitated using 800 µL of 0.1% ammonium acetate in methanol and kept at –20 ℃ for 8–12 h. The mixture was centrifuged at 10,000×g for 10 min at 4 ℃ and the precipitate was washed with fresh methanol for 3 times followed by 2 times wash with acetone. The isolated protein sample (900 µg) was then solubilized in rehydration buffer followed by isoelectric focusing using pH 4–7 isoelectric focusing strips (18 cm, 4–7 pH linear gradient; GE Healthcare, USA) for 12 h at 50 V for first separation. Rehydration and focusing was carried out in Ettan IPGphor II (GE Healthcare, USA) at 20 ℃, using the following program: 30 min at 500 V, 3 h to increase from 500 to 10,000 V and 6 h at 10,000 V (a total of 60,000 Vh). The IEF strips were then placed over 12 % acrylamide gel for the second separation through SDS PAGE at 300–500 V using standard protocol. After staining, the gels were scanned and the gel image obtained was analyzed though Image Master 2-D Platinum version 6 image analysis software (GE Healthcare, USA).

For the identification of proteins, MALDI TOF MS/MS was carried out through in-gel trypsin digestion and database searches (PMF and MS/MS) using MASCOT program [Software] (available from http://www.matrixscience.com). The similarity search for mass values was done with existing digests and sequence information from NCBInr and Swiss Prot database. The taxonomic category was set to Viridiplantae (green plants). The other search parameters were: fixed modification of carbamidomethyl (C), variable modification of oxidation (M), enzyme trypsin, peptide charge of 1+ and monoisotopic. Out of top ten most significantly identified proteins, results having highest score, peptide match, and similarity of molecular weights were considered.

2.20 Gene expression analysis by real time PCR

Total RNA was isolated from developing fresh seeds using Plant total RNA extraction kit by following manufacturer's instructions (Sigma Aldrich, USA). Primers for the required genes were designed from available *P. pinnata* transcriptome sequences (Sreeharsha et al., 2016). First strand cDNA was synthesized with 1µg of RNA using Revert aid first strand cDNA synthesis kit (Thermo-Fischer Scientific, USA). Expression analysis of selected genes were carried out on Realplex thermal cycler (Eppendorf, Germany) using SYBR FAST qPCR universal master mix (2X) (KAPA Biosystems, USA) with 50 ng of cDNA as template and following the program: 2 min at 95 °C, followed by 40 cycles of 15 seconds at 95 °C, 30 seconds at 55 °C annealing temperature and 20 seconds at 72 °C, followed by the dissociation (melting) curve. The mRNA expression level was calculated according to the $2^{-\Delta\Delta Ct}$ formula (Livak and Schmittgen. 2001). For the internal control, 18s ribosomal RNA gene was used.

2.21 Statistical analysis

Each seed collected at various stages was considered as one unit (n = 40), where n is the number of seeds for the verification of difference in its morphological and physiological measurements. All the experiments were carried out in triplicates for the developing stages. PCA and heat map analysis was carried out for metabolome data using MetaboAnalyst online tool version 4.0 [Software] (available from http: //www.metaboanalyst.ca/MetaboAnalyst /faces /home. xhtml). The mean values were compared and analysis of significance ($P<0.05$) was determined by student's t-test and one way ANOVA using Sigma plot 11.0.

.

CHAPTER 3

RESULTS

3.1 Stomatal characteristics and the photosynthetic gas exchange parameters

The lower epidermal peel of *P. pinnata* showed stomatal aperture, measuring ~3.45 µ with a density of 201±20 mm^{-2} of the leaf surface area (Figure 3.1; Table 3.1). Both SEM observed and the fluorescent images of the stomata showed a typical pattern similar with other C3 plants (Figure 3.1). However, the AO-stained epidermal peel clearly showed guard cells and other subsidiary cells surrounding it (Figure 3.1). Photosynthetic gas exchange parameters such as the rate of photosynthesis (A_{sat}), stomatal conductance (g_s), transpiration rate (E) and water use efficiency (WUE) were also successfully measured using IRGA (Table 3.1).

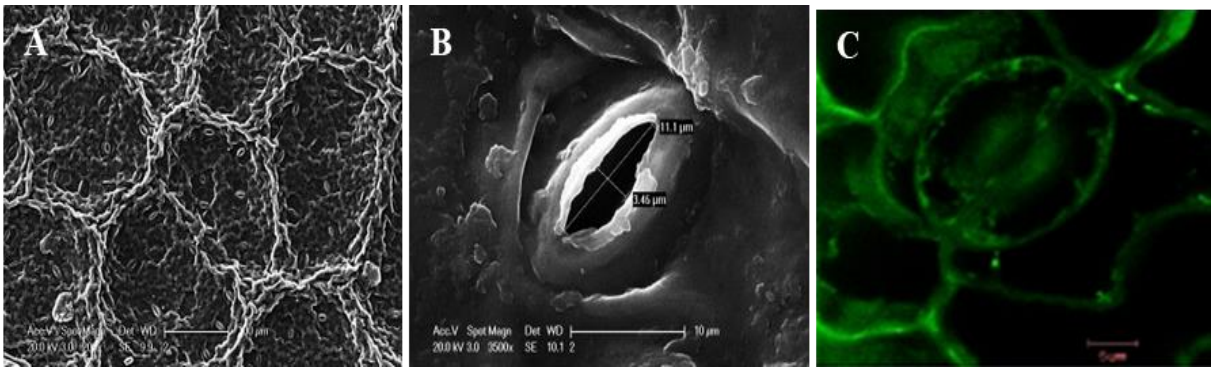


Figure 3.1: *Stomatal anatomy of the ventral surface of P. pinnata leaves: (A) SEM image wide view (Scale: 100 µm) (B) SEM image narrow view (Scale:10 µm) (C) Acridine orange–stained stomate observed under confocal microscope (Scale:10 µm).*

Gas exchange and photosynthetic parameters	Values
Stomatal aperture (diameter)	3.45±1.1 µ
Stomatal density (mm^{-2})	201±20
A_{sat} / Photosynthesis rate (µmol m^{-2} s^{-1})	7.8±2
g_s / Stomatal conductance (µmol m^{-2} s^{-1})	1.2±0.3
E / Transpiration rate (µmol m^{-2} s^{-1})	3.2±0.2
WUE / Water use efficiency (µmol m^{-2} s^{-1})	2.9±0.3

Table 3.1: *Leaf gas exchange characteristics along with stomatal aperture and stomatal density of P. pinnata leaves. Values are mean ± SD (n=3).*

Foliar biochemistry

The pigment content of mature *P. pinnata* were analyzed where the chl a content was more (2.84 ± 0.1 mg g^{-1} FW) than the chl b content (1.1 ± 0.1 mg g^{-1} FW). The ratio is more than 1 and the total carotenoids content ranges in 0.55 ± 0.05 mg g^{-1} FW (Table 3.2).

Chlorophyll a	2.84 ± 0.1 (mg g^{-1} FW)
Chlorophyll b	1.11 ± 0.01 (mg g^{-1} FW)
Total Chlorophyll	3.95 ± 0.8
Chlorophyll a/b ratio	2.56 ± 0.1
Carotenoids	0.55 ± 0.05

Table 3.2: *Pigment analysis in the leaves P. pinnata. Values are mean ± SD (n=3).*

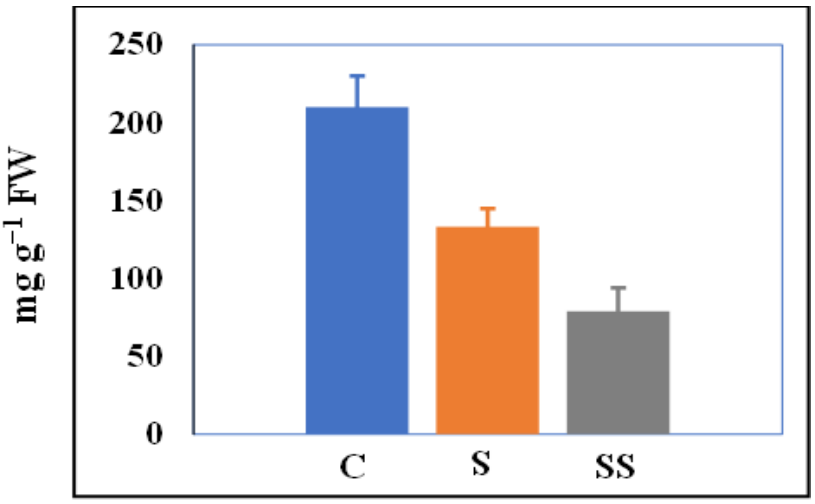


Figure 3.2: *Foliar content of total carbohydrate, starch and soluble sugars. Values are mean ± SD (n=3). C – Carbohydrates, S – Sucrose, SS – Soluble sugars.*

The foliar contents of total carbohydrates, starch and soluble sugars were also analyzed along with a complete profile of free amino acids (Figure 3.2, Table 3.3). Total carbohydrate content was much higher when compared to starch followed by total soluble sugars in the foliar parts of *P. pinnata* (Figure 3.2). among the free amino acids, asparagine (3.8±0.02 µM g^{-1} FW), glutamine (3.4±0.11 µM g^{-1} FW), glutamate (2.2±0.16 µM g^{-1} FW) and glycine (2.1±0.03 µM g^{-1} FW) were the most abundant followed by arginine, aspartate, serine, alanine, and threonine. Rest of the free amino acids detected were in traces and their concentration is very less (Table 3.3).

Amino acids	Concentration (µmole g^{-1}FW)	Amino acids	Concentration (µmol g^{-1}FW)
Aspartic acid	1.6±0.02	Alanine	1.4±0.11
Glutamic acid	2.2±0.16	Valine	0.7±0.04
Asparagine	3.8±0.02	Methionine	0.5±0.01
Serine	1.75±0.01	Tryptophan	0.3±0.01
Glutamine	3.4±0.11	Isoleucine	0.1±0.03
Histidine	0.5±0.1	Lysine	0.7±0.01
Arginine	1.9±0.12	Leucine	0.2±0.01
Threonine	1.2±0.02	Proline	0.9±0.01
Glycine	2.1±0.03		

Table 3.3: *List of free amino acids and their content detected in the foliar parts of P. pinnata leaves. Values are mean ± SD (n=3).*

3.3 Identification of leaf proteins using nano–LC–MSMS

Identified proteins of major metabolic pathways and processes of *P. pinnata* leaves were sorted and major important proteins are listed in the Table 3.4. Out of ~ 5000 known and unknown proteins, some of the key proteins are presented and grouped based on the main functional ontology taking part in various cellular processes (Table 3.4). Proteins controlling amino acid metabolism such as S–adenosylmethionine synthase, bifunctional UDP–glucose 4–epimerase and UDP–xylose 4–epimerase, ketol–acid reductor–isomerase, glutamine synthetase cytosolic isozyme, phenylalanine ammonia–lyase class 1, asparagine synthetase, glutathione S–transferase 3, and primary amine oxidase proteins were some of the prominent ones having characteristic scores (Table 3.4).

Carbon fixation related proteins detected mainly were glyceraldehyde–3–phosphate dehydrogenase A, ribulose bisphosphate carboxylase small chain 1, ribulose bisphosphate carboxylase large chain 1, ribulose bisphosphate carboxylase/oxygenase activase etc. Additionally, there were also key proteins controlling the cell division and cell cycle such as retinoblastoma–related protein, cell division control protein 2, G2/mitotic–specific cyclin, serine/threonine–protein phosphatase, ATP–dependent zinc metalloprotease FTSH, dynamin–related protein 12A, cell division control protein 2, retinoblastoma–related protein 1. Proteins controlling glycolysis are mainly fructose–1,6–bisphosphatase, glucose–6–phosphate 1–dehydrogenase, fructose–bisphosphate aldolase, glucose–1–phosphate dehydrogenase, phosphoglucomutase, pyruvate decarboxylase 2, pyruvate kinase etc. Major proteins controlling the process of light dependent photosynthesis are mainly oxygen–dependent coproporphyrinogen–III oxidase, oxygen–evolving enhancer protein 1, phosphoenolpyruvate carboxylase, photosystem I P700 chlorophyll a, apoprotein A, Photosystem II CP47 reaction center protein, photosystem II protein D1, phytochrome A, phytochrome B, protein STAY–GREEN and protochlorophyllide reductase. There were also proteins related to the metabolism secondary metabolites production (chalcone synthase chalcone—flavonone isomerase, dihydroflavonol 4–reductase, flavonoid 3–O–glucosyltransferase, glutathione gamma–glutamyl–cysteinyl–transferase, isoflavone 2'–hydroxylase and transport mechanism (porphobilinogen deaminase, probable ion channel SYM8, probable ureide permease A3, Sec–independent protein translocase protein TATB, secretory carrier–associated membrane protein, vacuolar–sorting receptor 1, V–type proton ATPase catalytic subunit A) (Table 3.4). All of the proteins listed above were selected based on the PGLS score, coverage and the matched peptides.

Accession	Description	Functional ontology	Molecular weight (da)	Score
A4PU48	S–adenosylmethionine synthase	Amino acid metabolism	43251	184.8239
B0M3E8	Bifunctional UDP–glucose 4–epimerase and UDP–xylose 4–epimerase	Amino acid metabolism	38935	174.694
O82043	Ketol–acid reductor–isomerase	Amino acid metabolism	62812	98.6399
O82560	Glutamine synthetase cytosolic isozyme	Amino acid metabolism	39183	94.6895
P07218	Phenylalanine ammonia–lyase class 1	Amino acid metabolism	55817	295.4599
P19251	Asparagine synthetase	Amino acid metabolism	66311	133.2889
P46417	Glutathione S–transferase 3	Amino acid metabolism	25884	131.2395
P49252	Primary amine oxidase	Amino acid metabolism	75510	228.0833
P50246	Adenosyl–homo–cysteinase	Amino acid metabolism	53116	138.6404
P56707	Selenocysteine methyltransferase	Amino acid metabolism	36691	177.7343
Q04708	Pyrroline–5–carboxylate reductase	Amino acid metabolism	28220	151.9674
Q3LRV4	Bifunctional nitrilase/nitrile hydratase NIT4B	Amino acid metabolism	38031	91.299
P12858	Glyceraldehyde–3–phosphate dehydrogenase A	Carbon fixation	43311	568.68
P00868	Ribulose bisphosphate carboxylase small chain 1	Carbon fixation	15873	175.7317
P27066	Ribulose bisphosphate carboxylase large chain 1	Carbon fixation	52576	638.6
O64981	Ribulose bisphosphate carboxylase/oxygenase activase	Carbon fixation	48170	577.9896
A9UL14	Retinoblastoma–related protein	Cell division	114126	219.9076
P24923	Cell division control protein 2	Cell division	33459	136.2476
P25011	G2/mitotic–specific cyclin	Cell division	50063	139.2522
P48488	Serine/threonine–protein phosphatase	Cell division	36225	217.5365
P52389	Cell division control protein	Cell division	33978	95.7174

Table 3.4: *List of the identified proteins with nano–LCMS along with their functional ontology.*

Accession	Description	Functional ontology	Molecular weight (da)	Score
P24923	Cell division control protein 2	Cell division	33459	136.2476
Q9SLZ4	Retinoblastoma–related protein 1	Cell division	114403	192.7375
Q9BAE0	ATP–dependent zinc metalloprotease FTSH	Cell division	75633	396.918
Q39821	Dynamin–related protein 12A	Cell division	68312	217.6662
P52416	Glucose–1–phosphate dehydrogenase	Glycolysis	55592	350.0692
Q42796	Fructose–1,6–bisphosphatase	Glycolysis	43869	124.8646
Q42919	Glucose–6–phosphate 1–dehydrogenase	Glycolysis	58886	123.2937
P46256	Fructose–bisphosphate aldolase	Glycolysis	38421	213.9801
Q9SM59	Phosphoglucomutase	Glycolysis	68531	178.6637
P51851	Pyruvate decarboxylase 2	Glycolysis	44050	349.2783
Q42806	Pyruvate kinase	Glycolysis	55266	290.0367
P35055	Oxygen–dependent coproporphyrinogen–III oxidase	Photosynthesis	43237	51.3363
P14226	Oxygen–evolving enhancer protein 1	Photosynthesis	34871	912.6872
P51061	Phosphoenolpyruvate carboxylase	Photosynthesis	110691	262.2128
P05310	Photosystem I P700 chlorophyll a apoprotein A1	Photosynthesis	84145	65.5891
A4GGD1	Photosystem II CP47 reaction center protein	Photosynthesis	56112	86.3835
P06585	Photosystem II protein D1	Photosynthesis	38936	49.0268

Table 3.4 (continued): *List of the identified proteins with nano–LCMS along with their functional ontology.*

Accession	Description	Functional ontology	Molecular weight (da)	Score
P42500	Phytochrome A	Photosynthesis	125653	148.4542
P42499	Phytochrome B	Photosynthesis	129003	147.0571
A7VLV1	Protein STAY–GREEN	Photosynthesis	29632	98.6671
Q01289	Protochlorophyllide reductase	Photosynthesis	42935	71.6007
P51109	Dihydroflavonol 4–reductase	Secondary metabolites	24373	125.5113
Q9MB42	Beta–amyrin synthase	Secondary metabolites	87459	267.4901
P51081	Chalcone synthase	Secondary metabolites	42885	251.7411
Q43056	Chalcone—flavonone isomerase	Secondary metabolites	23789	148.236
A6XNC6	Flavonoid 3–O–glucosyltransferase	Secondary metabolites	50125	148.8072
Q2TSC7	Glutathione gamma–glutamylcysteinyltransferase	Secondary metabolites	55483	126.892
P93147	Isoflavone 2'–hydroxylase	Secondary metabolites	57155	224.4688
Q43082	Porphobilinogen deaminase	Transport	39905	398.6277
Q4VY51	Probable ion channel SYM8	Transport	99633	303.2635
P51109	Dihydroflavonol 4–reductase	Secondary metabolites	24373	125.5113
Q94G16	Sec–independent protein translocase protein TATB	Transport	28417	204.289
Q41706	Probable ureide permease A3	Transport	41302	332.1842
Q9ZTX0	Secretory carrier–associated membrane protein	Transport	32536	434.0522
P93484	Vacuolar–sorting receptor 1	Transport	68851	150.2177
P13548	V–type proton ATPase catalytic subunit A	Transport	68637	157.279

Table 3.4 (continued): *List of the identified proteins with nano–LCMS along with their functional ontology.*

3.4 *P. pinnata* pods and morpho–physiology of seeds

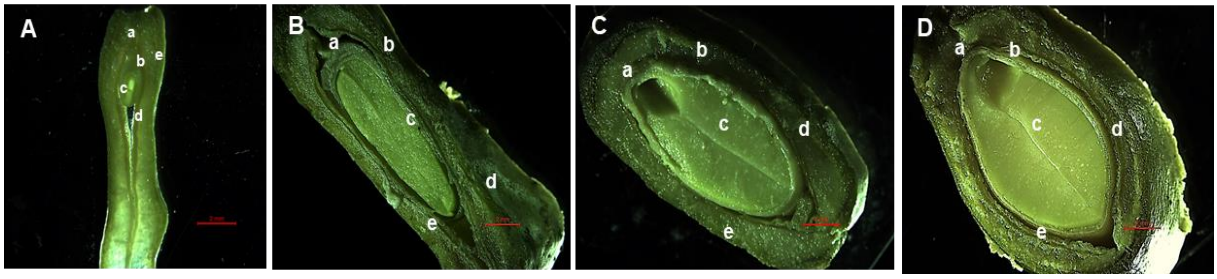


Figure 3.3: *The cross sections of randomly collected P. pinnata pods at different stages of development (very young to a matured stage);* ***"a"****– Funicle (attachment of the seed with the pod),* ***"b"*** *– Seed coat,* ***"c"*** *– Cotyledons,* ***"d"*** *– Inner layer of Pod wall,* ***"e"*** *– Outer layer of Pod wall. A – D; from a very young to a mature pod. (Scale – 2mm).*

The cross sections of intact *P. pinnata* pods showed a clearly differentiated seed coat, cotyledons and the tiny embryo housed under the inner and outer layer of the pod wall (Figure 3.3). The thickness of the pod increased at later stages of development which was maximum at the fully matured stage. The seeds were attached to the inner pod wall layer with the help of a tiny funicle (Figure 3.3 "a"). Figure 3.3 also shows the cross sections of four collected pods randomly at four developing stages. The cotyledons cannot be seen properly at the very young stage (A), whereas we can observe them clearly at later stages (B – D). The developing pods were collected at the interval of 30 days after flowering and their morphology was examined (Figure 3.4). The cotyledons and seed coats were separated (Figure 3.4 A) after the morphology of the collected seeds were observed (Figure 3.4 B). Later, based on the morphology of developing seeds, four stages were selected for all the further analysis (Figure 3.4 C).

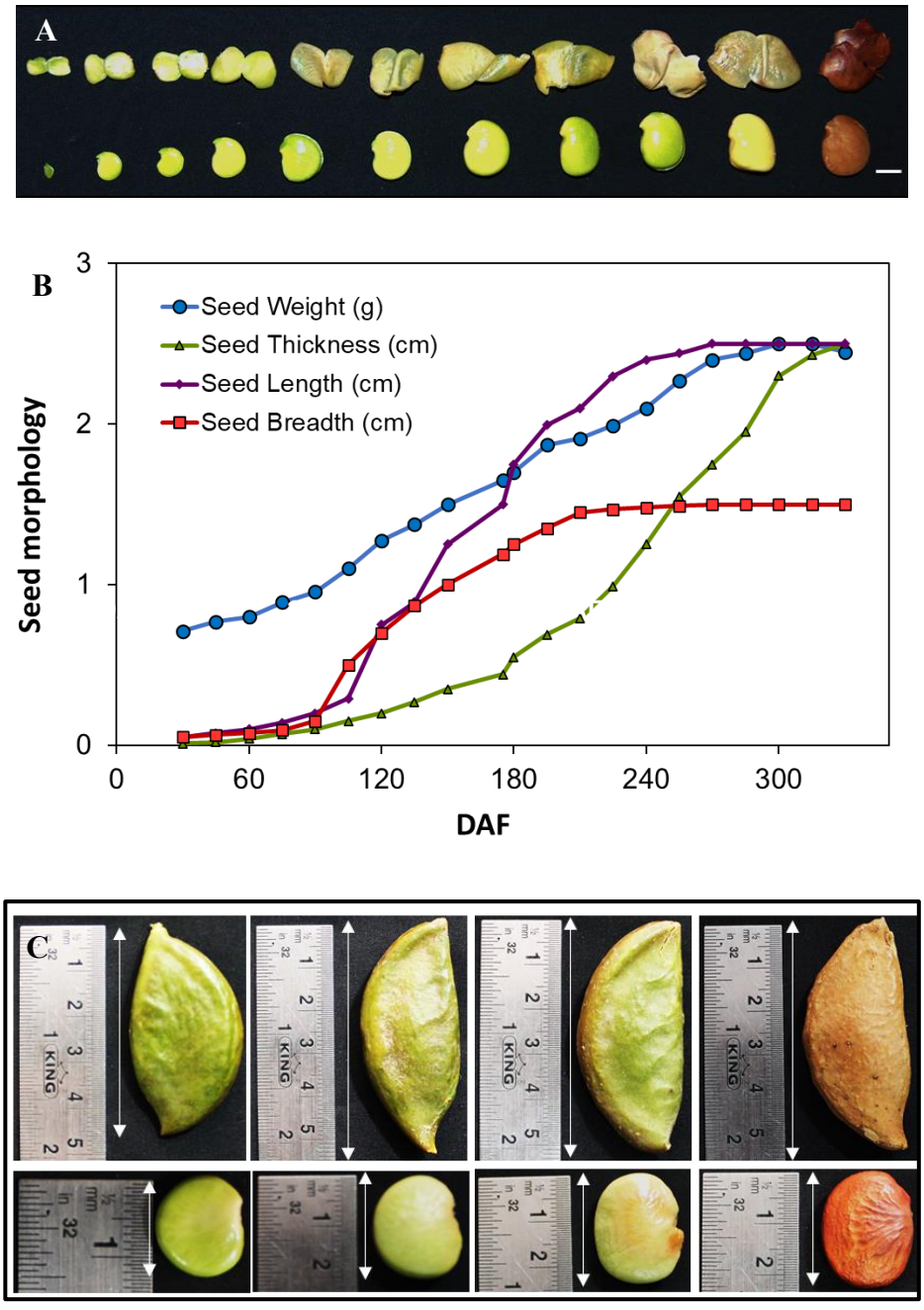


Figure 3.4: *The separated cotyledons from seeds collected at the interval of 30 days from the first appearance of pods* ***(A)****. Morphological measurements of developing seeds* ***(B)****. Selected stages for further analysis* ***(C)****.*

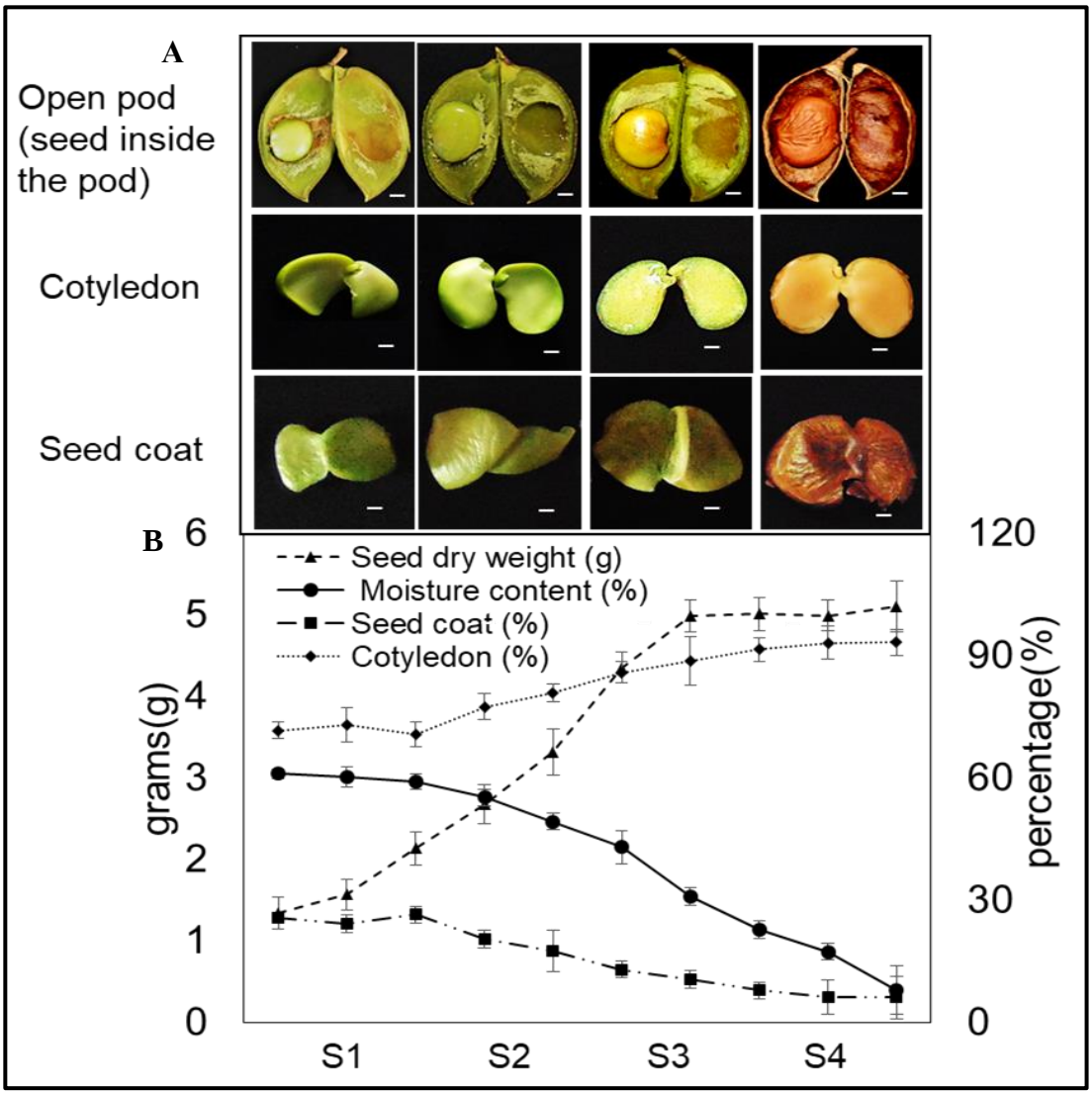


Figure 3.5: *Position of seed inside the pod and cotyledons separated from seed coat at each developmental stage* ***(A)****. Morphological and biochemical analysis of developing P. pinnata seeds. Changes in seed dry weight, moisture content, % cotyledon and seed coat throughout development* ***(B)****. Values are mean ± S.D, (n = 40); the scale is presented by white bar with the length of 0.3 cm.*

The young seeds were green in color initially, which turned completely brown upon maturation (Figure 3.5 A). There was a gradual decrease in the seed moisture content from S1 to S4 whereas, the seed dry weight increased till S3 and got stabilized by maturation (Figure 3.5B). On the other hand, dry weight was increased at S3 with no further increased at S4 (Figure 3.5 B). The percentage of cotyledons increased with maturation whereas seed coat percentage decreased gradually (Figure 3.5 B).

3.5 Pigment content and major storage compounds

Due to the green color of younger developing seeds, major photosynthetic pigments were analyzed in all the stages (S1 to S4). Chlorophyll a, chlorophyll b and total chlorophyll concentrations showed a distinct pattern when the seeds were green and their contents increased from S1 to S2 with a further decrease thereon at S3 and S4. It is clear that the photosynthetic pigments decreased when the seeds started to develop a brown color (Figure 3.6 A). Similarly, chlorophyll a/b ratio also consistently decreased with the progression of seed development (Figure 3.6 A). Relative gene expressions of some of the key photosynthetic genes were analyzed at different stages of development by considering S1 as reference (Figure 3.6 B). Transcripts were selected from the available data of transcriptome data of *P. pinnata* published earlier from our laboratory (Sreeharsha et al., 2016). All genes related to Photosystem I and II major subunits such as *psaA, psaD psbA, psbC, psbQ* and ATP synthase subunit (*atp1A* and *atp1B*) were upregulated at S2 which subsequently decreased from S3 to S4 (Figure 3.6 B). Starting from 120 DAF and keeping 60 days interval up to 300 DAF, i.e 120 DAF (S1), 180 DAF (S2), 240 DAF (S3) and 300 DAF (S4).

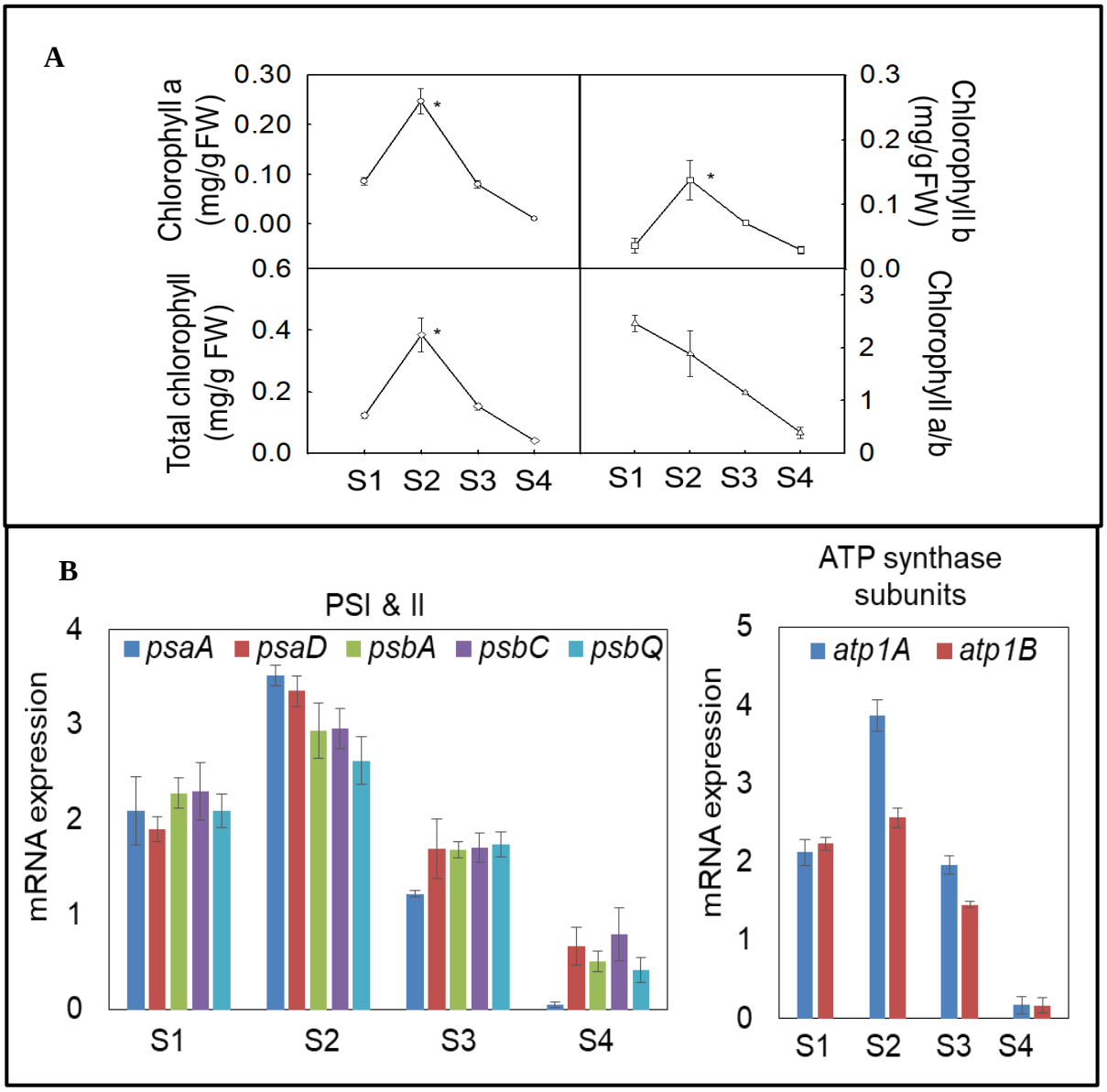


Figure 3.6: *Changes in the chlorophyll a, b, total chlorophyll and chlorophyll a/b* ***(A)****. Expression patterns of key regulatory genes represented in fold change related to photosynthesis* ***(B)****. Fold change >2 is considered as upregulation. Values are mean ± SD (n=3). Data are given as means ± SD; (n=3). *, ** represents significant differences within the stages analyzed by one–way ANOVA, P<0.05; P<0.01 respectively.*

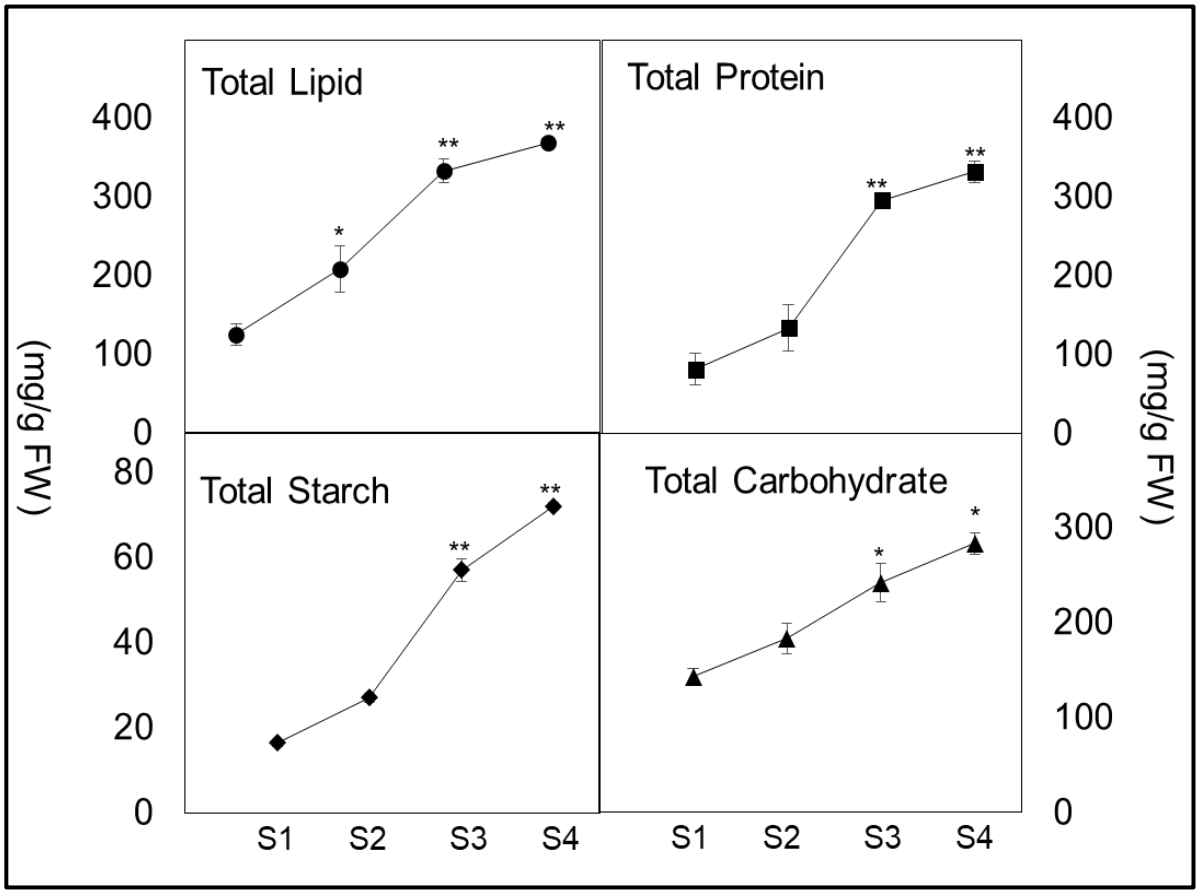


Figure 3.7: *Pattern of increasing storage products (lipid, protein, total carbohydrate and starch) with development. Data are given as means ± SD; (n=3). *, ** represents significant differences within the stages analyzed by one–way ANOVA, P<0.05; P<0.01 respectively.*

Protein and lipid contents were similar at younger stages (S1–S2), while the lipid content reached up to 37 % of total 100 g of seeds (~36.8 g oil/100g seed), whereas protein content reached to 33 % (~331 mg/g FW) by the end of the development. Similarly, the total carbohydrate content in the mature seeds (S4) was ~298 mg/g FW (1.9 folds higher when compared to S1), out of which the content of starch was only ~80 mg/g FW (4.3 folds higher when compared to S1) (Figure 3.7).

3.6 Anatomy of the cotyledonary sections and identification of storage granules

The surface morphology of *P. pinnata* cotyledonary cross sections analyzed by SEM at various stages showed an increase in storage granules from S1 to S4 (Figure 3.8A). Further the biochemical nature of these storage granules were studied through TEM analysis and identified large protein vacuoles at S1 which further increased during S2. The appearance of a large amount of oil body accumulation was observed in S3 and S4, where the lipid droplets were located at cellular periphery (Figure 3.8A, 3.9). For more clarification of the accumulation of oil bodies in developing stages, Nile red fluorescent staining was performed which specifically stains only the lipid containing substances. Oil body accumulation was very low at S1 and S2, but showed rapid increase at S3 to S4 (Figure 3.10). Moreover, the oil bodies were localized near the periphery of the cell with sizes varying from 1 to 2 µ while the protein storage vacuoles have unusually large diameter (~2–5 µ) and situated mostly in the central region (Figure 3.10). The size of the starch, protein granules and oil bodies ranges from 0.5–1 µ, 0.8–4.1 µ and 0.5–2 µ respectively (Figure 3.9). The number of protein granules and lipid bodies in a mature cotyledon range from 6 to 10 $cell^{-1}$ and 120 to 180 $cell^{-1}$ (Table 3.5).

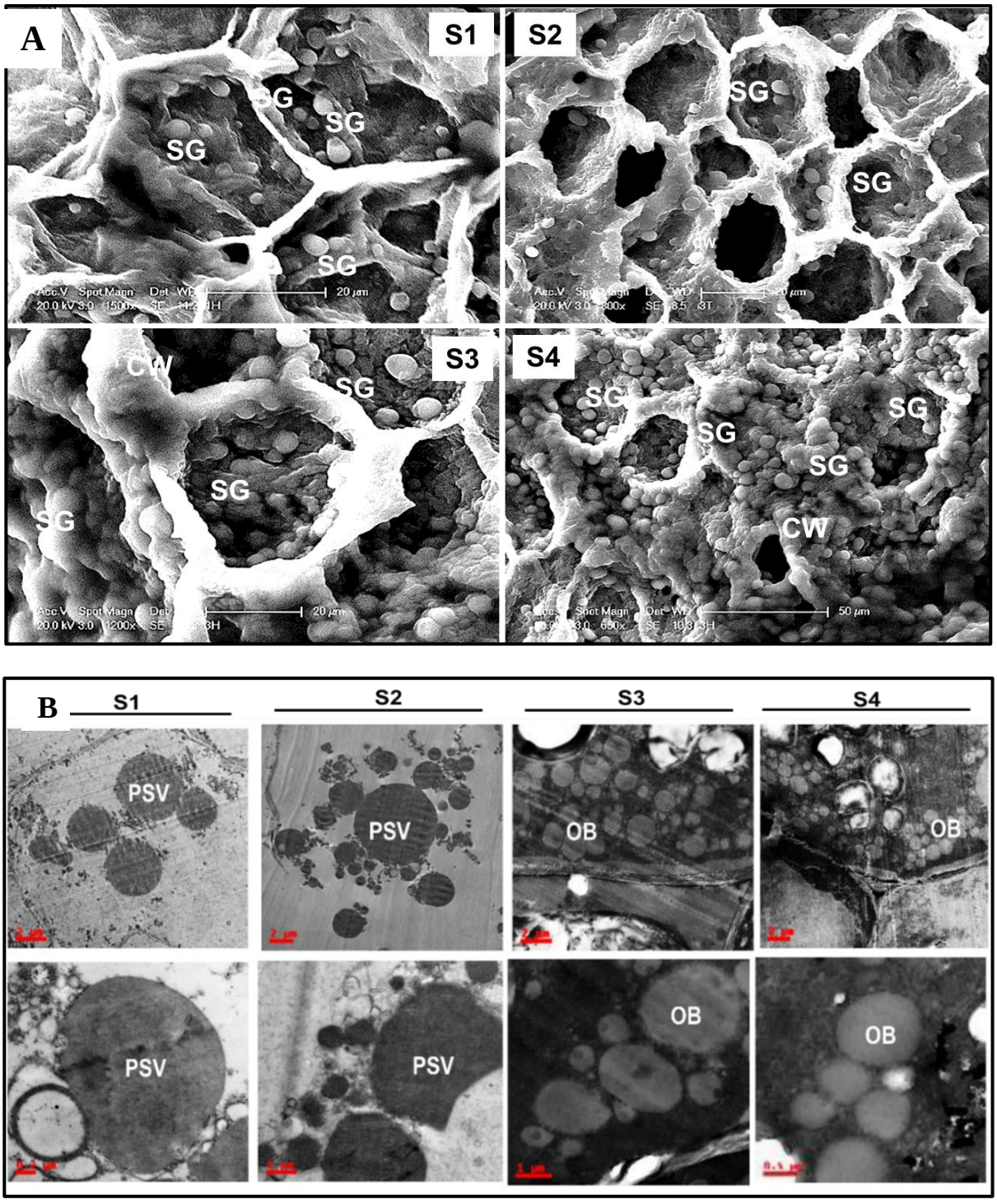


Figure 3.8: *Cotyledonary cross section under SEM from S1 to S4; SG – storage granules, CW – cell wall, Scale (20–50 µ)* **(A)**. *Cotyledonary cross section under TEM from S1 to S4; PSV – protein storage vacuoles, OB – oil bodies, Scale: 0.2–2 µ* **(B)**.

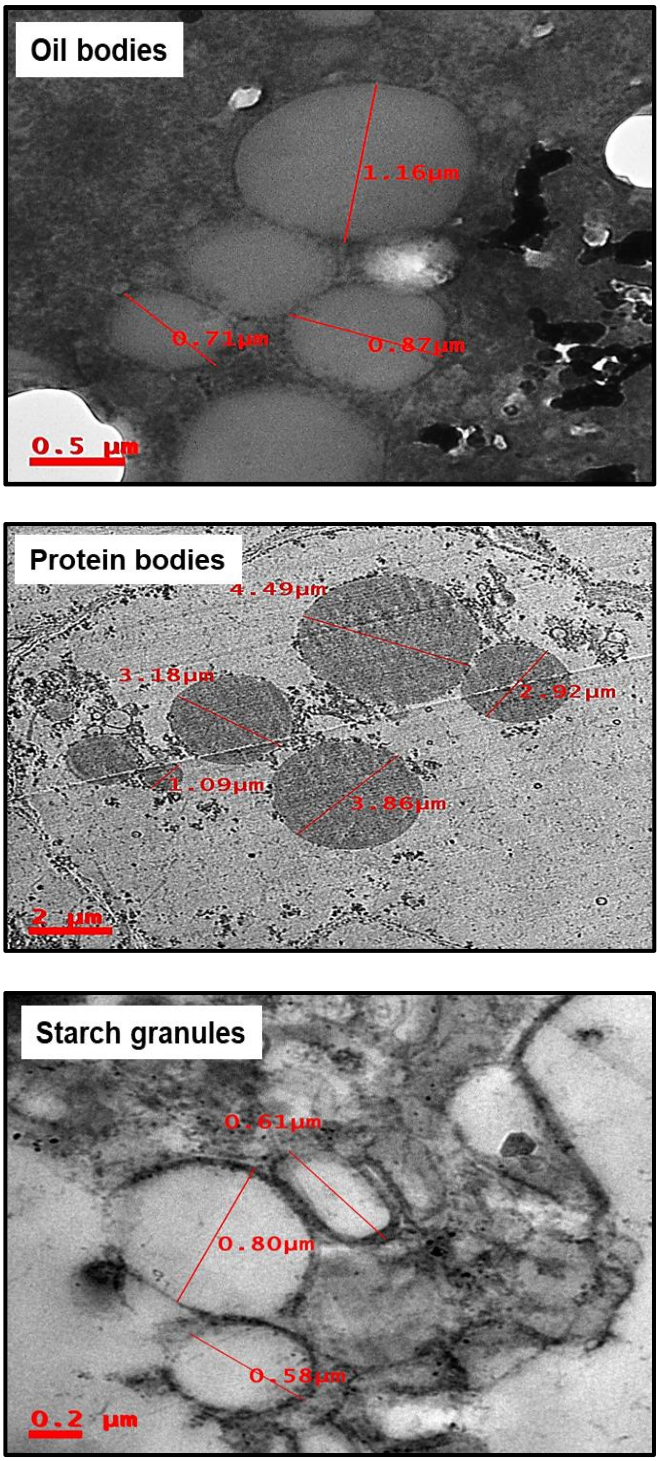


Figure 3.9: *The oil bodies, protein bodies, and starch granules of P. pinnata cotyledonary sections observed under TEM, Scale: 0.2–2 µ.*

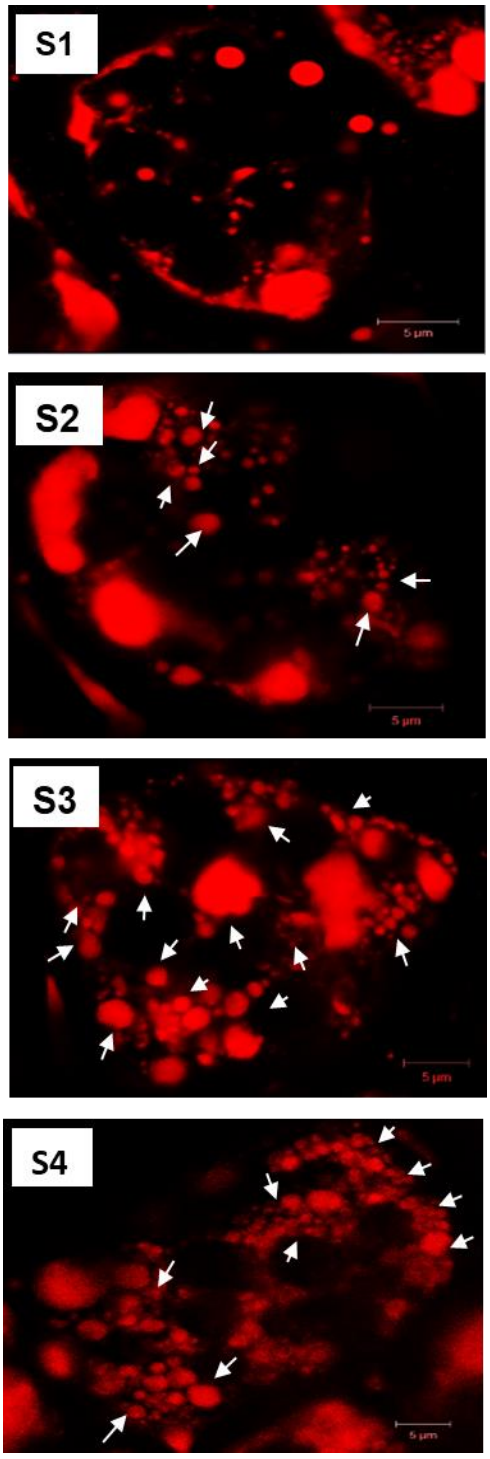


Figure 3.10: *Nile red staining of seed sections observed under confocal microscope. Oil bodies are marked by white arrows indicating their accumulation, Scale: (5µ).*

Stages	Oil bodies		Protein bodies		Starch granules	
	Size (µ)	Count Cell^{-1}	Size (µ)	Count Cell^{-1}	Size (µ)	Count Cell^{-1}
S1	—	—	0.8–3.9	4–5	0.5–0.7	0–2
S2	0.5–1.0	40.0–50.0	0.8–3.9	5–7	0.5–0.7	1–5
S3	0.5–1.5	70.0–100	0.8–4.1	5–10	0.5– 1	3–5
S4	0.5–2.0	120–180	0.8–4.1	6–10	0.5– 1	3–5

Table 3.5: *Stage wise changes in the size and the count per cell of different types of storage granules observed in P. pinnata cotyledonary sections (S1 to S4). The sizes of are in micron (µ) and values are represented as ranges observed in triplicates.*

3.7 Major soluble sugars of developing seeds

Sucrose, glucose and fructose content in the *P. pinnata* cotyledons as well as seed coat were quantified based on the HPLC analysis (Figure 3.11). Sucrose content was higher in seed coats than other sugars (fructose and glucose) at S1 when compared to that of cotyledons. Calculations were done based on the peaks and internal standards used for every sugar individually or combined (Figure 3.12). The concentration of sucrose increased and stabilized from S2 to S4 in cotyledons whereas, it decreased gradually in seed coats. Both glucose and fructose were high at younger stages (S1) which showed gradual decrease from S2 to S4 in both cotyledons and seed coats (Figure 3.12). The hexoses (total of fructose and glucose) showed a similar pattern in their concentrations but lesser in seed coats when compared to the cotyledons in the developing *P. pinnata* seeds. However, during the initial stage (S1) fructose and glucose concentrations were similar in both cotyledons and seed coats (~10 mg g^{-1} FW). Fructose content decreased abruptly from S2 to S4 in cotyledons (Figure 3.12). The total content of these three major sugars also decreased gradually with maturation. However, the concentration of major soluble sugar in seed coats were slightly higher when compared to that of cotyledons (Figure 3.12). Major differences in the concentrations of the sugars between seed coat and cotyledons were observed with sucrose content only (8 mg g^{-1} FW in cotyledons and 20 mg g^{-1} FW in seed coats). However, younger seed coats of the developing *P. pinnata* seeds (S1) were rich in sucrose content (Figure 3.12).

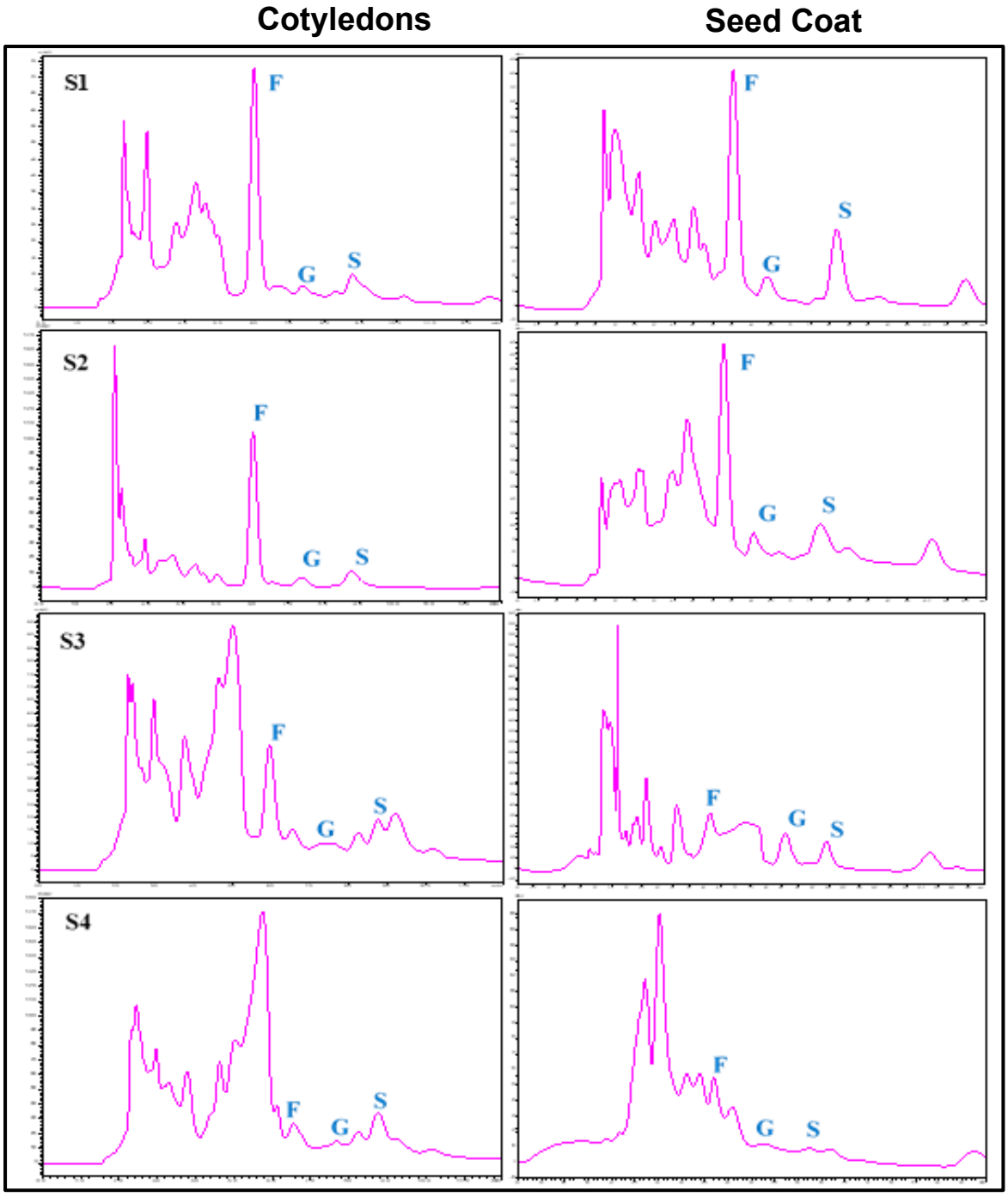


Figure 3.11: *HPLC peaks of the major soluble sugars of P. pinnata seeds (cotyledons and seed coat separated) from S1 to S4. "F" – Fructose, "G" – Glucose, "S" – Sucrose.*

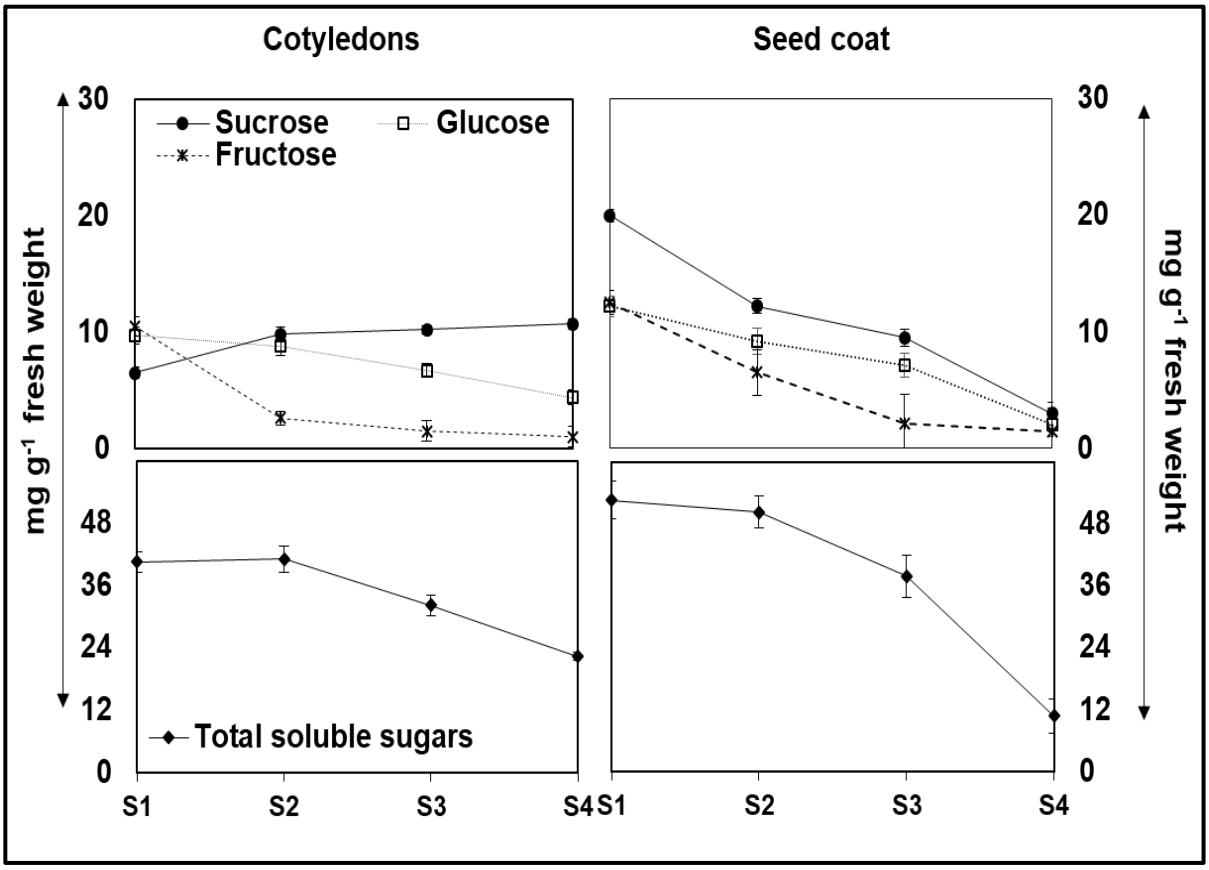


Figure 3.12: *Quantification of the soluble sugars of developing P. pinnata seeds (cotyledon and seed coat separated). Values are represented in mg g^{-1} FW of the sample analyzed. Data are given as means ± SD; (n=3).*

3.8 The stage wise changes in the activities of invertases

P. pinnata seed invertases were measured at different pH based on their types in each selected stages of development. The activity of these enzymes was quantified based on the fructose content observed at 0 min and 5 min of the essay mixtures (Figure 3.13). The activities of invertases varies in all the developing stages (Figure 3.14). Developing seeds at (S1) showed higher activity of acid invertases (pH – 4) than the neutral and alkaline invertases (pH – 7 to 7.5). Younger developing stages (S1 – S2) showed the highest activity of acid invertases which sharply decreased from S3 to S4. Although the pH of the neutral and alkaline invertases were only slightly different, their activities based on the concentrations varies largely between S1 to S4 (Figure 3.14). Both neutral and alkaline invertases decreases with maturation and the slightly alkaline invertases showed a sharp increase at S2 (Figure 3.14). The changes in the expression pattern of genes related to different invertases were analyzed with using the transcriptome data available and published earlier (Sreeharsha et al., 2016). Transcripts of some of the genes encoding the different types of invertases were successfully amplified and the expressions were studied in all the selected developing stages (S1 to S4). There were two cell wall bound, four vacuolar and one neutral invertase related transcripts whose expressions were studied (Figure 3.17). Cell wall bound invertases (PpcwINV1 and 2) and vacuolar invertases (PpvINV1, 2 and 3) showed an increased expression at younger stages (S1 to S2) whereas the neutral invertase transcript PpnINV1 showed much lower expression in all the stages (Figure 3.17). Downregulation of the expressed transcripts of invertases were observed in the more mature seeds (S4) (Figure 3.17).

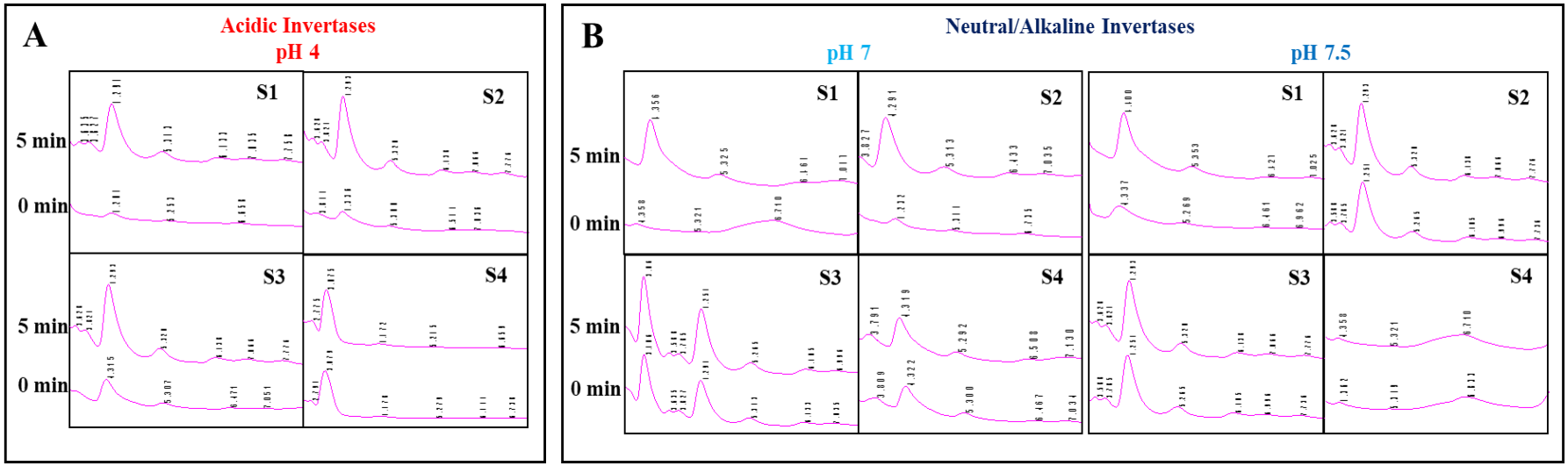


Figure 3.13: *HPLC peaks of sugars observed after the invertase assay at different pH. Calculations were based on the fructose content after 5 min of incubation for all the assay sample. Acid invertases pH 4 (A); neutral invertases pH 7, alkaline invertases pH 7.5 (B).*

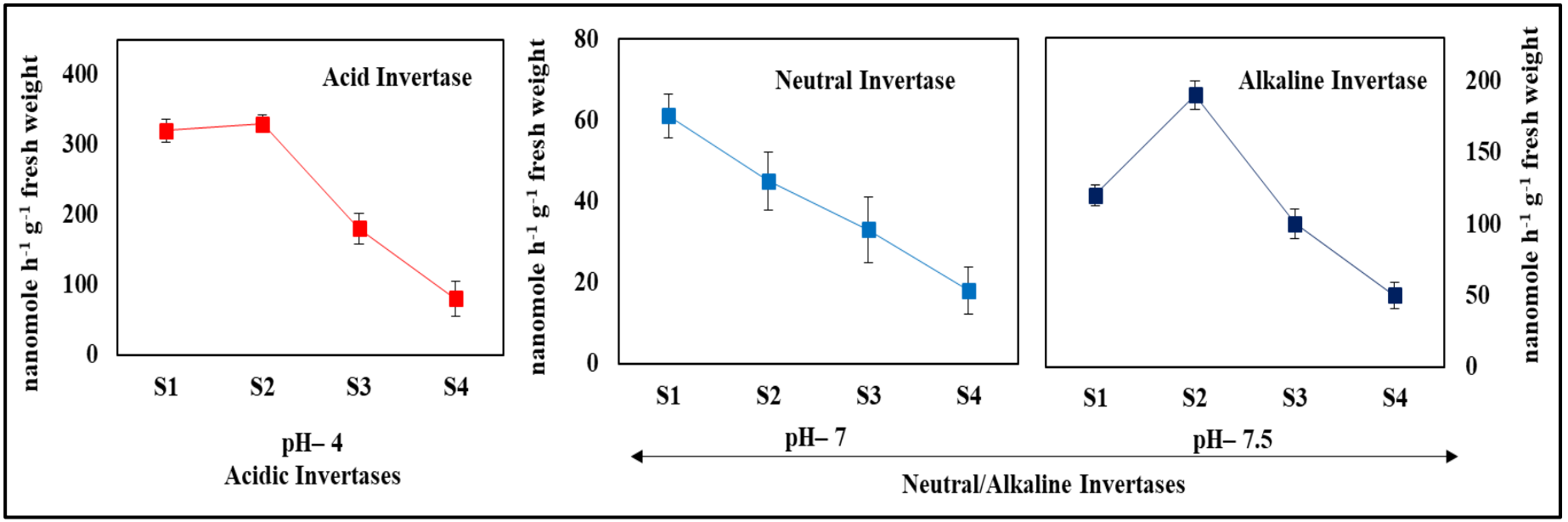


Figure 3.14: *Invertase activities (nanomole h^{-1} g^{-1} fresh weight of the seeds) at different pH in the developing P. pinnata seeds. Values are mean ± SD (n=3).*

3.9 Activities of sucrose synthase, sucrose phosphate synthase and hexokinases in *P. pinnata* seeds

The activities of SS and SPS in developing seeds were observed from S1 to S3, (Figure 3.15). SS activity increased from S1 to S2 and remained as such up to S3 whereas, the activity SPS was recorded to be gradually increased from S1 to S3 (Figure 3.15). A greater level of SS activity when compared to the SPS activity in developing seeds was observed. The expressions of transcripts related to SS (PpSUSy1, 2 and 3) and SPS (PpSPS1) upregulated from S2 to S3 (Figure 3.17).

Both the activities of hexose cleaving enzymes (GlucoK and FructoK) were increased from S2 to S3 with sharp decrease thereon at S4. The FructoK activity was approximately double of GlucoK in all the stages of seed development in *P. pinnata* (Figure 3.16). There were six and three transcripts for FructK and GlucK respectively and thier expressions were successfully recorded from S1 to S4 (Figure 3.17). The expression of the transcripts PpFRK1 and 2 were increased in S2 to S3 and the expression of PpFRK3 and 4 increased in S3 only. PpHXK2 and 3 were upregulated up to 2–fold during S2 – S3 (Figure 3.17).

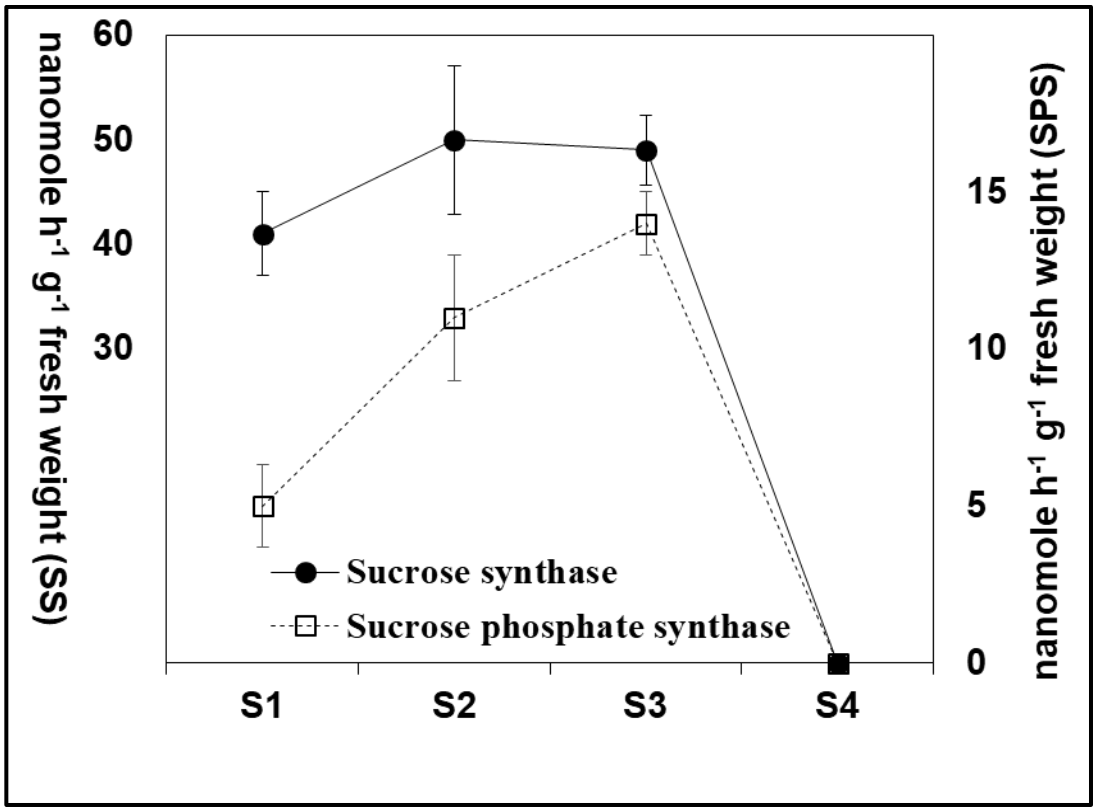


Figure 3.15: *Changes in the SS and SPS activity (nanomole h^{-1} g^{-1} fresh weight) from S1 to S4. Left and right axis represents SS and SPS respectively. Values are mean ± SD (n=3).*

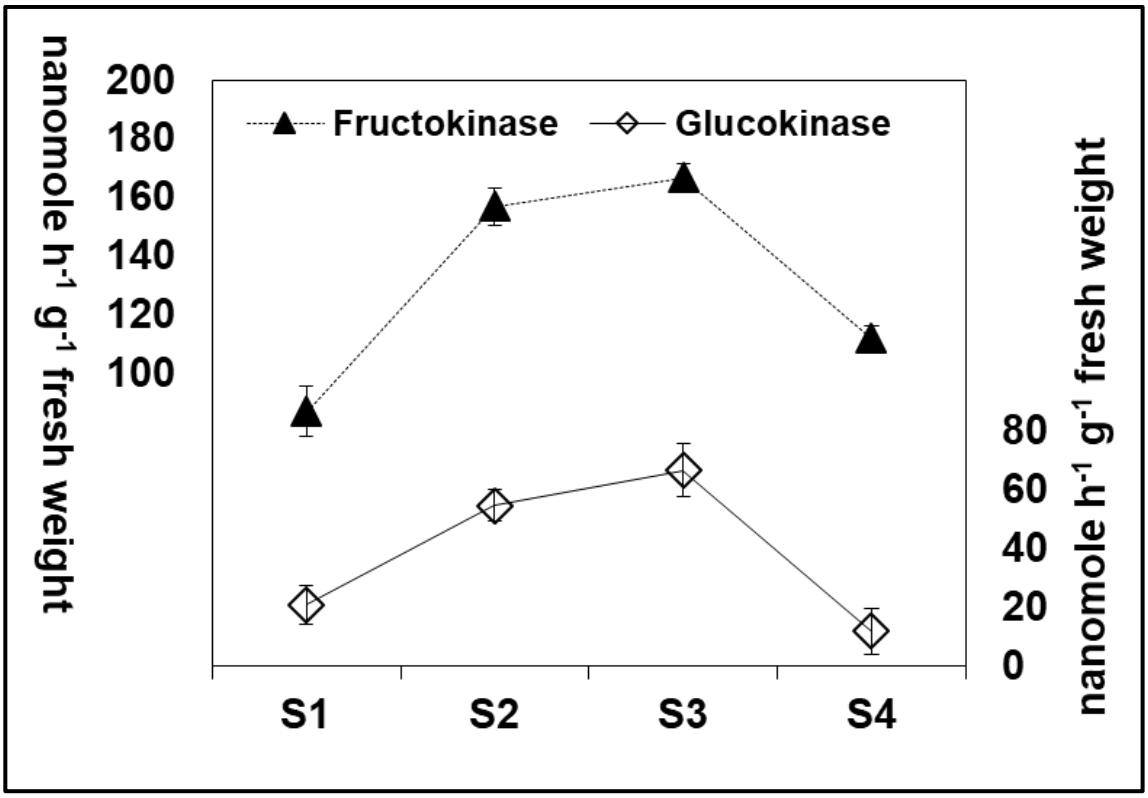


Figure 3.16: *Changes in the fructokinase and glucokinase activity (nanomole h^{-1} g^{-1} fresh weight) from S1 to S4. Values are mean ± SD (n=3).*

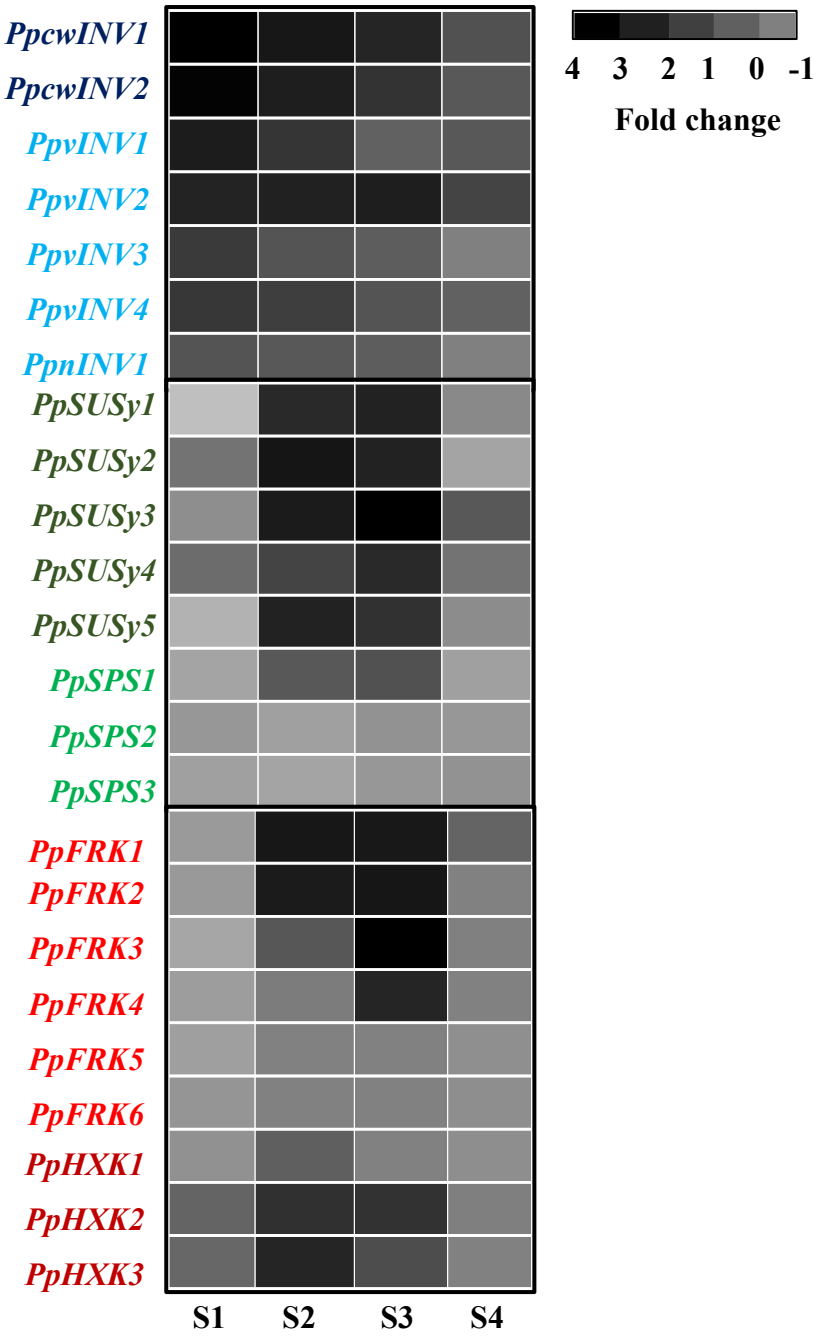


Figure 3.17: *Changes in the expression of transcripts belonging to the various enzymes of sugar metabolism in the developing P. pinnata seeds. Values are represented in log 2–fold change of gene expression (n=3).*

3.10 Complete metabolome analysis of different developmental stages of *P. pinnata* seed

GCMS analysis showed 61 metabolites which were identified in all the four stages. The metabolites were classified and documented based on their chemical structure (Figure 3.18 A) and biological functions (Figure 3.18 B). Metabolites were classified as polar and non–polar and further into fatty acids, amino acids, sugars and organic acids etc. depending on their respective chemical characteristics (Figure 3.18 A). The initial stages of development (S1 to S2) were rich in the polar metabolites, whereas in the maturing stages (S3 to S4), there was an increase in non–polar metabolites (Figure 3.18 A). The polar groups of metabolites were mainly amino acids, organic acids and sugars. Fatty acids and poly–phenols were categorized as non–polar metabolites. Majority of the metabolites detected were functionally related to protein metabolism (21 %), carbohydrate metabolism (18 %) and development (16 %). Rest of the metabolites belonged to energy (12 %), stress and defense (10 %), fatty acids (9 %), cell wall formation (7 %), hormonal responses (4 %) and secondary metabolism (3 %) (Figure 3.18 B). The dynamics of the metabolites were determined by their concentration during all the developing stages analyzed through heat map using Pearson's test in MetaboAnalyst online software (Figure 3.20). Among the sugars, ribose, glucose, erythrose, gluconic acid, inositol and ribono lactone showed higher concentrations at S1 which gradually decreased with the maturation. Glyceric acid, pinitol and mannose were abundant at S4 whereas sugars like cellobiose, arabinose and NADGA were higher during S1 to S2. In contrast to the sugars, most of the amino acids and organic acids were abundant during S3 and S4 (Figure 3.19). Concentration of Alanine, glutamate, leucine, threonine, tryptophan, valine was higher in matured stages (Figure 3.19). Organic acids including fumaric acid, keto–glutaric acid, malonic acid, pyruvic acid and succinic acid, were also abundant in S3 and S4 (Figure 3.19).

The contents of glycine, glutamine and proline were higher in S4 (Figure 3.19). Similarly, organic acids such as adipic acid, acetate, malic acid, citric acid, oxalic acid and propionic acid was high in the initial stage (S1). Also, there are certain other amino acids such as aspartic acid was more at S1, which gradually decreased to the least at S4 (Figure 3.19). Fatty acids including hexanoic acid, myristic acid and stearic acid was abundant in S3, while others such as butanoic acid and eicosanoic acid were more in S4 (Figure 3.19). The variation among distinctive metabolite profile was observed through PCA analysis where there was an overall 72.2 % (PC 1) separation among the developing stages (Figure 3.20). Separation or variation of seed metabolites among S2 and S3 was less when compared to S1 and S4 (Figure 3.20).

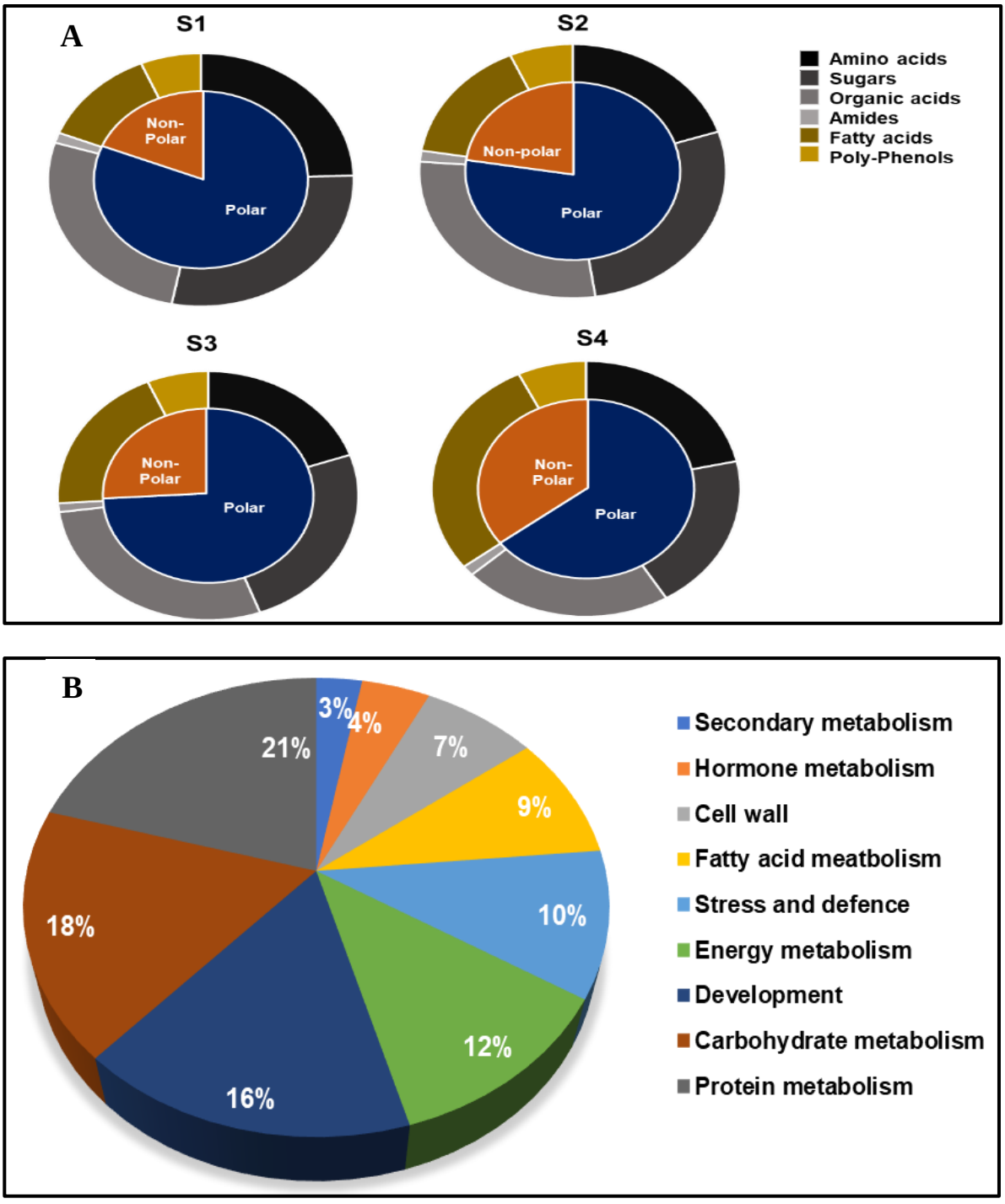


Figure 3.18: *Pie charts representing the metabolites based on the biochemical characteristics **(A)**, biological functions **(B)**.*

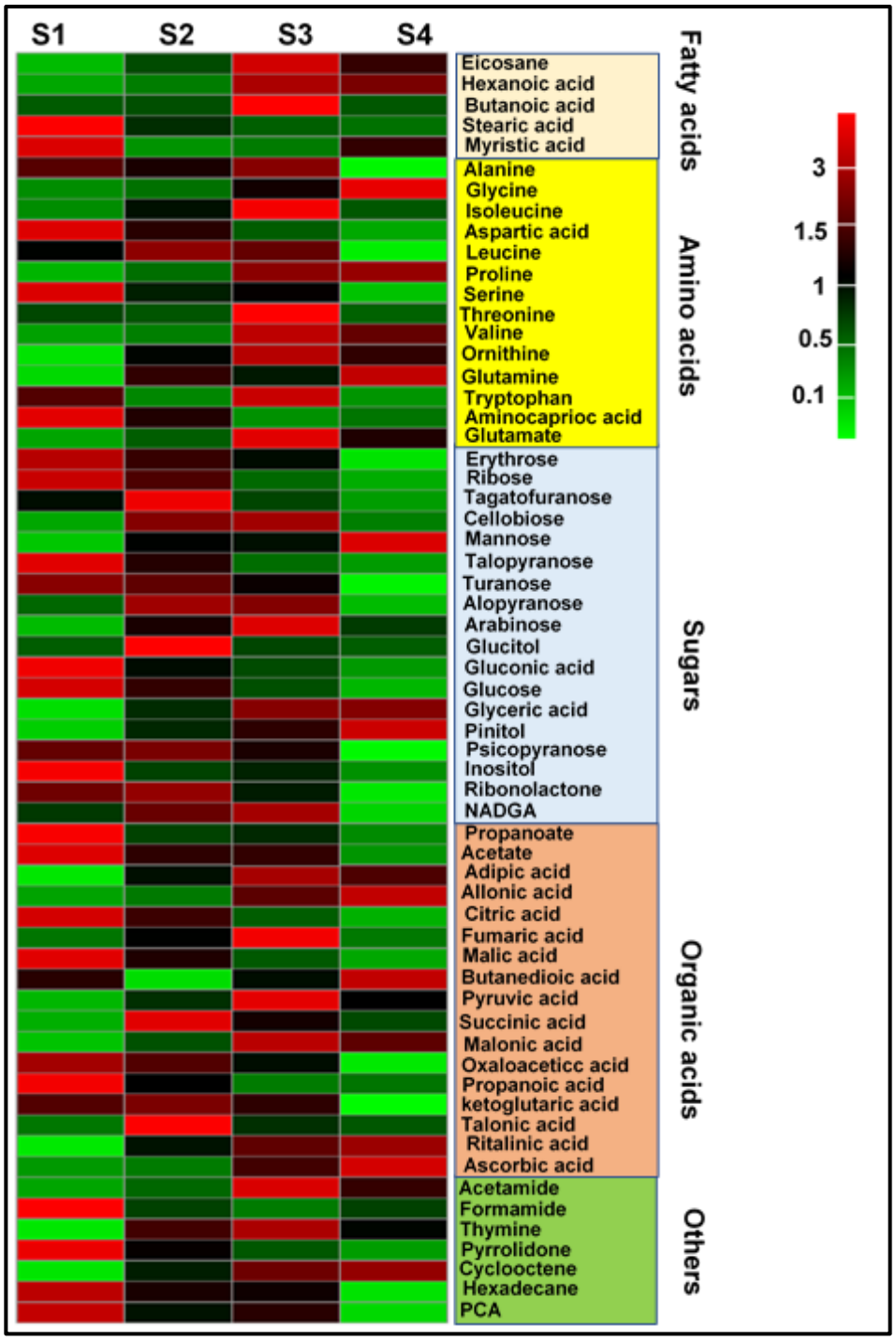


Figure 3.19: *Heat map of metabolite concentrations (µg g^{-1} FW) at S1, S2, S3 and S4. Analysis was carried out in Metaboanalyst 4.0.*

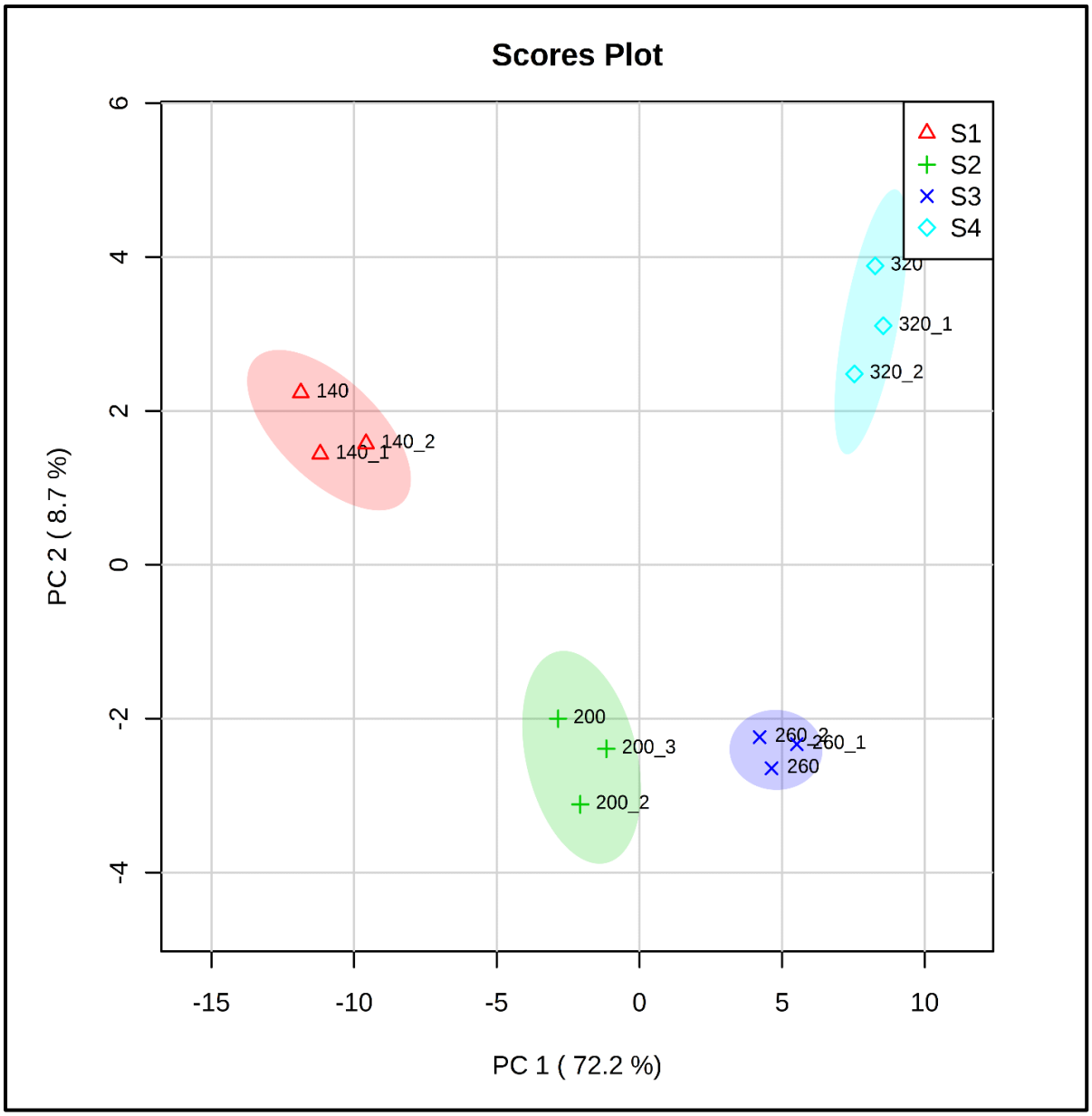


Figure 3.20: *The PCA biplot analysis of all the metabolites isolated from S1 to S4.*

3.11 Total lipid and fatty acid profile

The separation of phospholipids, DAGs and TAGs during developing stages were compared by the intensities on the TLC plates which have been loaded with equal quantity and separated with the same mobile phase. There was a visible change in the TAGs accumulation where it progressively increased as the seeds reached maturity (Figure 3.21 B). The fatty acid profile of the extracted TAGs from the TLC plates were examined with the help of GCMS analysis, which was compared with the FAMEs standards, based on which the percentage was also determined (Figure 3.21 A). The major fatty acids during initial stages of development (S1 and S2) were oleic and linoleic acid which were present in equal amounts (Figure 3.21 A). In later stages of development, oleic acid gradually increased with a significant ($P < 0.05$) decrease in linoleic acid (Fig. 7a). Furthermore, palmitic acid also decreased gradually from S1 to S4 (Fig. 7a) while there was no change in stearic acid content in all the stages. The GCMS peaks showing all the major fatty acids (a, b, c and d) of the FAMEs analyzed in all the stages were shown in Figure 3.22.

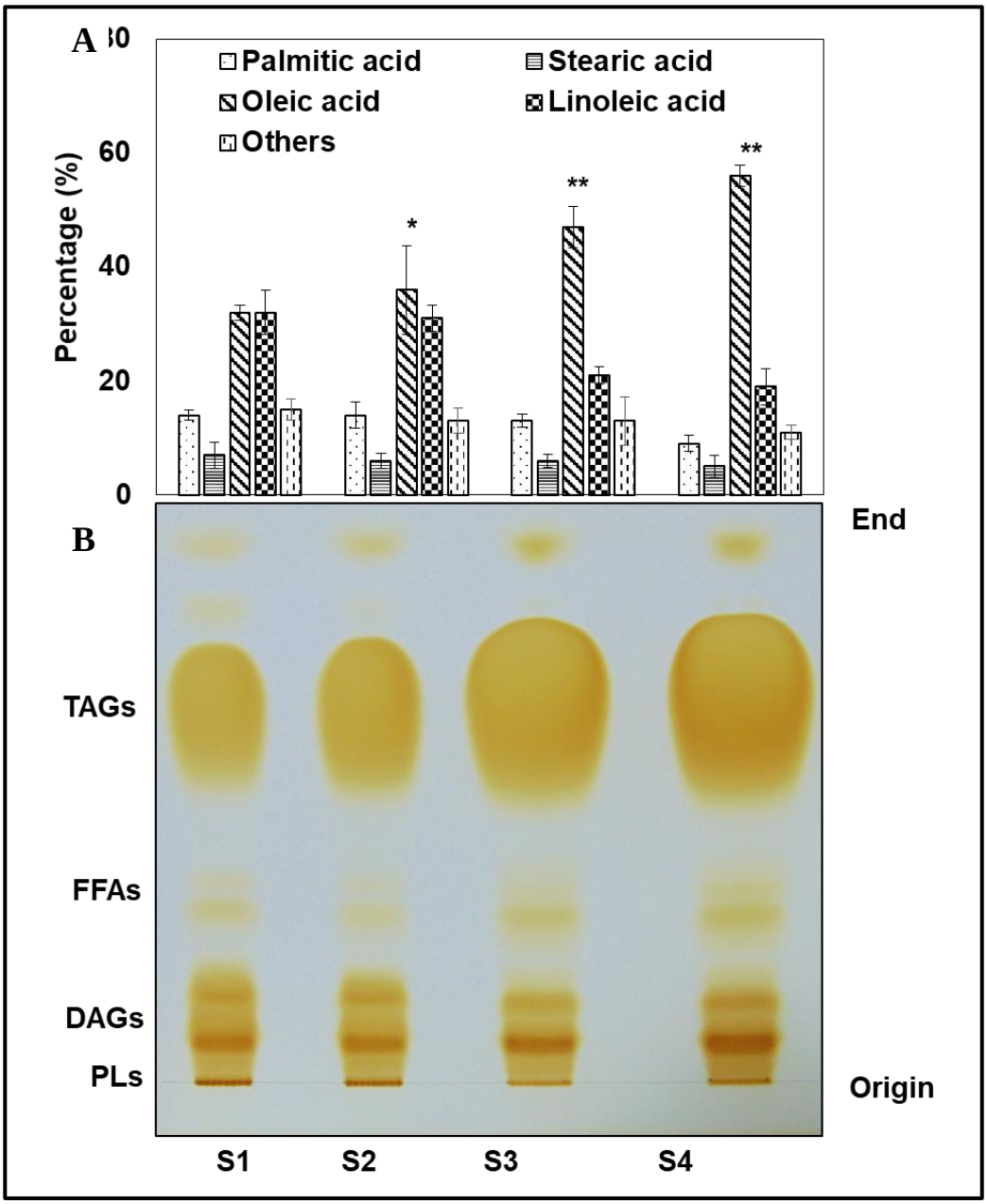


Figure 3.21: *Percentage of fatty acids accumulated in the TAGs at different developing stages* ***(A)****. Total lipid profiling through TLC* ***(B)****. Values are mean ± SD (n=3). *, ** represents significance difference (P < 0.05, P < 0.01) between the fatty acid content with respect to 120 DAA, analyzed by t–test.*

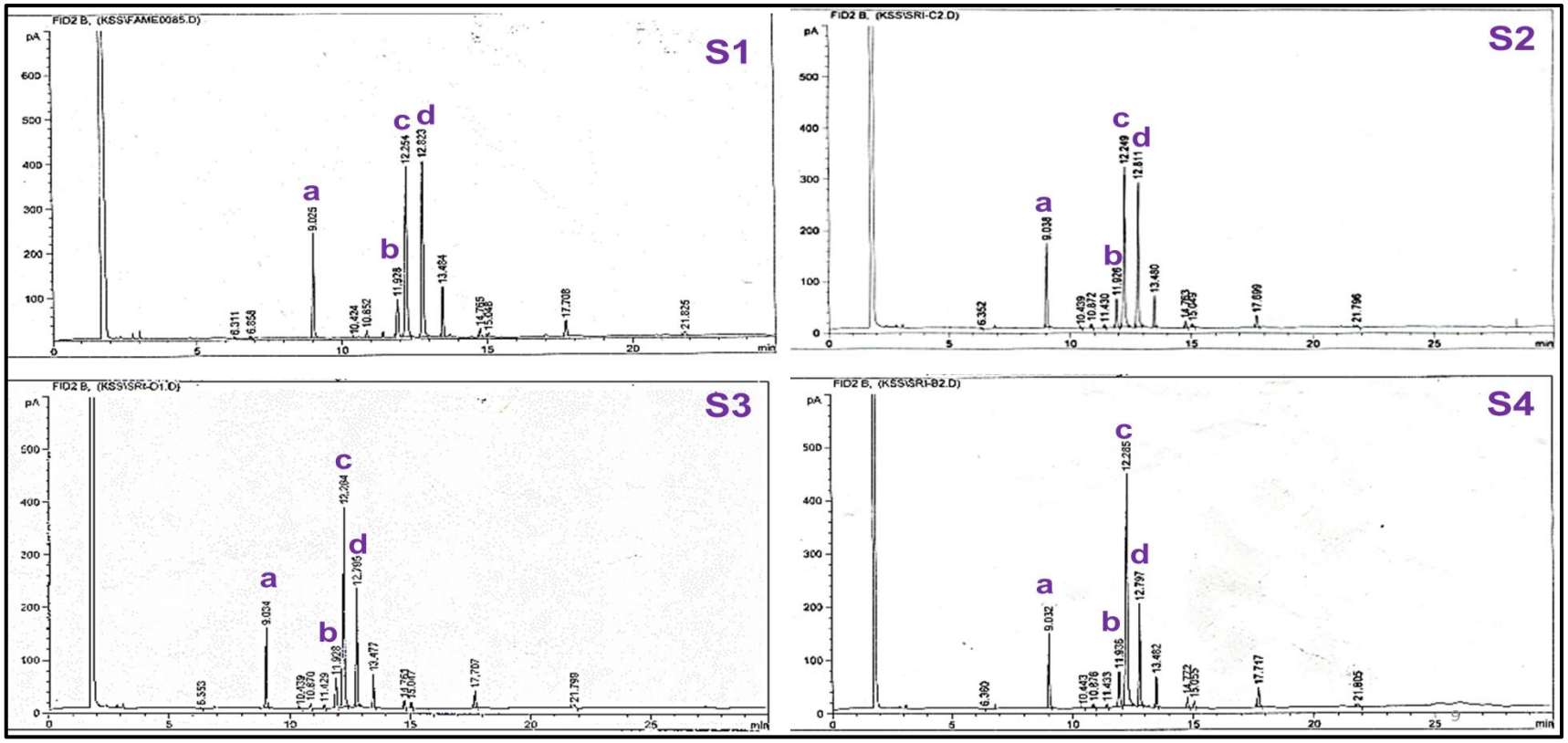


Figure 3.22: *GCMS peaks of the FAMEs analyzed.* ***"a"*** *– Palmitic acid; 16:0;* ***"b"*** *– Stearic acid; 18:0;* ***"c"*** *– Oleic acid, 18:1;* ***"d"*** *– Linoleic acid, 18:2.*

3.12 2D–proteomic profile of developing *P. pinnata* seed

The analysis of seed protein expression patterns using 2D SDS gel were quantified at different developing stages of *P. pinnata* seed. More than 300 spots were reproducibly detected in all the gels with Image Master 2D Platinum software and 125 spots matched in all the four stages. The spot distribution and intensity of those spots at each stage were analyzed through PCA analysis (using Metaboanalyst 4.0) which illustrated the variation in the form of percentage separation (Figure 3.24). There was overall 64.2 % separation among the spots observed in all the stages and the graph clearly showed the protein profile of S4 was different compared to the younger stages (S1 – S2). The profiles of S2 and S3 were more similar with least separation (Figure 3.24). Based on the percentage spot volume in the S1 stage, the significantly upregulated, unchanged or downregulated spots throughout developing stages were determined (Figure 3.23). In order to know regulatory proteins of seed development, some of the spots were selected and successfully identified using MALDI MSMS (Table 3.6). The selected regions (total of 6) in the gel pictures which represent all of the differentially regulated spots were shown and changes in some of the spots among the different stages were represented along with the reference gel in (Figure 3.25). The functions of the identified proteins were studied and further classified based on the available literature (Figure 3.26). Most of the proteins identified and matched with other plant species were related to *Arabidopsis* (27 %), *Brassica* (23 %), and *Glycine max* (21 %) (Figure 3.27 A). Functional classification of the identified spots was also presented in Figure 3.26 B. Based on their metabolic function the proteins were grouped in three major categories. Proteins controlling developmental mechanism (Figure 3.27 A), photosynthesis and energy (Figure 3.27 B) and lipid metabolism (Figure 3.27 C).

Overall, the upregulated spots from S1 to S4 were identified as Accase, lipid transfer like protein VAS, and stearoyl–ACP desaturase. Protein spots upregulated or unchanged during S1 to S3 were ATP synthase subunit beta, G3P dehydrogenase, PEP kinase, and GTP binding protein SAR1b. Gradually decreasing spots from S1 to S4 were Midasin and AP2 subunit alpha (Figure 3.27). Only few spots were observed in S1 and S4 specifically (Figure 3.28). The protein spot identified as ELF 4 was observed only in S1 whereas, Allergen Len and D19 (LEA_5) was observed only in S4 (Figure 3.28). Details of all the identified spots were documented in Table 3.6.

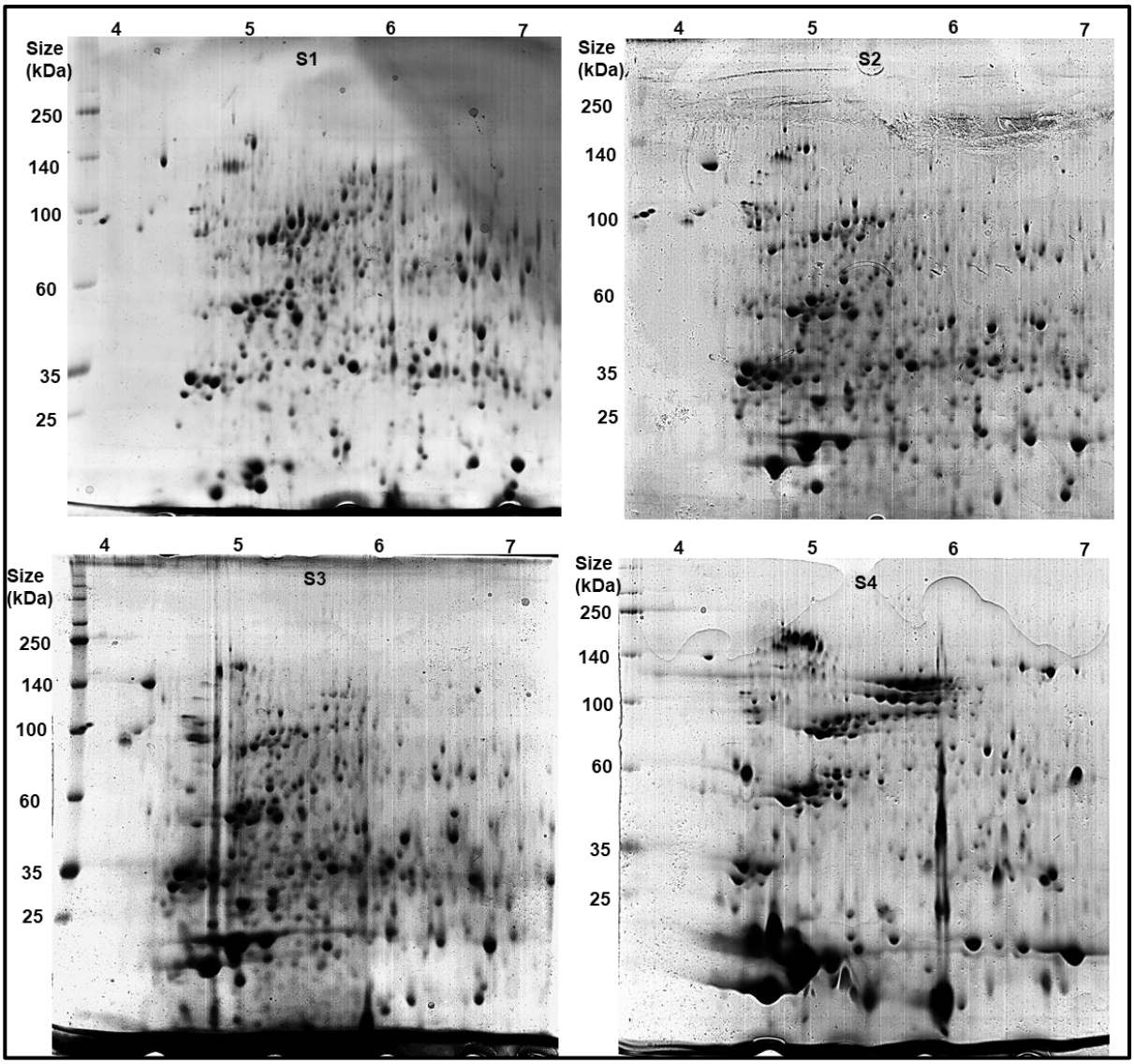


Figure 3.23: *2D SDS gel picture of all the developing stages (S1 to S4).*

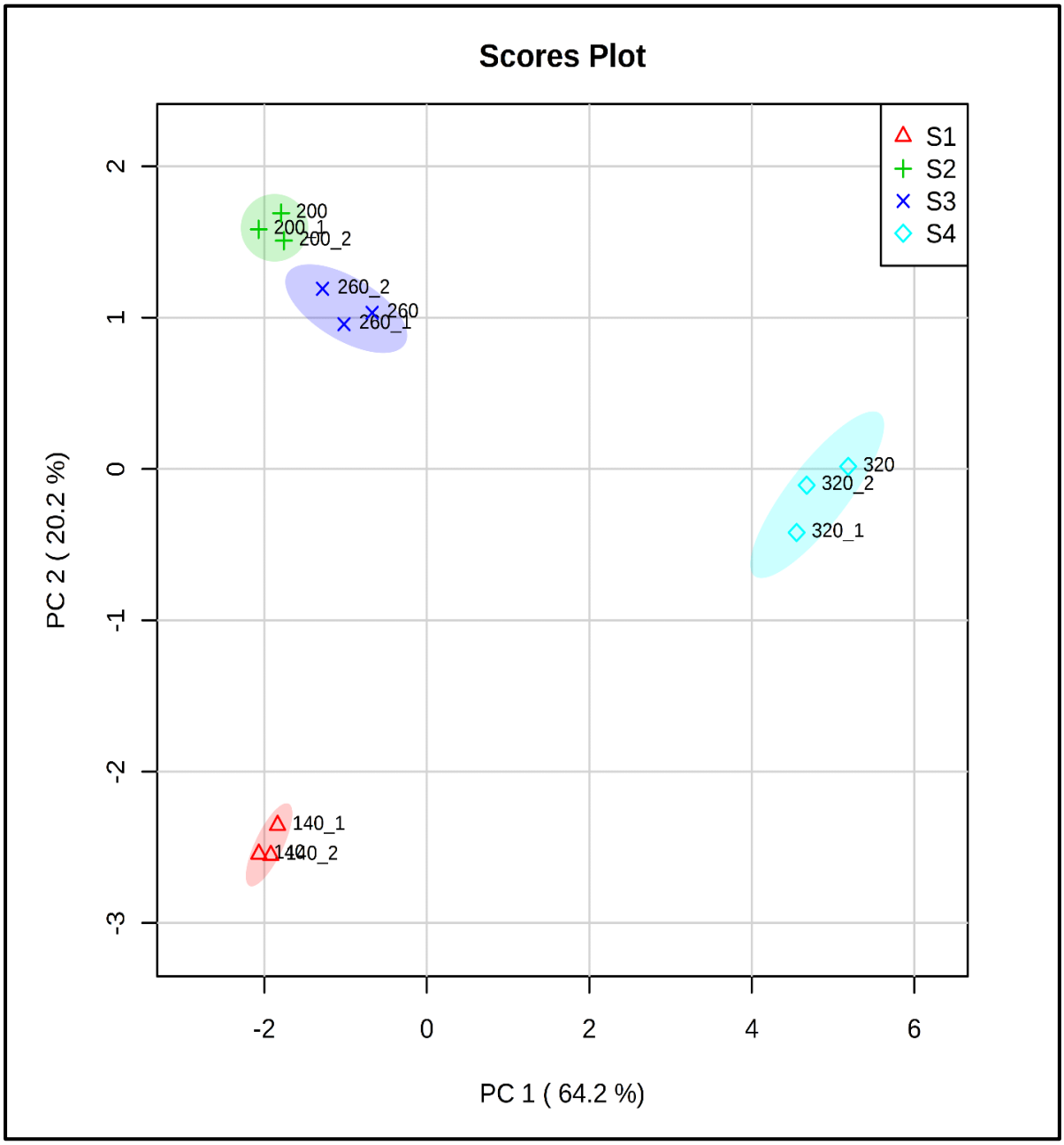


Figure 3.24: *The PCA biplot analysis of all the gel spots observed with Image G (S1 to S4).*

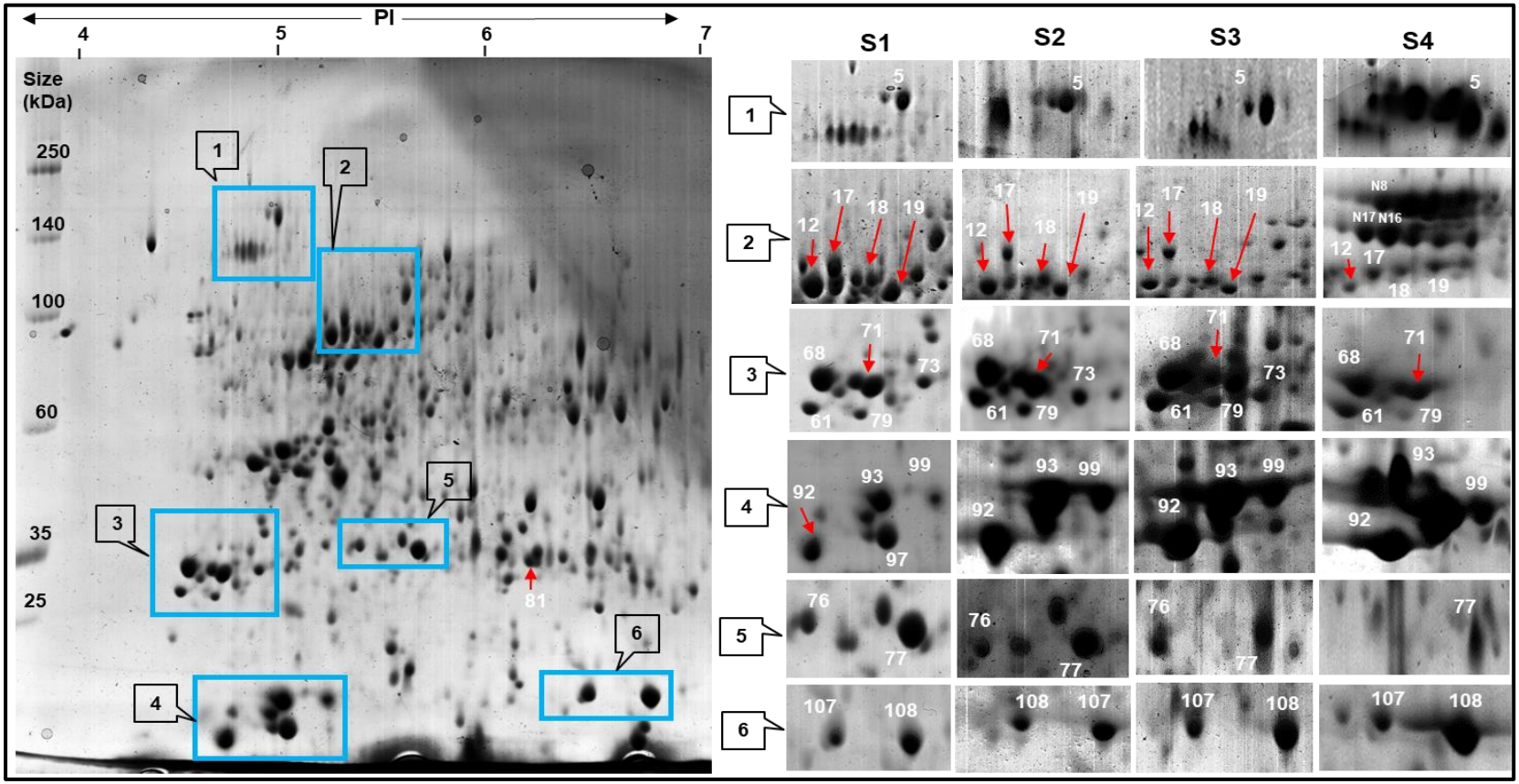


Figure 3.25: *Identified spots with the reference Gel. Selected regions with spot numbers are given for all the developing stages (S1 to S4).*

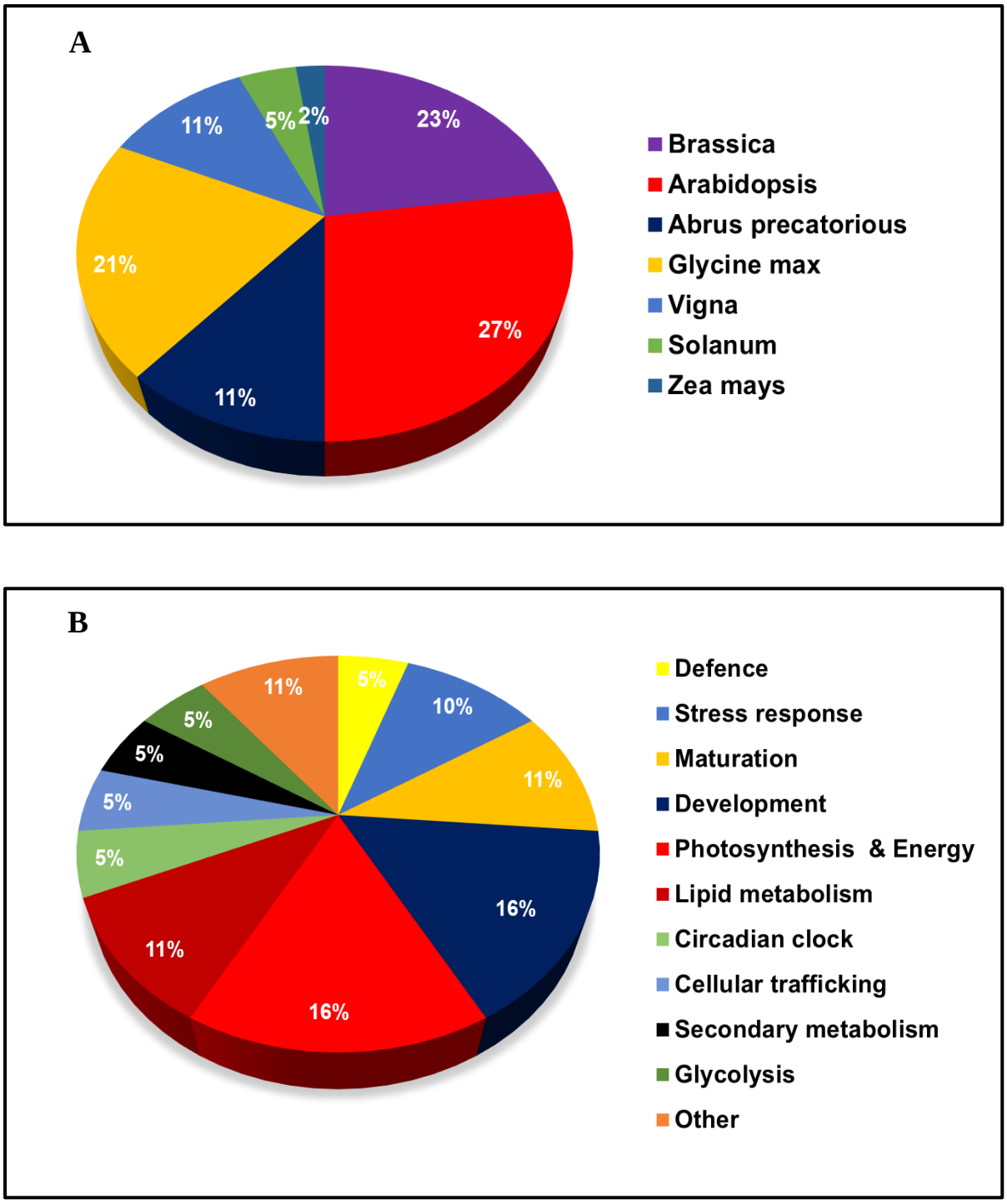


Figure 3.26: *Pie charts representing the identified proteins based on the similarly of protein accessions of the reference organisms* **(A)**. *Biological functions* **(B)**.

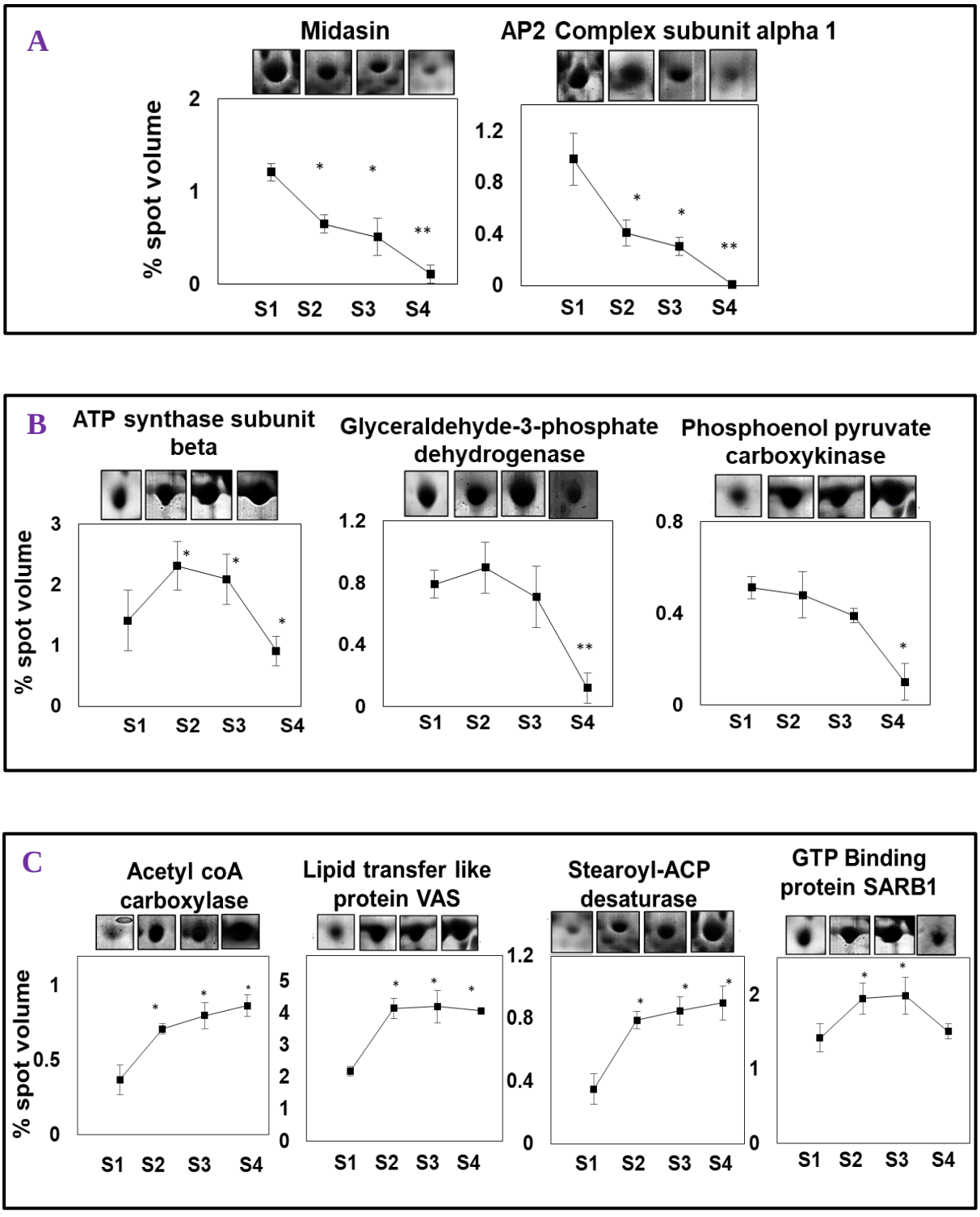


Figure 3.27: *Changes in the expression of key regulatory proteins from S1 to S4.*

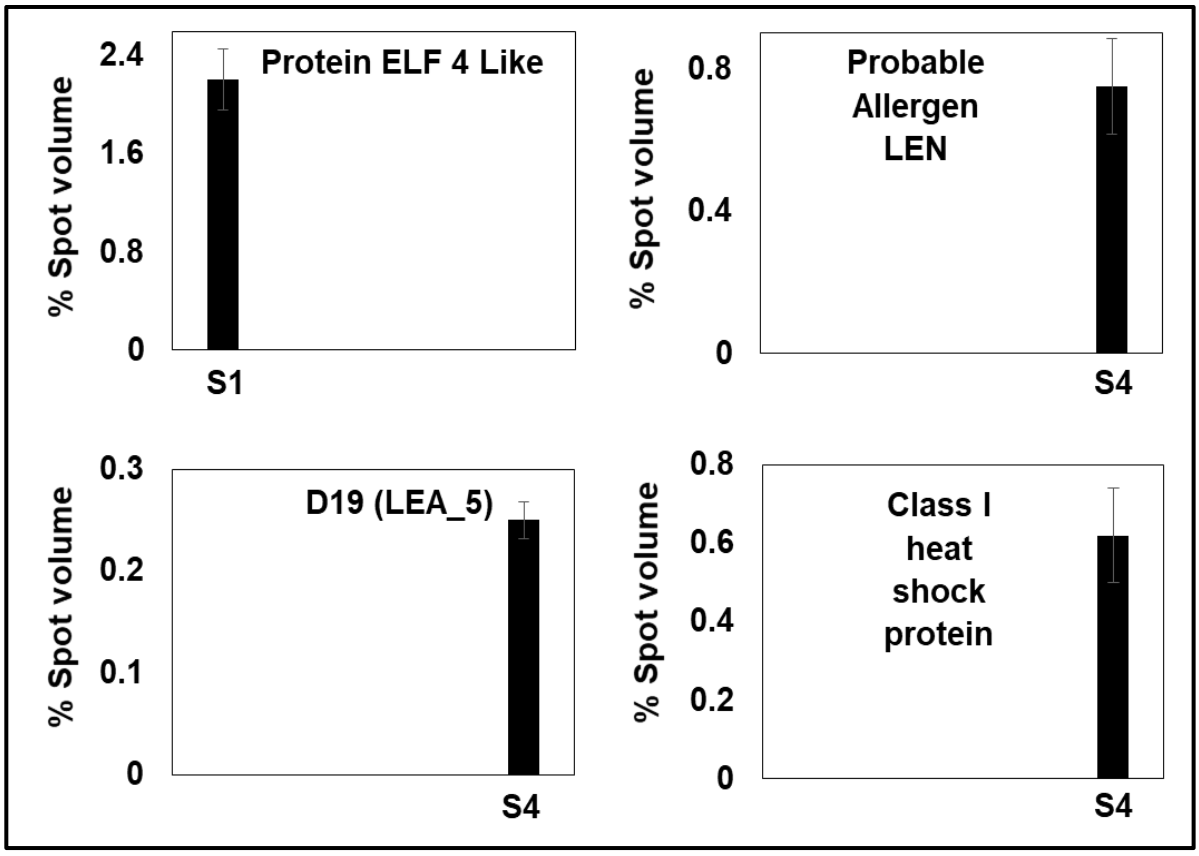


Figure 3.28: *Stage specific proteins observed at S1 and S4.*

Sp. ID	Protein identified	MW in gel	MW	Known Function	Organism related to	Accession no.	MS/MS Score	Peptide sequences matched
5	Acetyl coA carboxylase	200kD	226 kD	Fatty acid biosynthesis	*Sesamum indicum*	XP011083400.1	98	29
6	1-cys peroxiredoxin B	35kD	24.4kD	Vitamins	*Oryza sativa*	REHYB_ORYSI	67	8
12	Midasin	95kD	116kD	Female gamete and seed development	*Arabidopsis thaliana*	MDN1_ARANTH	94	34
17	AP2 Complex subunit alpha 1	91kD	112kD	Vesicular transport and development	*Arabidopsis thaliana*	AP2A1_ARATH	94	12
18	Phosphoenol pyruvate carboxykinase	95kD	73.7kD	Photosynthesis carbon fixation and energy	*Zea maize*	PCKA_MAIZE	64	12
19	Probable disease resistance protein	100kD	104.5kD	Defence and disease resistance	*Arabidopsis thaliana*	PX24L_ARATH	108	31
61	Lacto-glutathione Lyase	59 kD	55kD	Stress response	*Abrus precatorious*	XPOO9339247.1	96	7
64	Glyceraldehyde-3-phosphate dehydrogenase	45kD	36.5kD	Glycolysis and energy metabolism	*Arabidopsis thaliana*	G3PC1_ORYSJ	278	7
68	ATP synthase subunit beta	35kD	53.3kD	Photosynthesis electron transport and energy	*Brimeura amethystina*	ATPB_BRIAM	68	8
71	Ent-copalyl diphosphate synthase 1	35kD	99kD	Secondary metabolism	*Oryza sativa*	CPS1_ORYSJ	88	8
73	Soluble Inorganic Pyrophosphate	40kD	32.5kD	Vacuolar transport	*Glycine max*	XPO28204350.1	138	12
77	GTP Binding protein SARB1	30kD	22.02kD	Vesicular transport of storage bodies	*Arabidopsis thaliana*	SAR1B_ARATH	74	9
79	DUF4283 Domain containing protein	40kD	39.6kD	Stress and defence	*Glycine max*	GAV92775.1	184	14
80	Fructose-bisphosphate Aldolase	40kD	38.6kD	Glycolysis, Gluconeogenesis, Energy	*Vigna angularis*	TKY56421.1	255	13
81	Heat stress transcription factor c-1b	30kD	27.4 kD	Abiotic stress regulator	*Oryza sativa*	HFC1B_ORYSJ	97	17
82	Triose-phosphate Isomerase	35kD	27.4kD	Glycolysis	*Abrus precatorius*	XPO27339977.1	258	16
83	60S ribosomal protein	35kD	38.8kD	Translation	*Durio zibethinus*	XPO22727416.1	79	21

Table 3.6: *Details of the identified spots with MALDI MSMS.*

Sp. ID	Protein identified	MW in gel	MW	Known Function	Organism related to	Accession no.	MS/MS Score	Peptide sequences matched
84	Peptidyl serine alphagalctosyltransferase	35kD	84kD	Cell wall protein	Oriza sativa	SRGTILCHLRE	152	17
91	Annexin like protein	35kD	35.9kD	Stress and defence	Abrus precatorius	XPO27355675.1	159	10
92	Lipid transfer like protein VAS	10kD	17kD	Lipid transfer metabolism	Arabidopsis thaliana	VAS_ARATH	90	8
97	Protein ELF 4 Like	15kD	14.07kD	Circadian clock	Arabidopsis thaliana	EF4LT_ARATH	58	2
98	Integrin linked protein kinase 1	25kD	27.57kD	Signal transduction (Plant defence response)	Cucurbita maxima	XP_022970695.1	101	6
99	Stearoyl ACP Desaturase chloroplastic	35kD	43.5kd	Fatty acid desaturation	Arabidopsis thaliana	AIN52151	108	14
102	Alcohol dehydrogenase	35kD	41.6kD	Glycolysis	Glycine max	XPO19417469.1	176	4
103	Alcohol dehydrogenase2	35kD	41.7kD	Glycolysis	Vigna radiata	TKY74163.1	233	3
107	Class I heat shock protein	10.5kD	17.7kD	Heat shock and stress	Solanum pennelli	XP_015078018.1	88	7
108	Class I heat shock protein	10.3kD	17.8kD	Heat shock and stress	Solanum lycopersicum	HSP11_SOLLC	66	6
N8	4-hydroxyl-3-methyl diphosphate synthase	80kD	74kD	Secondary metabolite	Triticum	EMS56222.1	99	10
N12	ATP dependent RNA Helicase DExH	100kD	129.3kD	Stress Response	Arabidopsis thaliana	MER3_ARATH	66	11
N14	Maturin	100kD	119.08kD	Growth and development	Oryza sativa	SPT16_ORYSJ	67	8
N16	Calcium dependent Serine/Threonine Kinase	61kD	59kD	Signal transduction	Arabidopsis thaliana	KAB2594919.1	122	12
N17	Allergen LEN	90kD	47.7kD	Allergen protein in non-edible plants	Lens culineris	CAD87730.1	112	8

Table 3.6 (continued): *Details of the identified spots with MALDI MSMS.*

CHAPTER 4
DISCUSSION

4.1 The vegetative growth physiology and photosynthetic characteristics of *P. pinnata*

The overall vegetative development in *P. pinnata* has been initially studied and observed before carrying out the further objectives on reproductive development. The tropical tree, *P. pinnata* shows a dense canopy with compound leaves which bears large leaflets. Vegetative growth and development in *P. pinnata* can be observed annually with shading and re–emergence of leaflets on the tertiary branches. The rate of photosynthesis along with transpiration rate and quantum yield of PSII suggest that the tropical tree *P. pinnata* has a probability of high efficiency towards atmospheric CO_2 assimilation (Figure 3.2; Table 3.1). The biochemical analysis of foliar leaf samples showed an increased starch production when compare to the total soluble sugars. Carbohydrates are the major photo assimilates which are stored mainly in the form of starch in *P. pinnata* leaves (Figure 3.1). Apart from the carbohydrates, there are also significant amount of free amino acids detected in the foliar parts of *P. pinnata* (Table 3.3). Amino acids such as asparagine, glutamine are very common free amino acids of leafy parts in many plant species. The amino acids directly or indirectly regulate nitrogen metabolism and other signaling functions in the leaves (Ayalew et al., 2016; Guo et al., 2021). Amino acids such as arginine, glutamine, glycine and glutamic acids higher in concentration in the foliar parts of *P. pinnata* (Table 3.3) indicating their possible role in the various physiological processes. Moreover, the distribution of stomata on the ventral surface of the leaf surface area shares the characteristics of typical leaf anatomy showing high efficiency towards photosynthetic carbon assimilation (Figure 3.2). The major overall functions in the photosynthetic parts of *P. pinnata* tree species were thoroughly studied with protein profiling which also helped in the assorting the proteins to various crucial pathways such as

amino acid metabolism, carbon fixation, cell division, glycolysis, photosynthesis, transport and secondary metabolites (Table 3.4). The major proteins specifically belonging to certain metabolisms such as carbon fixation through photosynthesis are of great importance because of their unique nature and participation in many futuristic adverse climatic conditions such as drought, salinity or elevated CO_2 etc. Understanding the mechanisms related to carbon assimilation with photosynthetic efficiency determines the characteristics of the stress tolerance in many plant species (Rodriguez et al., 2008). Since *P. pinnata* has been considered as a multipurpose tree with many unique areas of research (Table 1.1), a detailed study on its foliar components will be a crucial to understand the basic physiological and biochemical characteristics associated with the growth and development. Further, these data on foliar physiological and biochemical characteristics provide crucial information on the reproductive developmental events in the tree species *P. pinnata.* Protein profiling of the green leaves which can be also considered as the source tissues, is another effort to expand the knowledge on various metabolisms controlling the growth and physiology of the oil yielding tree species. Overall, the successful identification of proteins extracted from the leaves of *P. pinnata* will assist a new approach for further relevant studies related to vegetative growth mechanism which proceeds the reproductive phase by the onset of flowering. However, being a prominent biofuel species, the focus of this thesis is on the developmental characteristics of oil producing seeds and their regulation throughout the process of seed development. After a brief understanding of vegetative growth and phenology, we have focused on the unfolding of various crucial mechanisms controlling the seed development in *P. pinnata* which is not yet mined or studied earlier (Table 4.1). The objectives of this thesis is to meet the gap of earlier findings and the targeted future research in oil biosynthesis of *P. pinnata*.

4.2 An overview on changes in the characteristic of developing *P. pinnata* seeds

Non–edible tree legumes and other non–legumes are distributed throughout the continent, and following their seed development is as fascinating as is laborious due to the lengthy life cycle and prolonged period of seed development. The limited available literature on this tropical tree legume *P. pinnata* has been a challenge as well an opportunity for conducting major experiments on its seed oil research. Our study on seed development took two sessions of 12 months each to be completed (Singha et al., 2019). Previous other findings from the available literature on characteristics of seed development have achieved few milestones on physiological and biochemical changes including genetic regulations (Table 4.1). Our current outcomes provided additional insights focusing on the seed oil biosynthesis right from the initiation to metabolic and proteomic regulations. A comparative table of the available literature and previous findings on various aspects of *P. pinnata* seed development is shown in Table 4.1. Being a tropical annual tree, flowers took 1 month to fully bloom and pods took another 10 months to completely mature for *P. pinnata* (Figure 3.1 B) (Scott et al., 2008). The non–endospermic seeds of *P. pinnata* resembles most of the leguminous species where the cotyledons are the main storage regions. The basic seed morphology and biochemistry of developing *P. pinnata* seed share certain similar developmental regulatory processes with other legumes where, the pods were nearly flat in younger stages with a tiny developing seed inside and the pod thickness increased as the seed developed, finally filling the pod (Wright and Lenssen 2013). The duration of seed development in *P. pinnata* is a prolonged process making it complex to understand the seed filling and oil accumulation patterns in this imminent biofuel tree species.

Thus, knowledge on physiological and molecular dynamics of this lengthy seed development will boost the attempts to improve the seed productivity towards potential biofuel feedstock. The current study thoroughly examined the seed developmental processes of *P. pinnata* and deduced the physiological, biochemical and molecular changes through metabolomics and proteomic approaches. Our present studies also provide insights into the photo-autotrophic nature of *P. pinnata* seed.

Legume seed development and differentiation is characterized by three rapid growth phases separated by two lag phases (Weber et al., 2005). Majorly, the growth phases are marked by the development of endosperm, seed coat and embryo associated with maturation and cell expansion. The differentiation of processes during seed development is a sequential process which involves active mitotic cell division, sucrose uptake, cell expansion, greening and gaining of photosynthetic activity as well as accumulation of storage products (Baud et al., 2002). For oil seeds, the differentiation phase is of special interest because during this stage a regulatory network initiates the accumulation of storage products. The developing seeds of *P. pinnata* are green in color at younger stages (S1–S2) which turned brown during maturation with a gradual increase in seed weight, storage products and lipid content but decrease in moisture content as well as the weight of seed coat (Figure 4.1). The importance of maternal seed coat in providing the necessary sources of filial cotyledons for development and controlling germination is well known in legumes (Weber et al., 2005). Seed coat percentage in terms of its structural weight is high during younger stages of development which indicated its active role in maintaining seed development. During natural seed maturation process, there is a gradual increase in seed desiccation, wherein the water content declined significantly, (Angelovici et al., 2010).

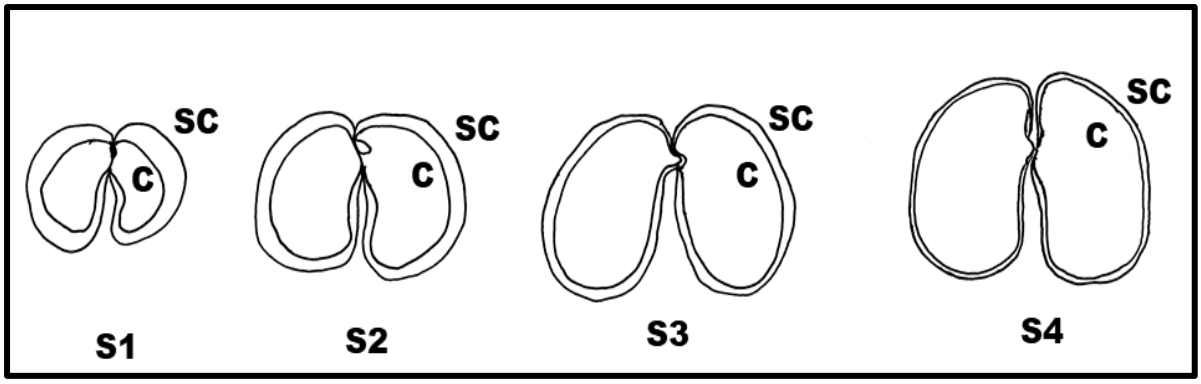


Figure 4.1: *Diagrammatic representation of developing P. pinnata seeds. SC – seed coat, C – cotyledons.*

In general, chlorophyll content and the ratio of chlorophyll a/b ratio play a significant role in developing green seeds and are directly related to the photosynthetic efficiency of the green seeds (Eastmond et al., 1996). In the current study, initial stages of *P. pinnata* seed showed higher chlorophyll a/b ratio (>1), which declined gradually with maturation. This clearly indicated the active photosynthesis in *P. pinnata* developing seeds during S1 and S2. To support our results, we have also analyzed the gene expression of some of the key photosynthetic electron transport genes. PS I and PS II related gene expression was reported in many green and oil yielding seeds and it is also considered to be an important factor for embryo development and fatty acid biosynthesis (Allen et al., 2009; Niu et al., 2009, Allorent et al., 2015). The expression of photosynthetic genes in *P. pinnata* was higher till S2 after which there was a steady decline in expression levels indicating very active photosynthetic process during the initial stages of seed development. The temporal induction of ATP synthase subunit gene (*atp1A*) could have also contributed for substantial production of ATP which is needed for the energy productions and fatty acid biosynthesis.

Criteria	Current results		Previous findings with references		
Duration of seed development	12 months		11–12 months		Scott et al., 2008; Sangwan et al., 2010.
Seed storage compounds	Lipid, protein, starch, other carbohydrates		Lipid, protein, starch, total carbohydrates		Pavithra et al., 2013; 2014
Seed oil yield	37 %		30–40 %		Kumar and Sharma 2011.
Fatty acid composition in matured seeds	Palmitic acid	9.9 % (slight decrease with maturation)	Palmitic acid	9–11 % (decreases with maturation)	Bala et al., 2011; Pavithra et al., 2012; Sharma et al., 2016.
	Stearic acid	5.1 % (no change)	Stearic acid	4–5 % (no change)	
	Oleic acid	61.7 % (increases with maturation)	Oleic acid	60–70 % (increases with maturation)	
	Linoleic acid	17.8 % (decreases with maturation)	Linoleic acid	13–17 % (slight decrease with maturation)	
	Others	5.5 % (no change)	Linolenic acid	2–5 % ((decreases with maturation)	
Transcripts/Genes of seed	Transcript analysis of seeds.		Seed transcripts, SSR markers		Huang et al., 2016, 2018.
Sugar profile in the developing seed	Sucrose and Fructose as the major sugars		No data		—
Metabolomics of seed	Crucial metabolites of major pathways		No data		—
Proteomics of seed	Key proteins for seed development and fatty acid metabolism		No data		—

Table 4.1: *Current information available on the seed development in P. pinnata*

The seed photosynthesis may provide energy for the fatty acid biosynthesis during subsequent stages and also helps in re-fixing the respiratory CO_2 release. The release of O_2 through the photosynthesis can aid in reducing $NO:O_2$ ratio, thereby preventing anoxia in seeds during development (Borisjuk and Rolletschek 2009). Active photosynthesis might also contribute for significant accumulation of storage products including protein, starch and other carbohydrates which was evident in this study with *P. pinnata* seeds.

The storage chemistry of *P. pinnata* seeds have been reportedly dominant with the high lipid and protein content followed by the least amount of starch content in a fully developed and matured seeds (Pavithra et al., 2014; Figure 3.7). To figure out their subcellular stay and basic characteristics, anatomical studies were carried out focusing mainly on the oil bodies (Figure 3.8). The storage products present inside legume seeds are distributed at confined locations inside the cotyledons. A typical legume seed stores protein in the form of protein bodies/protein storage vacuoles and lipids in the form of oil bodies (Shang et al., 2020). Protein bodies in legumes are mixture of various protein molecules which arrange themselves at different sub–cellular regions. The formation of protein storage vacuoles and protein bodies are regulated by phosphate rich phytates and globoids which give them the spherical globular structures (Madsen and Brinch–Pedersen 2020). The cotyledons of *P. pinnata* showed large globular protein storage vacuoles which were mainly distributed around the cellular space (Figure 3.9, 3.10). On the other hand, the seed oil storage involves fatty acid biosynthesis and assembly of triacylglycerides (TAGs), which is followed by packaging into oil bodies with the help of oleosin protein (Song et al., 2017). The number of oil bodies observed in a matured *P. pinnata* seed (S4) was much higher but smaller in sizes when compared to the protein bodies or protein storage vacuoles (Figure 3.10).

The count and size of oil bodies of a mature *P. pinnata* seeds ranged from 0.5 to 1.5 µ and 120 to 180 count per cell respectively (Table 3.5). Oil body formation and lipid accumulation in the matured seeds have to be clearly known by the cultivars in order to get the full benefit of the harvest (Pavithra et al., 2013). Additionally, knowledge on the size of oil body and pattern of its deposition during maturation is also important to unfold the mechanisms regulating its molecular processes.

4.3 Fate of sucrose and other storage compounds while picking the favorite hexoses to lipids

The sequential process of development and differentiation in seeds involves sucrose uptake, cellular division and differentiation followed by the end product biosynthesis (Baud et al., 2002). Interestingly for oil seeds, the differentiation phase is of special concern, because during this stage of development, as different metabolic events initiate the accumulation of storage products (Weber et al., 2005). The metabolism of storage accumulation starts with the entry of nutrients, majorly sucrose, peptides, hormones etc. through the funicular attachment of developing fruit and the mother plant (Borisjuk et al., 2003). The maternal tissues and filial parts are rich in sucrose content during the young active growth phase of seed development. Developing seeds completely depend on the maternal tissues for protective, physical and nutritional support (Savadi 2018). Sucrose content is rich in the seed coats of developing legumes, where the enclosed cotyledons further metabolize it to fructose and glucose throughout the different stages of seed development (Figure 4.2). In most of the developing seeds, the production of glucose and fructose are the dependent mechanisms of two enzymes, invertases and sucrose synthase (Weber et al., 2005).

Invertases were again classified spatially into three different types, the cell wall invertases (cwIn), the cytosolic invertases (cytIn) and the vacuolar invertases (vnIn). The cwIn and vnIn are acidic invertases whereas the cytIns are neutral/alkaline invertases. The specific role of these invertases is similar in all the species related to different plant families and mostly in legumes. The cytIns which are mostly neutral/alkaline invertases (Welham et al., 2009) and certain types of the acidic invertases were mostly cell wall bound invertases (Wei et al., 2020). The plant cell wall bound invertases have multiple roles including a major role in seed development (Bergareche et al., 2018; Liao et al., 2020), while neutral invertases have osmotic and abiotic stress regulation (Dahro et al., 2016). As we can see in *P. pinnata* seeds, the acidic invertases play a major role in metabolizing sucrose and the activity was very high in the younger seeds (S1 and S2; Figure 3.14). The activities of neutral/alkaline invertases altogether were also higher in the younger developing stages indicating an all–around high performance of invertases in younger stages when compared to the more matured stage (Figure 3.14).

Along with the invertases, higher activities of sucrose synthase (SS) and sucrose phosphate synthase (SPS) were also observed (Figure 3.15). The reversible breakdown of sucrose has been controlled exclusively by both SS and SPS. Sucrose, as a signaling molecule in legumes, controls most of the metabolism relating to overall seed development and carbon partitioning (Weber et al., 2005). The activity of both SS and SPS in *P. pinnata* seeds are coordinated from S1 to S4, where the steady state concentrations of sucrose and hexoses are also corelated (Figure 4.3). However, the breakdown of sucrose through invertases overshades that of SS in the developing seeds (Figure 4.3).

The preference of hexoses depends on the activity of the hexokinases to metabolize the hexose sugars specifically (Pego and Smeekens 2000). Hexokinases also plays an imminent role in storage biosynthesis and oil accumulation in legumes and other oil yielding species (Stein et al., 2017). However, the preference of fructose over glucose has been observed in multiple oil yielding species such as *Brassica* and *Arabidopsis* (Stein et al., 2017). The higher activities of fructokinase when compared to the hexokinases in *P. pinnata* seeds also confer the active role of these kinases in controlling the further metabolic events which will be channelized towards other developmental processes. However, both glucokinase and the fructokinase regulate the breakdown of sucrose as well as multiple other cellular responses related to growth and development in many plant species (Stein et al., 2017).

In this study, the stage wise activities of invertases, SS and hexokinases were shown and the sucrose supplied from mother plant is actively utilized by the developing cotyledons in phases in the younger stages (S1 – S2) and further utilization of hexoses was in the mid developing stages (S2 – S3) (Figure 3.16). Moreover, an increase in levels of sucrose in matured phase also raises the importance in maintaining its steady state level while controlling other developing machinery inside the seeds. Sugar analysis along with enzymes reveal the fate of sucrose and its further mobilization which ultimately leads to precursors of major metabolic pathways leading to storage production (Figure 4.4). Sucrose is channelized mainly into hexoses (glucose and fructose) which are further utilized during the accumulation of major storage products in the form of lipids and proteins through the respective pathways of fatty acids and amino acids (Figure 4.4).

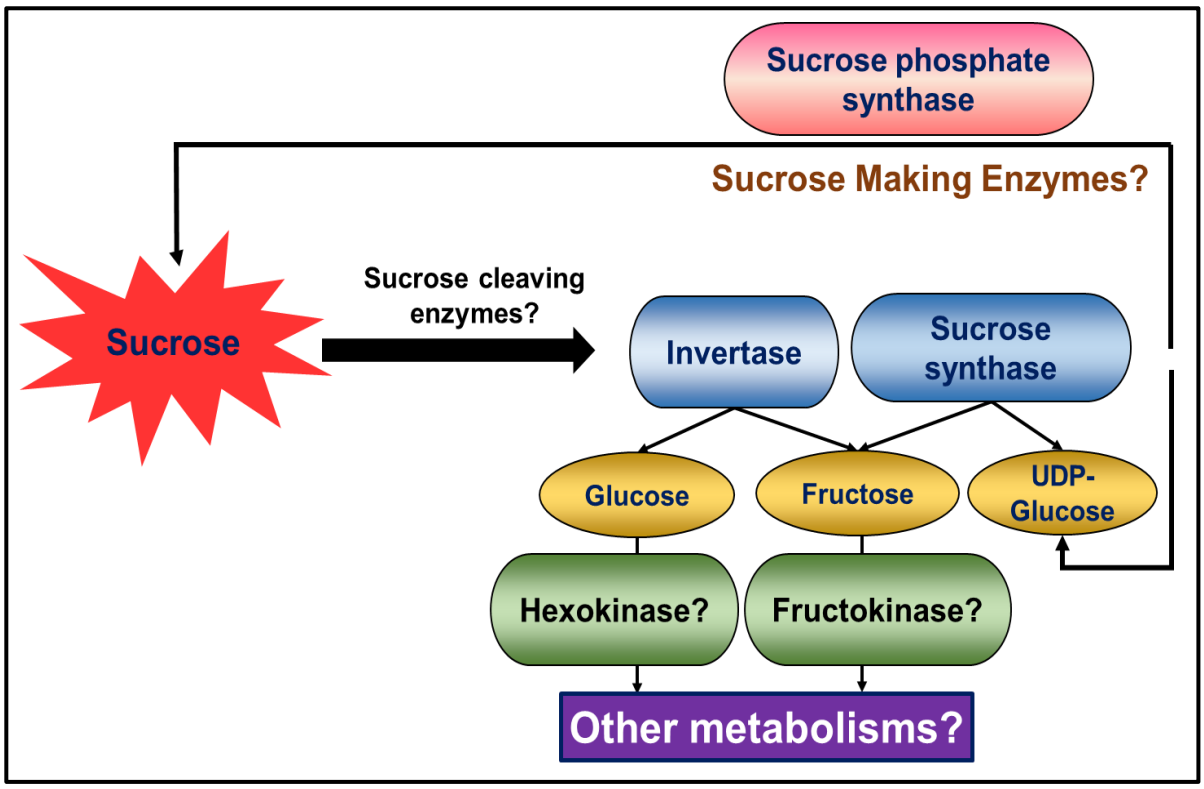


Figure 4.2: *Schematic representation of metabolism of sucrose into hexoses (glucose and fructose) and further into different metabolic events.*

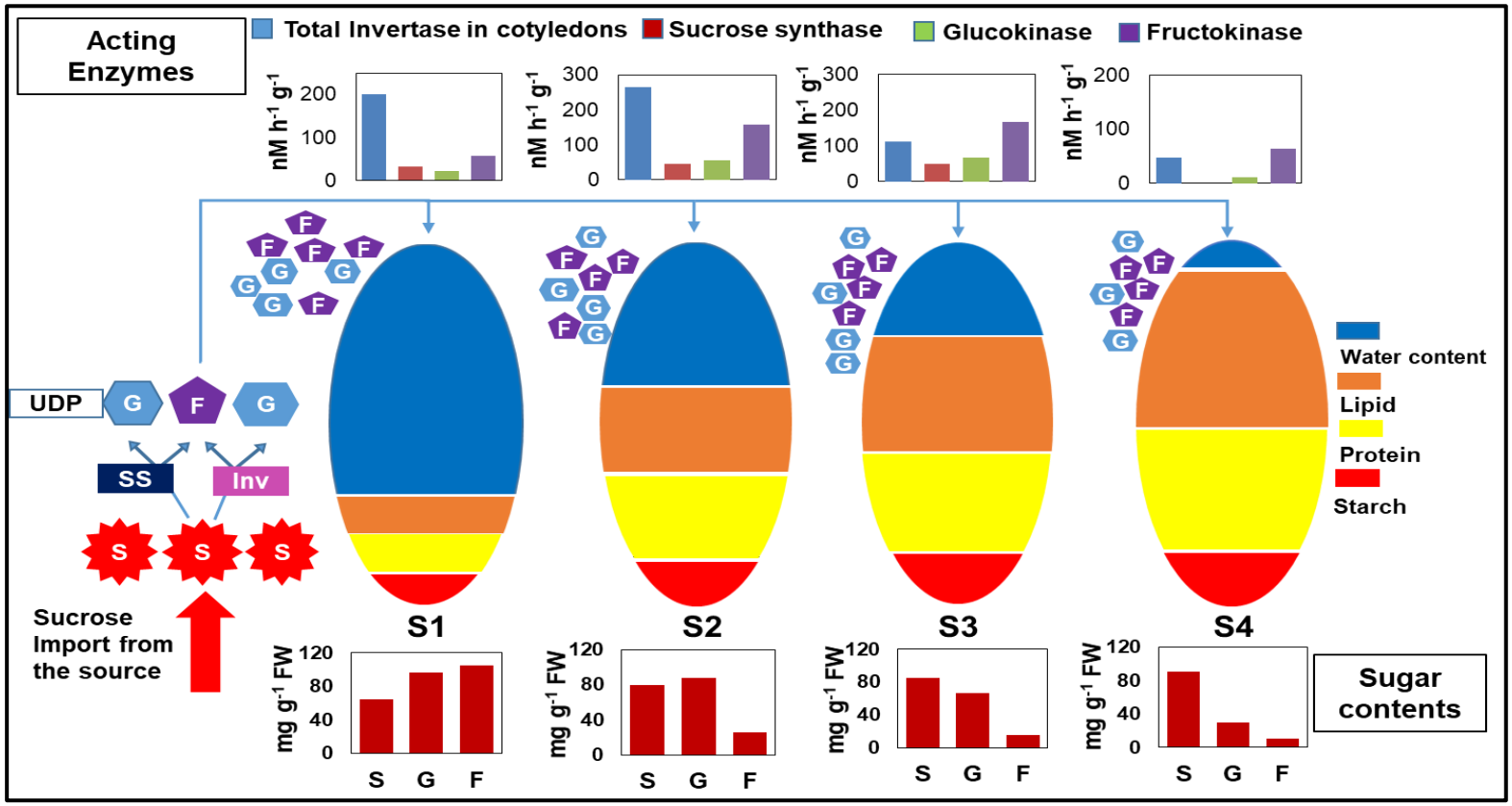


Figure 4.3: *Our proposed model for sugar–protein–lipid changes during seed development in P. pinnata.*

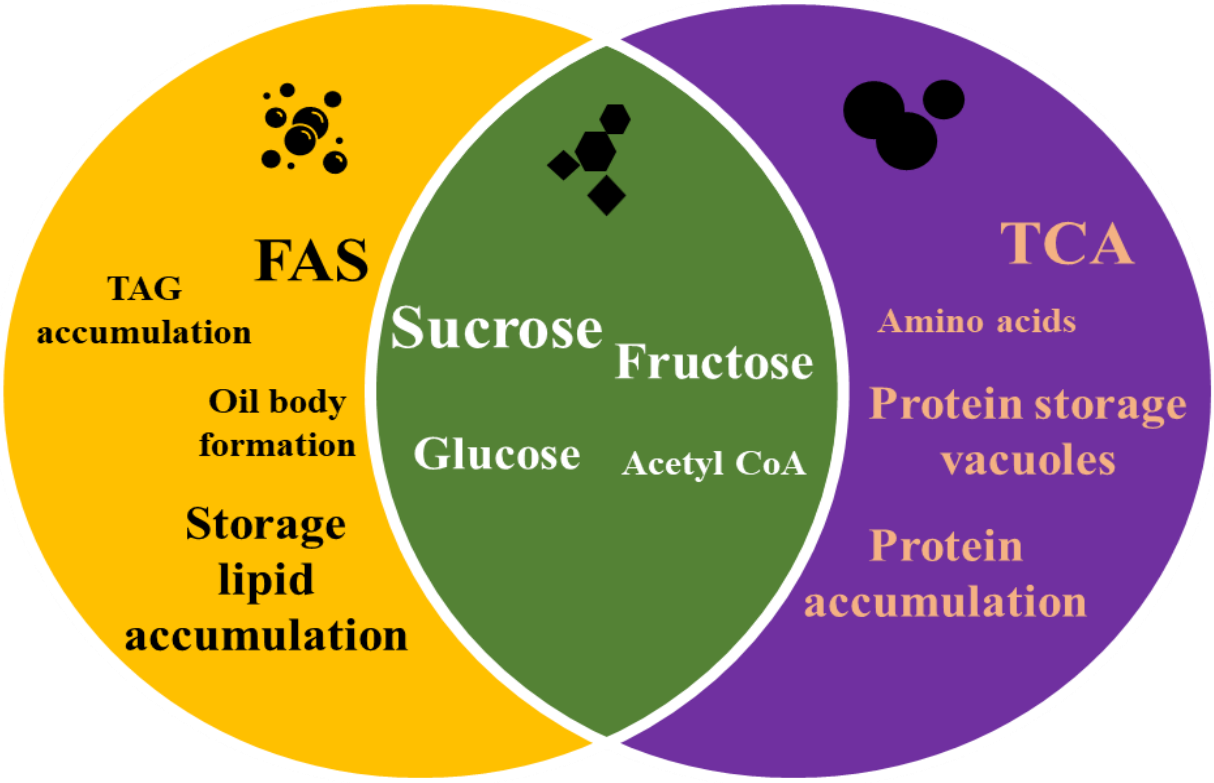


Figure 4.4: *Diagrammatic representation of metabolism of sucrose into hexoses (glucose and fructose) and further accumulation of major storage products in the form of lipids and proteins.*

4.4 Dynamic changes in primary metabolites with respect to different pathways in the developing seeds.

Metabolites are considered as very robust and sensitive as they maintain the seed physiology and storage biochemistry (Weselake et al., 2009). Understanding the dynamics of metabolic profile among various stages of development might give new insights into the mechanism of seed oil biosynthesis with respect to quantity and quality of lipids in this potential biofuel crop. The stage dependent metabolite changes of various crops including *Oryza sativa, Glycine max* and *Brassica napus* were recently reported which had enlightened the role of metabolites in regulating oil biosynthesis, cellular morphology, seed filling and maturity (Tan et al., 2015; Hu et al., 2016; Gupta et al., 2017). We have detected sugars, organic acids, free fatty acids and amino acids through our metabolite analysis in developing *P. pinnata* seed. Sugars and organic acids take part in energy metabolism thus providing reducing compounds as well as precursors for fatty acid biosynthesis, cell division and differentiation. Among the detected sugars, glucose is utilized as a signal for cellular multiplication and differentiation while other sugars like ribose and erythrose take part in nucleotide formation and energy metabolism such as Calvin cycle. Interestingly, sugars related to cell wall formation and cell division were higher in S2-S3, for example NADGA, cellobiose, pyranose and arabinose indicating that the active stages of cellular division and differentiation are from S1 to S3. Sugars such as mannose were most abundant in the mature stage S4 and it may contribute to the production of bioethanol from remaining seed residue after the extraction of oil.

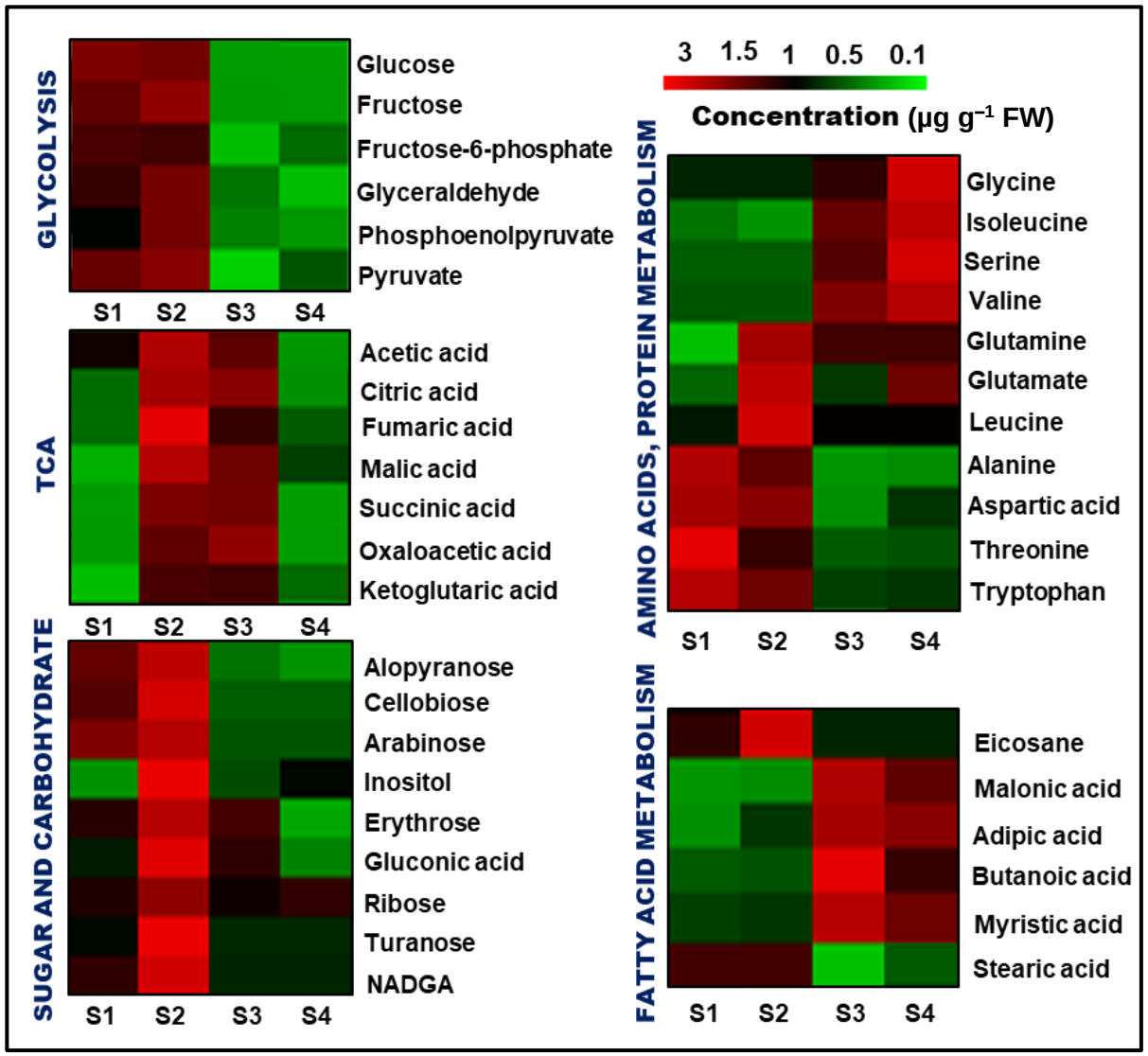


Figure 4.5: *Classification of metabolites based with their respective pathways and metabolism. All the detected and identified metabolites are clubbed into their respective known pathways and other related cellular metabolism.*

The major metabolites of glycolysis and organic acid components of TCA cycle were also detected in the analysis where, the stage wise quantitative changes were observed in *P. pinnata* seeds. The presence of a flux mode for respiratory energy synthesis through TCA may be the reason for the accumulation of these metabolites in higher concentration during initial stages (S1 – S3), which in turn can aid the fatty acid biosynthesis. Amino acids in the developing seeds contribute in various metabolic activities including storage protein accumulation, germination, stress responses and in respiratory catabolic processes to maintain optimum energy status in the developing seeds (Miranda et al., 2001; Galili et al., 2014). Most of the amino acids were abundant in S2 – S3, while glycine and proline which have stress responsive characteristics were abundant during S4, providing clues for stress tolerance in *P. pinnata* seed as the development progressed (Hayat et al., 2012; Czolpinska and Rurek 2018). The other important amino acids such as glutamate, aspartate, glutamine were also detected which play a central role in the synthesis of proteins during seed development (Jander and Joshi. 2009). Pathway based classification of the metabolites and their changes in concentrations shows a clear distinction among the stages (S1 – S4) (Figure 4.5). Metabolite content of glycolysis was higher in S1 and S2, while the respective metabolites of TCA cycle were high in S2 and S3, showing a stage dependent expression of these pathways ultimately controlling the production of precursors for the storage biosynthesis (Figure 4.5).

4.5 Quality of *P. pinnata* seed oil at maturation and the end of the long spanned developmental phase.

We have analyzed the control of fatty acids, TAG production and oil body accumulation after channelizing the metabolites towards the fatty acid synthesis pathway. The accelerating accumulation of oil bodies in developing seeds could help in determining the best harvest time of *P. pinnata* seeds. However, the size and pattern of oil body accumulation during seed development is important for understanding the dynamics of molecular regulation behind this process. It was evident from our data that TAG levels were increased gradually as the seed development progressed and the fate of most of the TAGs accumulated is to ultimately form oil bodies (Hills et al., 2004). The increasing number of oil bodies with maturation suggest that majority of storage product observed were lipids in *P. pinnata* seed. The FAMEs composition of biofuel and its blend can determine its efficiency for fuel capacity (Sbihi et al., 2018; Xiong et al., 2018). In *P. pinnata*, the fatty acid composition of accumulated TAGs at each stage had shown that oleic acid was predominantly incorporated only after S3 making the seed oil of *P. pinnata* rich with unsaturated fatty acids and making it a valuable source of biodiesel which meets all the standard criteria for biodiesel production (Karmee and Chadha 2005). This result also suggests the tropical legume *P. pinnata* as a valuable and relevant source of biodiesel meeting all the standard criteria for better productivity. During the course of seed development in *P. pinnata*, decreased accumulation of fatty acids such as palmitic acid and linoleic acid has been predominantly observed (Figure 3.21). A schematic representation of the events interlinking the precursors for oil body formation has been presented where acetic acid enters the malonyl CoA pathway. Later, following the FAS cycle in plastids, the fatty acids are exported out for the oil body accumulation (Figure 4.6).

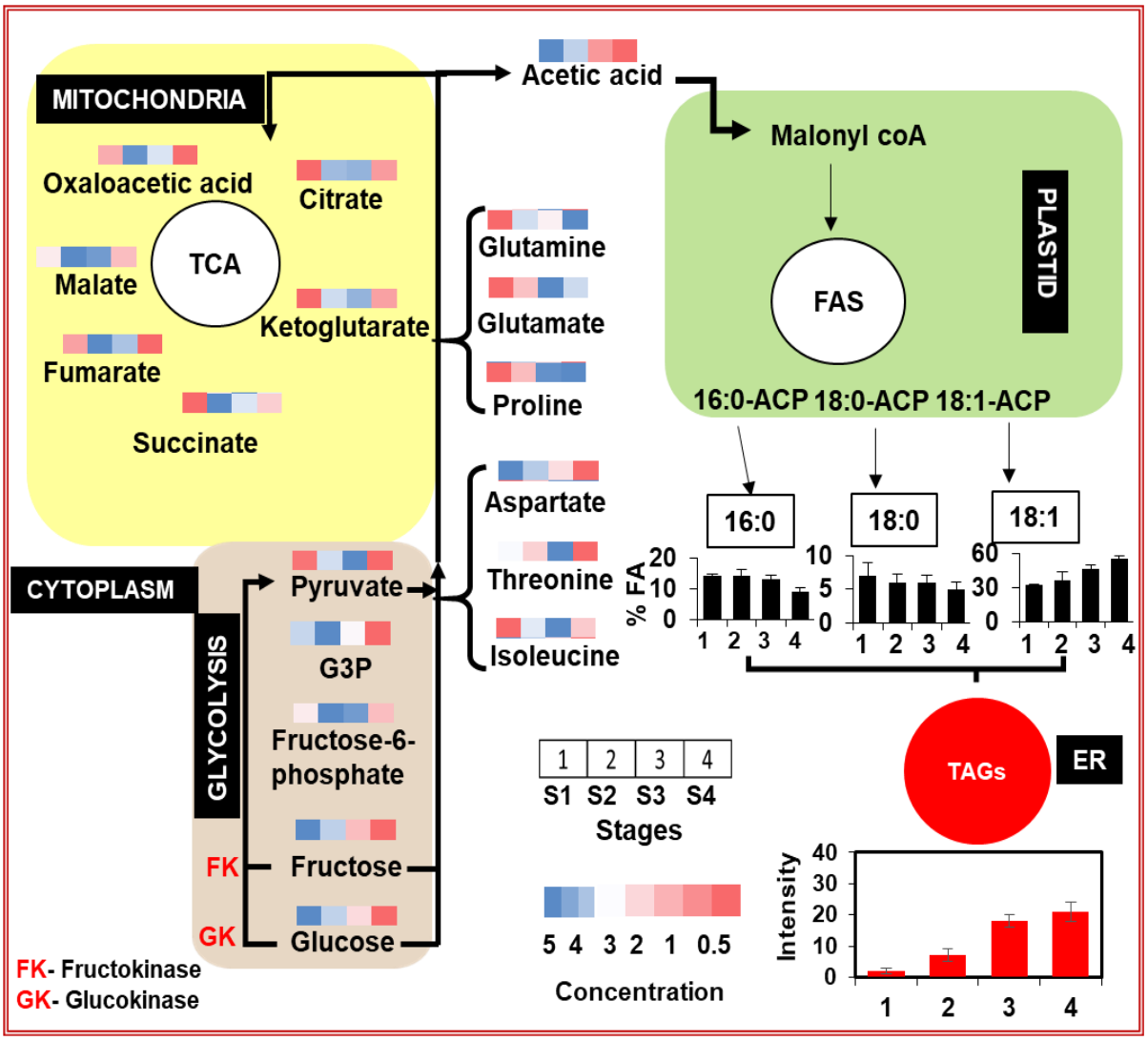


Figure 4.6: *Schematic presentation on the regulatory events during oil biosynthesis in P. pinnata seeds (S1 to S4). Stage wise changes in the metabolite concentrations belonging to various pathways in their subcellular locations, which later channelized towards the fatty acid biosynthesis and ultimately the TAG accumulations.*

4.7 Few proteins are identified which help in undermining the pathways related to storage production in *P. pinnata*.

Total seed protein profile analyzed through 2D gel electrophoresis provide the overall picture of stage wise variation in protein expression and regulation, where some key regulatory proteins have been successfully identified. The proteins, related specifically to seed development where Midasin and ap2 complex subunit alpha which take part in the development of female gametophyte as well as the cotyledonary development (Chantha et al., 2010). On the other hand, glyceraldehyde–3–phosphate dehydrogenase, a glycolytic enzyme showed higher expression during younger stages which decreased significantly at S4 indicating an active cellular mechanism such as glycolysis at younger stages of S1 – S2 which gradually decreased with maturation. ATPase subunit beta of photosynthetic electron transport and PEP carboxykinase of carbon metabolism were upregulated at S2 – S3. Our data infer that photosynthesis in *P. pinnata* seed plays a critical role in carbon recycling and energy metabolism which is crucial for fatty acid biosynthesis. Significantly upregulated spots from S1 to S4 included acetyl CoA carboxylase (*accase*), the first catalyzing enzyme of fatty acid biosynthesis. In addition, Steroyl–ACP–desaturase (SAD) the enzyme for desaturation of stearic acid (C18:0) to oleic acid (C18:1) and Lipid transfer like protein (LTPs). SAD is the crucial enzyme in *P. pinnata* for production of first unsaturated fatty acid and for maintaining its quality as a biodiesel source (Bates et al., 2013). The LTPs are low molecular weight (9-10 kD) proteins and have a significant role in plant development such as transporting lipids for membrane biosynthesis, defense responses and vesicular transports (Liu et al., 2015).

A subcellular localization study in seeds also showed the involvement of LTPs with protein as well as lipid bodies (De O Carvalho et al., 2004). In addition, GTP binding protein SAR1b, which is well known for taking part in vesicular protein traffic from Endoplasmic reticulum to Golgi leading to storage body formation was detected in *P. pinnata* seeds (Memon 2004). Moreover, recently in liver cells, phosphorylation of SAR1b had been proven to be an important step for release of lipid chylomicrons (Siddiqi and Mansbach. 2012). However, in plants particularly the high oil yielding species like *P. pinnata*, it still remains a question to whether this protein carries a role in the synthesis of lipid bodies. Our data in the present demonstrated that growth and development are mostly active in younger stages followed by the energy metabolism at mid developing stage where, on the other hand, the lipid metabolism is increased from mid developing phase till the matured phase (Figure 4.7). Young phase (S1), mid–developmental phases (S2–S3) and the maturation phases (S4) are categorized based on the functions of identified proteins in *P. pinnata* seeds (Figure 4.7).

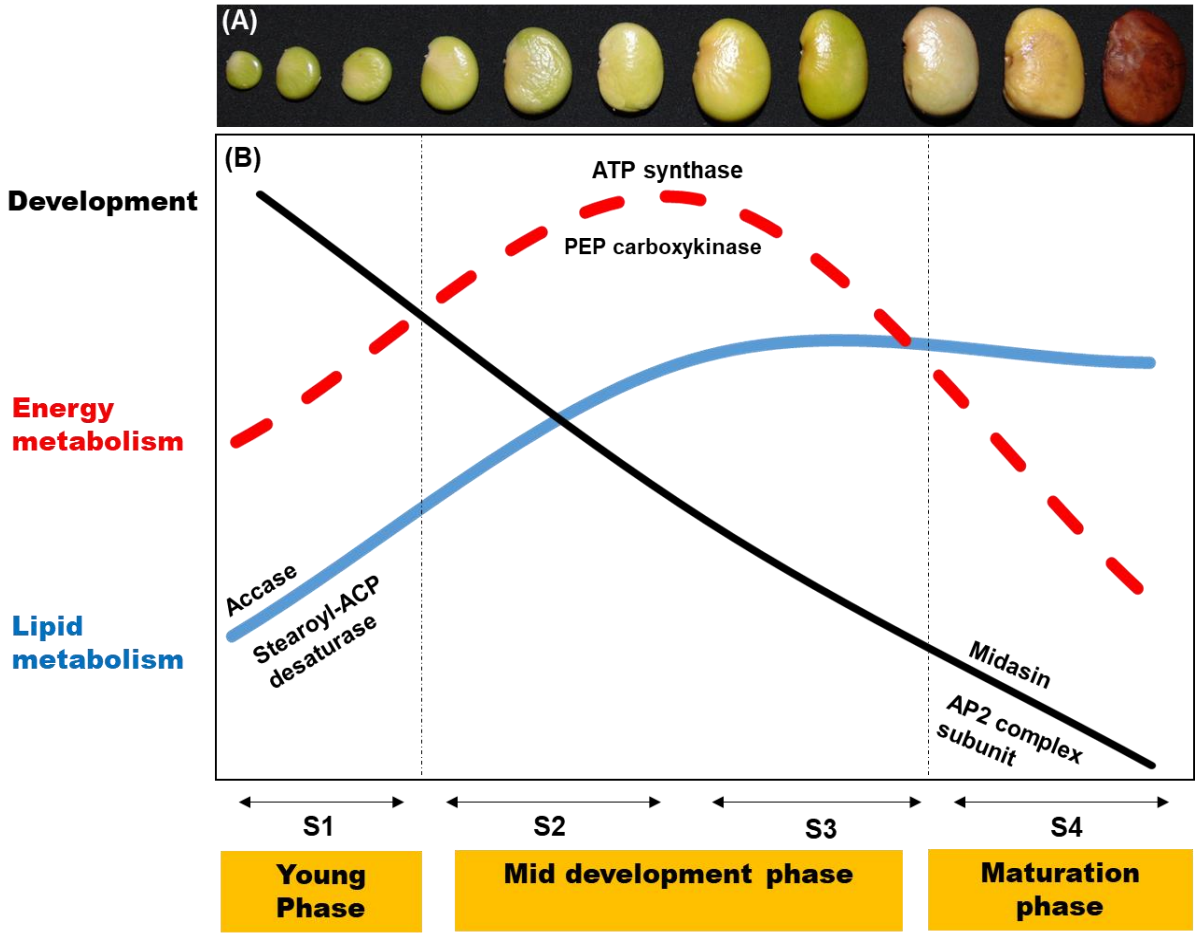


Figure 4.7: *Stage wise expression profile of crucial Proteins controlling key regulations of overall seed development. Developing seeds collected from 120 to 300 DAA (A). S1 to S4 has been categorized into three main phases.*

CHAPTER 5

SUMMARY AND CONCLUSIONS

Our current findings have created a new insightful understanding and information in relation with the regulation of seed development in *P. pinnata*, specifically on seed oil biogenesis. The vegetative growth physiology of *P. pinnata* trees have shown its physiological features along with the photosynthetic properties during the seed-bearing seasons. Unlike the other legume crops this tropical tree has a huge scope of versatility and adaptability due to which there is a rising attention on its productivity. Therefore, the present study on this versatile biofuel tree species also includes detail information on fatty acid biosynthesis, oil accumulation in matured seeds and the role of sugars as the initial carbon skeleton coming from the source tissues. In addition, evidence of probable seed photosynthesis in the green young developing seeds of *P. pinnata* has also been documented. The period of developing process in *P. pinnata* seeds is lengthy, which makes it very complex and crucial to unravel the patterns of seed filling related to oil accumulation. Hence, the knowledge on physiological and molecular regulations controlling these processes will eventually helpful and further accelerates the efforts to enhance productivity towards feedstock productions of biofuels. The control of sugars and other different metabolites plays a key role as a signaling element in the cotyledonary tissue formation and the accumulation of the storage products. Present study minutely examined the developmental processes of *P. pinnata* seeds in order to understand the physiological, biochemical and molecular dynamics through the analysis of sugars, crucial enzymes with metabolomics and proteomic analysis. The data also emphasizes on few but thoughtful insights towards understanding the nature of its regulation in seeds which include the photosynthetic intermediates as the additional sources of precursors of lipid metabolism using seed photosynthesis and energy regulations.

Further, the data in this thesis provide a thorough understanding towards complex regulations of sugars, metabolites and key proteins in maintaining the regulatory developmental processes as well as the storage biosynthesis with special attention to the oil body formation.

- ***Morphology and anatomy resemble a typical legume but very rich in lipid content***

Seed development in *P. pinnata* initiates with the formation of the embryo and the cotyledonary tissues. Due to seed desiccation, the protective seed coat around cotyledons, which is very thick in S1 become negligible in biomass when it reaches maturation. Inside the cotyledonary tissues, empty cells at very young stage becomes the cells filled with various storage granules mainly the oil/lipid bodies at fully developed stage. The greening and photosynthetic activity in seeds are also of greater interest in *P. pinnata* due to high lipid content. Generally, legume seeds are very rich in the starch and protein. However, the high oil containing legume seeds are also found to be regulated by many other metabolic events including seed photosynthesis and heteromorphic regulation of storage production. Now, the crucial issue has been about the source of carbon for lipid production in developing seeds and how the carbon conversion was maintained and controlled inside the cotyledonary tissues which is discussed in detail below.

- ***Sucrose is the source provider and preferred hexose sugar is fructose***

The sucrose supplied by the mother plant to the seeds is cleaved by the sucrolytic enzymes such as the invertases and the sucrose–synthase. While sucrose synthase has a reversible process of action, the invertases act in non-reversible way. There are two different types of invertases present inside the developing seeds depending on the cellular pH, which are classified as below: –

- Neutral/Alkaline Invertase– cytosolic invertase (pH 6.5–7.5).
- Acidic invertase– present mainly in the cell wall and the vacuoles (pH < 5).

High activities of invertases at the younger stages (S1 and S2) lead to the rapid breakdown of sucrose into glucose and fructose continuously. During these stages in *P. pinnata* seeds, sucrose has been readily entered and continuously been metabolized further. Most of the invertases preferred an acidic environment than neutral/alkaline conditions. Both glucose and fructose have to be broken down by specific hexokinases in order to move towards further metabolic processes. While both of these sugars undergo mainly towards the catabolic pathways, the developing seeds of *P. pinnata* prefer fructokinases over hexokinases. Where it is converted to form fructose–6–phosphate to go to the glycolysis and other following pathways. Moreover, sucrose synthase and sucrose phosphate synthase also may play a critical role in the developing machinery of the *P. pinnata* seeds. The reversible conversion of sucrose to hexoses is also controlled by the level of sucrose in many legumes including the developing seeds of *P. pinnata*. Additionally, it has also been observed that expression of genes encoding invertases is jointly controlled by these two enzymes (SS and SPS) maintaining the equilibrium of sucrose in the seed. Maturation and various other developmental regulations inside the seeds are also controlled by the levels of sucrose. Additionally, it has also been observed that expression of genes encoding invertases and fructokinases are specific and few important transcripts have been studied here using the real time gene expression analysis.

Accordingly, a coordinated regulation of stage wise utilization of sucrose and then the hexoses (mainly fructose) has been observed in developing seeds of *P. pinnata* and these findings give a new insight towards the source–sink relationship and metabolism of carbon skeleton distribution in eminent biofuel legume tree, *P. pinnata*.

- ***Metabolic dynamics and key regulatory proteins in the process of storage lipid biosynthesis***

Once the hexoses are hydrolyzed into respective byproducts, they can move through glycolysis which will be initiated by either hexokinase/fructokinase to glucose–6–phosphate or fructose–6–phosphate. By both ways, the end product of the glycolysis is pyruvate which will enter into TCA or fatty acyl synthase pathway in order to provide the substrates for amino acids and fatty acid biosynthetic processes respectively. The metabolic regulation in the developing seeds will be through the same course where amino acids and fatty acids are the building blocks of the protein and lipids in the developing *P. pinnata* seeds. Desaturation of fatty acids has also led to formation of carbon–carbon double bond which ultimately produces the unsaturated fatty acids followed by glycerol–lipid metabolism related to synthesis of TAGs and oil body formation. Further, this study demonstrated the synthesis of fatty acids till the process of unsaturation and oil body formation which is a coordinated network starting from breakdown of the sucrose. Stage wise metabolic profile of pathway specific intermediates and fatty acids of the TAGs accumulated has linked the precursor–product relationship which makes a great deal towards understanding the production of oleic acid and other unsaturated fatty acids which are key regulatory factors for biodiesel production. Moreover, the present study shows that cellular development and seed photosynthesis also may control lipid metabolism which is stage wise regulated.

Proteomics of *P. pinnata* seeds have clearly indicated the phases (young, mid-development and mature) and related metabolisms (development, energy and lipid metabolism) of seed development. After the active growth phase at S1, photosynthesis in the green seeds increased which might give the energy for fatty acid biosynthesis and lipid accumulation during the maturation phase. The metabolites together with regulatory proteins dynamics present in each developing stage clearly demonstrate the overall controlling elements of the storage metabolism which are useful for pathway mapping and metabolite targeting for improvement in the seed oil biosynthesis.

CHAPTER 6
LITERATURE CITED

Aadi Moolam R, Singh A, Shelke RG, Scott PT, Gresshoff PM and Rangan L. (2016). Identification of two genes encoding microsomal oleate desaturases (FAD2) from the biodiesel plant *Pongamia pinnata* L. *Trees*, 30(4): **1351–1360.**

Abugre S and Sam SJQ. (2010). Evaluating the allelopathic effect of *Jatropha curcas* aqueous extract on germination, radicle and plumule length of crops. *International Journal of Agriculture and Biology*, 12: **769–772.**

Aguilera–Alvarado GP, Guevara–García ÁA, Estrada–Antolín SA and Sánchez–Nieto S. (2019). Biochemical properties and subcellular localization of six members of the HXK family in maize and its metabolic contribution to embryo germination. *BMC plant biology*, 19(1): **1–18.**

Aguirre M, Kiegle E, Leo G and Ezquer I. (2018). Carbohydrate reserves and seed development: An overview. *Plant Reproduction***,** 31(3): **263–290.**

Akhtar S, Khalid N, Ahmed I, Shahzad A, Suleria HA. (2014). Physicochemical characteristics, functional properties, and nutritional benefits of peanut oil: a review. *Critical Reviews in Food Science and Nutrition***,** 54(12):**1562–1575.**

Ali F, Qanmber G, Li F and Wang Z. (2022). Updated role of ABA in seed maturation, dormancy, and germination. *Journal of Advanced Research*, 35: **199–214.**

Allen DK, Ohlrogge JB and Shachar-Hill Y. (2009). The role of light in soybean seed filling metabolism. *The Plant Journal*, 58(2): **220–234**.

Allorent G, Osorio S, Ly Vu J, Falconet D, Jouhet J, Kuntz M, Fernie AR, Lerbs-Mache S, Macherel D, Courtois F and Finazzi G. (2015). Adjustments of embryonic photosynthetic activity modulate seed fitness in *Arabidopsis thaliana*. *New Phytologist*, 205: **707–719.**

Angelovici R, Galili G, Fernie AR and Fait A. (2010). Seed desiccation: a bridge between maturation and germination. *Trends in Plant Science*, 15: **211–218.**

Arnon D. (1949). Copper enzymes isolated chloroplasts, polyphenoloxidase in *Beta vulgaris*. *Plant Physiology*, 24: **1–15.**

Arpiwi NL, Yan G, Barbour EL and Plummer JA. (2013). Genetic diversity, seed traits and salinity tolerance of *Millettia pinnata* (L.) Panigrahi, a biodiesel tree. *Genetic Resources and Crop Evolution*, 60 (2): **677–692.**

Atabani AE, Silitonga AS, Ong HC, Mahlia TMI, Masjuki HH, Badruddin IA and Fayaz, H., (2013). Non–edible vegetable oils: A critical evaluation of oil extraction, fatty acid compositions, biodiesel production, characteristics, engine performance and emissions. *Renewable and Sustainable Energy Reviews*, 18: **211–245.**

Ayalew Y, Retta N, Desse G, Mohammed A and Mellesse A. (2017). Amino acid profile and protein quality in tuber and leaf of *Coccnia abyssinica* (Lam.) (Cogn.) accessions of Ethiopia. *Food science & nutrition*, 5(3): **722–729.**

Bala M, Nag TN, Kumar S, Vyas M, Kumar A and Bhogal NS. (2011). Proximate Composition and Fatty Acid Profile of *Pongamia pinnata*, a Potential Biodiesel Crop. *Journal of the American Oil Chemists' Society*, 88: **559–562.**

Bates PD, Stymne S and Ohlrogge J. (2013). Biochemical pathways in seed oil synthesis. *Current Opinion in Plant Biology*, 16: **358–364.**

Baud S, Boutin J, Miquel M, Lepiniec L and Rochat C. (2002). An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiology and Biochemistry*. 40(2): **151–160.**

Baud S, Dubreucq B, Miquel M, Rochat C and Lepiniec L. (2008). Storage reserve accumulation in *Arabidopsis*: metabolic and developmental control of seed filling**.** *The Arabidopsis Book*. 6: **e0113.**

Baud S. (2018). Seeds as oil factories. *Plant Reproduction*, 31: **213–235.**

Bergareche D, Royo J, Muñiz LM, Hueros G. (2018). Cell wall invertase activity regulates the expression of the transfer cell-specific transcription factor ZmMRP-1. *Planta***,** 247: **429–442.**

Bieleski RL and Turner NA. (1966). Separation and estimation of amino acids in crude plant extracts by thin-layer electrophoresis and chromatography. *Analytical biochemistry*, 17(2): **278–293**.

Biswas B, Kazakoff SH, Jiang Q, Samuel S, Gresshoff PM and Scott PT. (2013). Genetic and genomic analysis of the tree legume Pongamia pinnata as a feedstock for biofuels. *The Plant Genome*, 6(3): **2013–05.**

Bligh EG and Dyer WJ. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology***.** 37: **911–917.**

Boelt B, Julier B, Karagić Đ and Hampton J. (2015). Legume seed production meeting market requirements and economic impacts. *Critical Reviews in Plant Sciences*, 34(1–3): **412–427.**

Borisjuk L and Rolletschek H. (2009). The oxygen status of the developing seed. *New Phytologist*, 182: **17–30.**

Borisjuk L, Nguyen TH, Neuberger T, Rutten T, Tschiersch H, Claus B, Feussner I, Webb AG, Jakob P, Weber H and Wobus U. (2005). Gradients of lipid storage, photosynthesis and plastid differentiation in developing soybean seeds. *New Phytologist*, 167(3): **761–776.**

Borisjuk L, Rolletschek H, Wobus U and Weber H. (2003). Differentiation of legume cotyledons as related to metabolic gradients and assimilate transport into seeds. *Journal of Experimental Botany*, 54(382): **503–512.**

Bradford MM. (1976). A Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Analytical Biochemistry*, 72: **248–254.**

Cao Y, Liu W, Xu X, Zhang H, Wang J and Xian M. (2014). Production of free monounsaturated fatty acids by metabolically engineered *Escherichia coli*. *Biotechnology for Biofuels and Bioproducts*, 7: **59.**

Catusse J, Job C and Job D. (2012). Proteomics reveals a potential role of the perisperm in starch remobilization during sugar beet seed germination. In Seed Development. *OMICS Technologies toward Improvement of Seed Quality and Crop Yield*: **27–41.**

Chaitanya BSK, Kumar S, Anjaneyulu E, Prasad RBN, Sastry PS and Reddy AR. (2015). Pivotal role of sugar fluxes between the inner integument and endosperm in lipid synthesis during seed ontogeny in *Jatropha curcas* L. *Industrial Crops and Products*, 76: **1106–1113.**

Chantha SC, Gray–Mitsumune M, Houde J and Matton DP. (2010). The MIDASIN and NOTCHLESS genes are essential for female gametophyte development in *Arabidopsis thaliana*. *Physiology and Molecular Biology of Plants*, 16(1): **3–18.**

Chauhan JS, Tomar YK, Badoni A, Singh NI, Ali S, Rawat AS and Nautiyal VP. (2009). Morphology and influence of various plant growth substances on germination and early seedling growth in *Macrotyloma uniflorum* (Lam.). *Journal of American Science*, 5(6): **43–50.**

Chen L, Zheng Y, Dong Z, Meng F, Sun X, Fan X, Zhang Y, Wang M and Wang S. (2018). Soybean (*Glycine max*) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. *Molecular Genetics and Genomics*, 293: **401–415.**

Coetzee R, Labuschagne MT and Hugo A. (2008). Fatty acid and oil variation in seed from kenaf (*Hibiscus cannabinus* L.). *Industrial Crops and Products*, 27(1): **104–109.**

Cousin R. (1997). Peas (*Pisum sativum* L.). *Field crops research*. 53(1-3): **111–130**.

Czolpinska M and Rurek M. (2018). Plant Glycine–Rich Proteins in Stress Response: An Emerging, Still Prospective Story. *Frontiers in Plant Science*, 9: **302.**

Dahro B, Wang F, Peng T, Liu JH. (2016). PtrA/NINV, an alkaline/neutral invertase gene of Poncirus trifoliata, confers enhanced tolerance to multiple abiotic stresses by modulating ROS levels and maintaining photosynthetic efficiency. *BMC plant biology*, 16: **1–8.**

De Cillis F, Leoni B, Massaro M, Renna M and Santamaria P. (2019). Yield and Quality of Faba Bean (*Vicia faba* L. var. *major*) Genotypes as a Vegetable for Fresh Consumption: A Comparison between Italian Landraces and Commercial Varieties. *Agriculture*, 9: **253.**

De O Carvalho A, De S Teodoro CE, Da Cunha M, Okorokova-Façanha AL, Okorokov LA, Fernandes KV and Gomes VM. (2004). Intracellular localization of a lipid transfer protein in *Vigna unguiculata* seeds. *Physiologia Plantarum*, 122(3): **328–336.**

de Vries SC, and Weijers D. (2017). Plant embryogenesis. *Current Biology*, 27(17): **R870–R873.**

Degani E, Prasad MVR, Paradkar A, Pena R, Soltangheisi A, Ullah I, Warr B and Tibbett M. 2022. A critical review of *Pongamia pinnata* multiple applications: From land remediation and carbon sequestration to socioeconomic benefits. *Journal of Environmental Management*, 324:**116297.**

Demirbas A. (2008). Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. *Energy conversion and management*, 49(8): **2106–2116.**

Doidy J, Vidal U and Lemoine R. (2019). Sugar transporters in Fabaceae, featuring SUT MST and SWEET families of the model plant *Medicago truncatula* and the agricultural crop *Pisum sativum*. *PLoS One*, 14(9): **e0223173.**

Duranti M. (2006). Grain legume proteins and nutraceutical properties. *Fitoterapia*, 77(2): **67–82.**

Eastmond PJ, Kolacna L. (1996). Photosynthesis by developing embryos of oilseed rape *Brassica napus*. *Journal of Experimental Botany*, 47: **1763–1769.**

Eastmond PJ and Rawsthorne S. (2000). Coordinate changes in carbon partitioning and plastidial metabolism during the development of oilseed rape embryos. *Plant Physiology*, 122(3): **767–774.**

Ekman Å, Hayden DM, Dehesh K, Bülow L and Stymne S. (2008). Carbon partitioning between oil and carbohydrates in developing oat (*Avena sativa* L.) seeds. *Journal of experimental botany*, 59(15): **4247–4257.**

Elevitch CR and Francis JK. (2006). *Gliricidia sepium* (gliricidia). Species Profiles for Pacific Island. *Agroforestry*, 2(1): **1–8**.

Eskandari M, Cober ER and Rajcan I. (2013). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high–oil parents. *Theoretical and applied genetics*, 126(2): **483–495.**

Fernando R. (2012). Variation in seed dormancy and storage behavior of three liana species of *Derris* (Fabaceae, Faboideae) in Sri Lanka and ecological implications. *Research Journal of Seed Science*, 5(1): **1–8**.

Finkelstein RR, Gampala SS and Rock CD. (2002). Abscisic acid signaling in seeds and seedlings. *The Plant cell*, 14 Suppl (Suppl): **S15–S45.**

Fotopoulos VA. (2005). Plant invertases: structure, function and regulation of a diverse enzyme family. *Journal of Biological Research-Thessaloniki*, 4: **127–137**.

Gacek K, Bayer PE, Bartkowiak–Broda I, Szala L, Bocianowski J, Edwards D and Batley J (2017). Genome–Wide Association Study of Genetic Control of Seed Fatty Acid Biosynthesis in *Brassica napus*. *Frontiers in Plant Science*, 7: **2062.**

Gahlaut A, Dahiya M, Gothwal A, Kulharia M, Chhillar AK, Hooda V, Dabur R. Proteomics & metabolomics: Mapping biochemical regulations. *Drug Invention Today*, 5(4): **321–6.**

Galili G, Avin–Wittenberg T, Angelovici R and Fernie AR. (2014). The role of photosynthesis and amino acid metabolism in the energy status during seed development. *Frontiers in plant science*, 5: **447**

Gallant DJ, Bouchet B, Buléon A and Perez S. (1992). Physical characteristics of starch granules and susceptibility to enzymatic degradation. *European journal of clinical nutrition*, 46(2): **S3–S16.**

Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD and Burstin J. (2003). Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiology*, 133(2): **664–682.**

Gallardo K, Thompson R and Burstin J. (2008). Reserve accumulation in legume seeds. *Comptes rendus biologies*, 331(10): **55–762.**

Golombek S, Rolletschek H, Wobus U and Weber H. (2001). Control of storage protein accumulation during legume seed development. *Journal of plant physiology*, 158(4): **457–464.**

Goldberg RB, De Paiva G, Yadegari R. (1994). Plant embryogenesis: zygote to seed. *Science.* 266(5185): **605–14.**

Gomes AMF, Draper D, Nhantumbo N, Massinga R, Ramalho JC, Marques I and Ribeiro–Barros AI. (2021). Diversity of cowpea [*Vigna unguiculata* (L.) walp] landraces in Mozambique: New opportunities for crop improvement and future breeding programs. *Agronomy*, 11(5): **991.**

Granot D, Kelly G, Stein O and David–Schwartz R. (2014). Substantial roles of hexokinase and fructokinase in the effects of sugars on plant physiology and development. *Journal of experimental botany*, 65(3): **809–819.**

Graham PH and Vance CP. (2003). Legumes: importance and constraints to greater use. *Plant physiology*, 131(3): **872–877.**

Greenspan P, Mayer EP and Fowler SD. (1985). Nile red: a selective fluorescent stain for intracellular lipid droplets. *The Journal of cell biology*, 100(3): **965–973.**

Guo N, Zhang S, Gu M and Xu G. (2021). Function, transport, and regulation of amino acids: What is missing in rice? *The Crop Journal*, 9(3): **530–542.**

Gupta M, Bhaskar PB, Sriram S and Wang PH. (2017). Integration of omics approaches to understand oil/protein content during seed development in oilseed crops. *Plant cell reports*, 36(5): **637–652.**

Hayat S, Hayat Q, Alyemeni MN, Wani AS, Pichtel J and Ahmad A. (2012). Role of proline under changing environments: a review. *Plant signaling & behavior*, 7(11): **1456–1466.**

Hedge JE and Hofreiter BT. (1962). In: Methods in Carbohydrate chemistry 17. (Eds. Whistler, R.L. and Be Miller, J. N.) *Academic Press, New York.*

Herath HT, Tharuka KG, Gunathilake Eashwarage I, Sivakumaran K, Ranathunga RA. (2018). Physico-chemical and in vitro glycemic indices of popular pulse varieties grown in Sri Lanka. *International Journal of Food Sciences and Nutrition*, 3(5): **137–1343.**

Hills MJ. (2004). Control of storage–product synthesis in seeds. *Current opinion in plant biology*, 7(3): **302–308.**

Hiscox JD and Israelstam GF. (1979). A method for the extraction of chlorophyll from leaf tissue without maceration. *Canadian journal of botany*, 57(12) **1332–1334.**

Hobbs DH, Flintham JE and Hills MJ. (2004). Genetic control of storage oil synthesis in seeds of *Arabidopsis*. *Plant physiology*, 136(2): **3341–3349.**

Hu C, Tohge T, Chan SA, Song Y, Rao J, Cui B ... and Shi, J. (2016). Identification of conserved and diverse metabolic shifts during rice grain development. *Scientific Reports*, 6(1): **1–12.**

Huang J, Guo X, Hao X, Zhang W, Chen S, Huang R, Gresshoff PM, Zheng Y. (2016). De novo sequencing and characterization of seed transcriptome of the tree legume *Millettia pinnata* for gene discovery and SSR marker development. *Molecular Breeding***,** 36**: 1–5.**

Huang J, Hao X, Jin Y, Guo X, Shao Q, Kumar KS, Ahlawat YK, Harry DE, Joshi CP, Zheng Y. (2018). Temporal transcriptome profiling of developing seeds reveals a concerted gene regulation in relation to oil accumulation in Pongamia (*Millettia pinnata*). *BMC plant biology***,** 18**: 1–6.**

Hudson KA and Hudson ME. (2021). Genetic Variation for Seed Oil Biosynthesis in Soybean. *Plant Molecular Biology Reporter*, 39: **700–709.**

Jain S. (2019). The production of biodiesel using Karanja (*Pongamia pinnata*) and Jatropha (*Jatropha curcas*) Oil. *Biomass, Biopolymer–Based Materials, and Bioenergy*: **397–408.**

Jander G and Joshi V. (2009). Aspartate–derived amino acid biosynthesis in *Arabidopsis thaliana*. *The Arabidopsis book/American Society of Plant Biologists*, **7.**

Jiang H, Wu P, Zhang S, Song C, Chen Y, Li M, ... and Wu G. (2012). Global analysis of gene expression profiles in developing physic nut (*Jatropha curcas* L.) seeds. *PLOS one*, 7(5): **e36522**.

Jin Y, Liu L, Hao X, Harry DE, Zheng Y, Huang T and Huang J. (2019). Unravelling the microRNA–mediated gene regulation in developing Pongamia seeds by high–throughput small RNA profiling. *International journal of molecular sciences*, 20(14): **3509.**

Kapilan N and Reddy RP. (2008). Evaluation of methyl esters of mahua oil (*Madhuca indica*) as diesel fuel. *Journal of the American Oil Chemists' Society*, 85: **185–188.**

Karmee SK and Chadha A. (2005). Preparation of biodiesel from crude oil of *Pongamia pinnata*. *Bioresource technology*, 96(13): **1425–1429.**

Kesari V and Rangan L. (2011). Coordinated changes in storage proteins during development and germination of elite seeds of *Pongamia pinnata*, a versatile biodiesel legume. *AoB plants*: **2011.**

Kesari V, Ramesh AM and Rangan L. (2012). High frequency direct organogenesis and evaluation of genetic stability for in vitro regenerated *Pongamia pinnata*, a valuable biodiesel plant. *Biomass and Bioenergy*, 44: **23–32.**

Khayoon MS, Olutoye MA and Hameed BH. (2012). Utilization of crude karanj (*Pongamia pinnata*) oil as a potential feedstock for the synthesis of fatty acid methyl esters. *Bioresource Technology*, 111: **175–179.**

Khurana E and Singh JS. (2000). Influence of seed size on seedling growth of *Albizia procera* under different soil water levels. *Annals of Botany*, 86(6): **1185–1192.**

Kong Q, Yang Y, Guo L, Yuan L and Ma W (2020). Molecular Basis of Plant Oil Biosynthesis: Insights Gained from Studying the WRINKLED1 Transcription Factor. *Frontiers in Plant Science*, 11: **24.**

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, ... and Steinhauser D. (2005). GMD@ CSB. DB: the Golm metabolome database. *Bioinformatics*, 21(8): **1635–1638.**

Kühnreich R and Holzgrabe U. (2016). High-performance liquid chromatography evaluation of the enantiomeric purity of amino acids by means of automated precolumn derivatization with ortho-phthalaldehyde and chiral thiols. *Chirality*, 28(12): **795–804.**

Kumar A and Sharma S. (2011). Potential non-edible oil resources as biodiesel feedstock: an Indian perspective. *Renewable and Sustainable Energy Reviews*, 15(4): **1791–1800.**

Kumar D, Tripathi DK, Liu S, Singh VK, Sharma S, Dubey NK, ... and Chauhan DK. (2017). *Pongamia pinnata* (L.) Pierre tree seedlings offer a model species for arsenic phytoremediation. *Plant Gene*, 11: **238–246.**

Kumar N, Chaudhary A, Singh D and Teotia S. (2020). Transcriptional regulation of seed oil accumulation in *Arabidopsis thaliana*: role of transcription factors and chromatin remodelers. *Journal of Plant Biochemistry and Biotechnology*, 29(4): **754–768.**

Kumar R and Pal P. (2021). Lipase immobilized graphene oxide biocatalyst assisted enzymatic transesterification of *Pongamia pinnata* (Karanja) oil and downstream enrichment of biodiesel by solar–driven direct contact membrane distillation followed by ultrafiltration. *Fuel Processing Technology*, 211: **106577.**

Laux T, Wurschum T, Breuninger H. (2004). Genetic regulation of embryonic pattern formation. *The Plant Cell*, 16**: 190–202.**

Le Brandon H, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ and Goldberg RB. (2007). Using genomics to study legume seed development. *Plant physiology*, 144(2): **562–574.**

Le Signor C, Aimé D, Bordat A, Belghazi M, Labas V, Gouzy J, ... and Gallardo K. (2017). Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. *New Phytologist*, 214(4): **1597–1613.**

Li J, Jiang Z, Li X, Hou Y, Liu F, Li N ... and Chen G. (2015). Natural therapeutic agents for neurodegenerative diseases from a traditional herbal medicine *Pongamia pinnata* (L.) Pierre**.** *Bioorganic & Medicinal Chemistry Letters*, 25(1): **53–58.**

Liao S, Wang L, Li J, Ruan YL. (2020). Cell wall invertase is essential for ovule development through sugar signaling rather than provision of carbon nutrients. *Plant Physiology*. 183(3): **1126–1144.**

Lin H, Li Y, Hill RT. (2022). Microalgal and bacterial auxin biosynthesis: implications for algal biotechnology. *Current Opinion in Biotechnology*. 73: **300–307.**

Liu B, Hua C, Song G, Wu M, Cui R, Zhang A, ... and Gan Y. (2017). The SPATULA transcription factor regulates seed oil content by controlling seed specific genes in *Arabidopsis thaliana*. *Plant Growth Regulation*, 82(1): **111–121.**

Liu F, Zhang X, Lu C, Zeng X, Li Y, Fu D and Wu G. (2015). Non–specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. *Journal of experimental botany*, 66(19): **5663–5681.**

Livak KJ and Schmittgen TD. (2001). Analysis of relative gene expression data using real–time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 25(4): **402–408.**

Luthra R, Munshi SK and Sukhija PS. (1991). Relationship of carbohydrate metabolism with lipid biosynthesis in developing sunflower (*Helianthus annuus* L.) seeds. *Journal of plant physiology*, 137(3): **312–318.**

Madsen CK and Brinch-Pedersen H. (2020). Globoids and phytase: the mineral storage and release system in seeds. *International Journal of Molecular Sciences*, 21(20): **7519.**

McAtee P, Karim S, Schaffer R and David K. (2013). A dynamic interplay between phytohormones is required for fruit development, maturation, and ripening**.** *Frontiers in plant science*, 4: **79.**

Melkus G, Rolletschek H, Radchuk R, Fuchs J, Rutten T, Wobus U ... and Borisjuk L. (2009). The metabolic role of the legume endosperm: a noninvasive imaging study *Plant physiology*, 151(3): **1139–1154.**

Memon AR. (2004). The role of ADP–ribosylation factor and SAR1 in vesicular trafficking in plants. *Biochimica et Biophysica Acta (BBA)–Biomembranes*, 1664(1): **9–30.**

Miklaszewska M, Zienkiewicz K, Inchana P, Zienkiewicz A. (2021). Lipid metabolism and accumulation in oilseed crops. *OCL - Oilseeds and fats, Crops and Lipids*, 28: **50.**

Miranda M, Borisjuk L, Tewes A, Heim U, Sauer N, Wobus U and Weber H. (2001). Amino acid permeases in developing seeds of *Vicia faba* L.: expression precedes storage protein synthesis and is regulated by amino acid supply. *The Plant Journal*, 28(1): **61–71.**

Miron D and Schaffer AA. (1991). Sucrose phosphate synthase, sucrose synthase, and invertase activities in developing fruit of *Lycopersicon esculentum* Mill. and the sucrose accumulating *Lycopersicon hirsutum* Humb. and Bonpl. *Plant Physiology*, 95(2): **623–627.**

Monteiro MS, Carvalho M, Bastos ML and Guedes de Pinho P. (2013). Metabolomics analysis for biomarker discovery: advances and challenges. *Current medicinal chemistry*, 20(2): **257–271.**

Moriguchi T, Abe K, Sanada T, Yamaki S. (1991). Levels and role of sucrose synthase, sucrose-phosphate synthase, and acid invertase in sucrose accumulation in fruit of Asian pear. *Journal of the American Society for Horticultural Science*, 117(2): **274–278.**

Muktham R, Ball AS, Bhargava SK and Bankupalli S. (2016). Bioethanol production from non–edible de–oiled *Pongamia pinnata* seed residue–optimization of acid hydrolysis followed by fermentation. *Industrial Crops and Products*, 94: **490–497.**

Murray DR. (1987). "Nutritive role of seedcoats in developing legume seeds." *American Journal of Botany*, 74.7: **1122–1137**.

Nagar DS, Jha SK and Jani J. (2015). Direct adventitious shoot bud formation on hypocotyls explants in *Millettia pinnata* (L.) Panigrahi–a biodiesel producing medicinal tree species. *Physiology and Molecular Biology of plants*, 21(2): **287–292.**

Nwafor CC, Li D, Qin P, Li L, Zhang W, Zhou Y, Xu J, Yin Y, Cao J, He L, Xiang F. (2022). Genetic and biochemical investigation of seed fatty acid accumulation in *Arabidopsis*. *Frontiers in Plant Science*;13: **942054.**

Nayak SK and Mishra PC. (2017). Emission from a dual fuel operated diesel engine fuelled with *Calophyllum Inophyllum* biodiesel and producer gas. *International Journal of Automotive and Mechanical Engineering*, 14(1): **3954–3969**.

Nayak SK, Behera GR, Mishra PC and Sahu SK. (2017). Biodiesel vs diesel: A race for the future. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 39(14):**1453–1460.**

Niu Y, Wu GZ, Ye R, Lin WH, Shi QM, Xue LJ, ... and Xue HW. (2009). Global analysis of gene expression profiles in Brassica napus developing seeds reveals a conserved lipid metabolism regulation with *Arabidopsis thaliana*. *Molecular plant*, 2(5): **1107–1122.**

Obendorf RL and Górecki RJ. (2012). Soluble carbohydrates in legume seeds. *Seed Science Research*, 22(4): **219–242.**

Ohlrogge J and Browse J. (1995). Lipid biosynthesis. *The Plant cell*, 7(7): **957–970.**

Ortiz–Martínez VM, Salar–García MJ, Palacios–Nereo FJ, Olivares–Carrillo P, Quesada–Medina J, De Los Ríos AP and Hernández–Fernández FJ. (2016). In–depth study of the transesterification reaction of *Pongamia pinnata* oil for biodiesel production using catalyst–free supercritical methanol process. *The Journal of Supercritical Fluids*, 113: **23–30.**

Parreira JR, Branco D, Almeida AM, Czubacka A, Agacka-Mołdoch M, Paiva JA, Tavares-Cadete F, Araújo SD. (2016). Systems biology approaches to improve drought stress tolerance in plants: state of the art and future challenges. Drought Stress Tolerance in Plants**.** *Molecular and Genetic Perspectives* 2: **433–71.**

Patnaik P, Abbasi T and Abbasi SA. (2017). Prosopis (*Prosopis juliflora*): blessing and bane. *Tropical Ecology*, 58(3): **455–483.**

Patwardhan M. Magical *Pongamia pinnata* (L.). *Medicinal Plant Treasures of India*.:**21**.

Pavithra HR, Gowda B and Shivanna MB. (2014). Biochemical changes in the composition of developing seeds of *Pongamia pinnata* (L.) Pierre. *Industrial Crops and Products*, 53: **199–208.**

Pavithra HR, Gowda B, Prasanna KT, Shivanna MB. (2013). Pod and seed traits in candidate plus trees of *Pongamia pinnata* (L.) Pierre from southern peninsular India in relation to provenance variation and genetic variability. *Journal of Crop Science and Biotechnology*, 16: **131–142**.

Pavithra HR, Gowda B, Rajesh Kumar K, Prasanna KT and Shivanna MB. (2012). Oil, fatty acid profile and karanjin content in developing *Pongamia pinnata* (L.) Pierre seeds. *Journal of the American Oil Chemists' Society*, 89(12): **2237–2244.**

Pego JV and Smeekens SC. (2000). Plant fructokinases: a sweet family get-together. *Trends in plant science*. 5(12): **531–536.**

Puhan S, Saravanan N, Nagarajan G and Vedaraman N. (2010). Effect of biodiesel unsaturated fatty acid on combustion characteristics of a DI compression ignition engine. *Biomass and Bioenergy*, 34(8): **1079–1088.**

Rajarajan K, Sakshi S, Taria S, Prathima PT, Radhakrishna A, Anuragi H, Ashajyothi M, Bharati A, Handa AK, Arunachalam A. (2022). Whole plant response of *Pongamia pinnata* to drought stress tolerance revealed by morpho-physiological, biochemical and transcriptome analysis. *Molecular Biology Reports*, 49(10): **9453–63.**

Rajavel A, Klees S, Schlüter JS, Bertram H, Lu K, Schmitt AO and Gültas M. (2021). Unravelling the Complex Interplay of Transcription Factors Orchestrating Seed Oil Content in *Brassica napus* L. *International journal of molecular sciences*, 22(3): **1033.**

Ramesh AM, Kesari V and Rangan L. (2014). Characterization of a stearoyl–acyl carrier protein desaturase gene from potential biofuel plant, *Pongamia pinnata* L**.** *Gene*, 542(2): **113–121.**

Rende U, Wang W, Gandla ML, Jönsson LJ, Niittylä T. (2017). Cytosolic invertase contributes to the supply of substrate for cellulose biosynthesis in developing wood. *New Phytologist*, 214(2): **796–807.**

Renz A and Stitt M. (1993). Substrate specificity and product inhibition of different forms of fructokinases and hexokinases in developing potato tubers. *Planta*, 190(2): **166–75.**

Rodionova MV, Poudyal RS, Tiwari I, Voloshin RA, Zharmukhamedov SK, Nam HG, Zayadan BK, Bruce BD, Hou HJ and Allakhverdiev SI. (2017). Biofuel production challenges and opportunities. *International Journal of Hydrogen Energy*, 42(12): **8450–8461.**

Rodrigues–Junior AG, Santos MT, Hass J, Paschoal BS and De–Paula OC. (2020). What kind of seed dormancy occurs in the legume genus Cassia? *Scientific Reports*, 10(1): **1–11.**

Rodriguez RJ, Henson J, Van Volkenburgh E, Hoy M, Wright L, Beckwith F, ... and Redman RS. (2008). Stress tolerance in plants via habitat–adapted symbiosis. *The ISME journal*, 2(4): **404–416.**

Rolletschek H, Radchuk R, Klukas C, Schreiber F, Wobus U and Borisjuk L. (2005). Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. *New Phytologist*, 167(3): **777–786.**

Ruuska SA, Schwender J, Ohlrogge JB. (2004). The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. *Plant physiology*, 136(1):**2700–2709.**

Sameer Kumar CV, Satheesh Naik SJ, Mohan N, Saxena RK, Varshney RK. (2017). Botanical description of pigeonpea [*Cajanus cajan* (L.) Millsp.]. *The pigeonpea genome*. 2017:**17–29.**

Sangwan S, Rao DV and Sharma RA. (2010). A review on *Pongamia pinnata* (L.) Pierre: A great versatile leguminous plant. *Nature and Science*, 8(11): **130–139.**

Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M and Lepiniec L. (2008). Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *The Plant Journal*, 54(4): **608–620.**

Sanz L, Albertos P, Mateos I, Sánchez–Vicente I, Lechón T, Fernández–Marcos M and Lorenzo O. (2015). Nitric oxide (NO) and phytohormones crosstalk during early plant development. *Journal of experimental botany*, 66(10). **2857–2868.**

Savadi S. (2018). Molecular regulation of seed development and strategies for engineering seed size in crop plants. *Plant Growth Regulation*, 84(3): **401–422**.

Sbihi HM, Nehdi IA, Mokbli S, Romdhani–Younes M and Al–Resayes SI. (2018). Study of oxidative stability and cold flow properties of *Citrillus colocynthis* oil and *Camelus dromedaries* fat biodiesel blends. *Industrial Crops and Products*, 122: **133–141.**

Schiltz S, Munier–Jolain N, Jeudy C, Burstin J and Salon C. (2005). Dynamics of exogenous nitrogen partitioning and nitrogen remobilization from vegetative organs in pea revealed by 15N in vivo labeling throughout seed filling. *Plant Physiology*, 137(4): **1463–1473.**

Schrimpe–Rutledge AC, Codreanu SG, Sherrod SD and McLean JA. (2016). Untargeted metabolomics strategies—challenges and emerging directions. *Journal of The American Society for Mass Spectrometry*, 27(12): **1897–1905.**

Schwender J, Goffman F, Ohlrogge JB and Shachar–Hill Y. (2004). Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, 432(7018): **779–782.**

Schwender J, Hebbelmann I, Heinzel N, Hildebrandt T, Rogers A, Naik D, ... and Rolletschek H. (2015). Quantitative multilevel analysis of central metabolism in developing oilseeds of oilseed rape during in vitro culture. *Plant Physiology*, 168(3): **828–848.**

Scott PT, Pregelj L, Chen N, Hadler JS, Djordjevic MA and Gresshoff PM. (2008). *Pongamia pinnata*: an untapped resource for the biofuels industry of the future**.** *Bioenergy Research*, 1(1): **2–11.**

Sengupta D, Kannan M and Reddy AR. (2011). A root proteomics–based insight reveals dynamic regulation of root proteins under progressive drought stress and recovery in *Vigna radiata* (L.) Wilczek. *Planta*, 233(6): **1111–1127.**

Shang X, Zhu Y, Chen X, Wang XD, Rose RJ, Song Y. (2020). Seed oil storage in three contrasted legume species: Implications for oil improvement. *Acta Physiologiae Plantarum*, 42**:1–2.**

Shao Q, Liu X, Su T, Ma C and Wang P. (2019). New insights into the role of seed oil body proteins in metabolism and plant development. *Frontiers in Plant Science*, 10: **1568.**

Sharma S, Guleria S, Gill BS and Munshi SK. (2011). Lipid accumulation in developing soybean: influence of seed position on stem axis. Genetic *Plant Physiology*, 1: **56–67.**

Sharma SS, Islam MA, Malik AA, Kumar K, Negi MS, Tripathi SB. (2016). Seed traits, fatty acid profile and genetic diversity assessment in *Pongamia pinnata* (L.) Pierre germplasm. *Physiology and Molecular Biology of Plants*, 22:**193–205.**

Sharma A, Kaushik N and Rathore H. (2020). Karanja (*Milletia pinnata* (L.) Panigrahi): a tropical tree with varied applications. *Phytochemistry Reviews*, 19(3): **643–658.**

Shevkani K, Singh N, Chen Y, Kaur A, Yu L. (2019). Pulse proteins: Secondary structure, functionality and applications. *Journal of food science and technology*. 56: **2787–98.**

Shewry PR, Napier JA and Tatham AS. (1995). Seed storage proteins: structures and biosynthesis. *The plant cell*, 7(7): **945.**

Shilpi JA, Islam ME, Billah M, Islam KMD, Sabrin F, Uddin SJ, ... and Sarker SD. (2012). Antinociceptive, anti–inflammatory, and antipyretic activity of mangrove plants: A mini review. *Advances in pharmacological sciences*, 2012.

Shoaib M, Hussain S, Cheng X, Cui Y, Liu H, Chen Q, ... and Yu X. (2021). Synergistic anti–oxidative effects of *Pongamia pinnata* against nickel mediated by *Rhizobium pisi* and *Ochrobacterium pseudogrignonense*. *Ecotoxicology and Environmental Safety*, 217: **112244.**

Siddiqi S and Mansbach CM. (2012). Phosphorylation of Sar1b protein releases liver fatty acid–binding protein from multiprotein complex in intestinal cytosol enabling it to bind to endoplasmic reticulum (ER) and bud the pre–chylomicron transport vesicle. *Journal of Biological Chemistry*, 287(13): **10178–10188.**

Sims RE, Mabee W, Saddler JN, Taylor M. (2010). An overview of second-generation biofuel technologies. *Bioresource technology*, 101(6): **1570–1580.**

Singh N, Sandhu KS, Kaur M. (2004). Characterization of starches separated from Indian chickpea (*Cicer arietinum* L.) cultivars. *Journal of Food Engineering.* 63(4): **441–449.**

Singh M, Jatoliya JK and Saharan V. (2016). Direct organogenesis of *Pongamia pinnata* (L.) Pierre from cotyledonary segments: a valuable *Brassica napus* biodiesel tree. *Journal of Cell and Tissue Research*, 16(1): **5399.**

Singh Y, Sharma A, Singh GK, Singla A and Singh NK. (2018). Optimization of performance and emission parameters of direct injection diesel engine fueled with Pongamia methyl esters–response surface methodology approach. *Industrial crops and products*, 126: **218–226.**

Singha KT, Sreeharsha RV, Mariboina S and Reddy AR. (2019). Dynamics of metabolites and key regulatory proteins in the developing seeds of *Pongamia pinnata*, a potential biofuel tree species. *Industrial Crops and Products*, 140: **111621.**

Singha WR, Kurmi B, Sahoo UK, Sileshi GW, Nath AJ and Das AK. (2021). *Parkia roxburghii*, an underutilized tree bean for food, nutritional and regional climate security. *Trees, Forests and People*, 4: **100065.**

Sivakumar V, Anandalakshmi R, Warrier RR, Singh BG, Tigabu M and Nagarajan B. (2013). Discrimination of Acacia seeds at species and subspecies levels using an image analyzer. *Forest Science and Practice*, 15(4): **253–260.**

Snyder CL, Yurchenko OP, Siloto RM, Chen X, Liu Q, Mietkiewska E and Weselake RJ. (2009). Acyltransferase action in the modification of seed oil biosynthesis. *New biotechnology*, 26(1–2): **11–16.**

Song Y, Wang XD and Rose RJ. (2017). Oil body biogenesis and biotechnology in legume seeds. Plant Cell Reports, 36(10): **1519–1532.**

Sreeharsha RV, Mudalkar S, Singha KT and Reddy AR. (2016). Unravelling molecular mechanisms from floral initiation to lipid biosynthesis in a promising biofuel tree species, *Pongamia pinnata* using transcriptome analysis. *Scientific reports*, 6(1): **1–15.**

Sreeharsha RV, Mudalkar S, Reddy AR. (2022). Genome sequencing and analysis uncover the regulatory elements involved in the development and oil biosynthesis of *Pongamia pinnata* (L.)–A potential biodiesel feedstock. *Frontiers in Plant Science*, 13: **747783.**

Stein O, Avin-Wittenberg T, Krahnert I, Zemach H, Bogol V, Daron O, Aloni R, Fernie AR, Granot D. (2017). Arabidopsis fructokinases are important for seed oil accumulation and vascular development. *Frontiers in plant science*, 7: **2047.**

Stein O and Granot D. (2018). Plant fructokinases: evolutionary, developmental, and metabolic aspects in sink tissues. *Frontiers in Plant Science*, 9: **339.**

Sun C, Jia L, Xi B, Wang L and Weng X. (2017). Natural variation in fatty acid composition of *Sapindus spp*. seed oils. *Industrial Crops and Products*, 102: 97–104.

Tan H, Xie Q, Xiang X, Li J, Zheng S, Xu X, ... and Ye W. (2015). Dynamic metabolic profiles and tissue–specific source effects on the metabolome of developing seeds of. *Brassica napus. PloS one*, 10(4): **e0124794.**

Tan SN, Tee CS and Wong HL. (2018). Multiple shoot bud induction and plant regeneration studies of *Pongamia pinnata*. *Plant Biotechnology*, 18: **0711.**

Tanaka, Y., Sano, T., Tamaoki, M., Nakajima, N., Kondo, N. and Hasezawa, S., 2006. Cytokinin and auxin inhibit abscisic acid-induced stomatal closure by enhancing ethylene production in Arabidopsis. *Journal of experimental botany*, 57(10), **2259-2266.**

under elevated CO_2 using transcriptome sequencing and assembly. *Scientific reports*, 7(1): **1–15.**

Tang CH and Sun X. (2011). A comparative study of physicochemical and conformational properties in three vicilins from Phaseolus legumes: implications for the structure–function relationship. *Food Hydrocolloids*, 25(3): **315–324.**

Teh HF, Neoh BK, Ithnin N, Daim LDJ, Ooi TEK and Appleton DR. (2017). Omics and strategic yield improvement in oil crops. *Journal of the American Oil Chemists' Society*, 94(10): **1225–1244.**

Tharanathan RN and Mahadevamma S. (2003). Grain legumes—a boon to human nutrition. *Trends in Food Science & Technology*, 14(12): **507–518.**

Thompson R, Burstin J and Gallardo K. (2009). Post–genomics studies of developmental processes in legume seeds. *Plant Physiology*, 151(3): **1023–1029.**

Tilman D, Hill J and Lehman C. (2006). Carbon–negative biofuels from low–input high–diversity grassland biomass. *Science*, 314(5805): **1598–1600.**

Tohge T, Scossa F and Fernie AR. (2015). Integrative approaches to enhance understanding of plant metabolic pathway structure and regulation. *Plant Physiology*, 169(3): **1499–1511.**

Tong L and Harwood Jr HJ. (2006). Acetyl-coenzyme A carboxylases: versatile targets for drug discovery. *Journal of cellular biochemistry*, 99(6): **1476–1488.**

Tullu ABE, Kusmenoglu ISMAIL, McPhee KE and Muehlbauer FJ. (2001). Characterization of core collection of lentil germplasm for phenology, morphology, seed and straw yields. *Genetic Resources and Crop Evolution*, 48(2): **143–152.**

Tyson KS. 2006. Biodiesel handling and use guidelines **(Vol. 540, No. 40555)**. US Department of Energy, *Energy Efficiency and Renewable Energy*.

Vakshasya RK, Rajora OP and Rawat MS. (1992). Seed and seedling traits of *Dalbergia sissoo* Roxb.: seed source variation studies among ten sources in India. *Forest Ecology and Management*, 48(3–4): **265–275.**

Verdier J and Thompson RD. (2008). Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant and Cell Physiology*, 49(9): **1263–1271.**

Wang H, Guo J, Lambert KN and Lin Y. (2007). Developmental control of *Arabidopsis* seed oil biosynthesis. *Planta*, 226(3): **773–783.**

Wang TL, Domoney C, Hedley CL, Casey R and Grusak MA. (2003). Can we improve the nutritional quality of legume seeds? *Plant Physiology*, 131(3): **886–891.**

Wang C, Hai J, Yang J, Tian J, Chen W, Chen T, Luo H, Wang H. (2016). Influence of leaf and silique photosynthesis on seeds yield and seeds oil quality of oilseed rape (*Brassica napus* L.). *European Journal of Agronomy*, 74: **112–118.**

Wang X, Liang H, Guo D, Guo L, Duan X, Jia Q and Hou X. (2019). Integrated analysis of transcriptomic and proteomic data from tree peony (*P. ostii*) seeds reveals key developmental stages and candidate genes related to oil biosynthesis and fatty acid metabolism. *Horticulture Research*, 6: 111.

Wani IA, Sogi DS, Hamdani AM, Gani A, Bhat NA and Shah A. (2016). Isolation, composition, and physicochemical properties of starch from legumes: A review. *Starch-Stärke*, 68(9–10): **834–845.**

Weber H, Borisjuk L and Wobus U. (1997). Sugar import and metabolism during seed development. *Trends in plant science*, 2(5): **169–174.**

Weber H, Borisjuk L and Wobus U. (2005). Molecular physiology of legume seed development. *Annual Review on Plant Biology*, 56: **253–279.**

Weber H, Golombek S, Heim U, Borisjuk L, Panitz R, Manteuffel R and Wobus U. (1998). Integration of carbohydrate and nitrogen metabolism during legume seed development: implications for storage product synthesis. *Journal of plant physiology*, 152(6): **641–648.**

Weber H, Sreenivasulu N and Weschke W. (2010). Molecular physiology of seed maturation and seed storage protein biosynthesis. *In Plant developmental biology–biotechnological perspectives***: 83–104.** Springer, Berlin, Heidelberg

Wei H, Chai S, Ru L, Pan L, Cheng Y, Ruan M, Ye Q, Wang R, Yao Z, Zhou G, Chen Y. (2020). New insights into the evolution and expression dynamics of invertase gene family in *Solanum lycopersicum. Plant Growth Regulation*, 92: **205–217.**

Welham T, Pike J, Horst I, Flemetakis E, Katinakis P, Kaneko T, Sato S, Tabata S, Perry J, Parniske M, Wang TL. (2009). A cytosolic invertase is required for normal growth and cell development in the model legume, *Lotus japonicus*. *Journal of Experimental Botany*. 60(12): **3353–3365.**

Weselake RJ, Taylor DC, Rahman MH, Shah S, Laroche A, McVetty PB and Harwood JL. (2009). Increasing the flow of carbon into seed oil. *Biotechnology advances*, 27(6): **866–878.**

Winarto HP, Liew LC, Gresshoff PM, Scott PT, Singh MB and Bhalla PL. (2015). Isolation and characterization of circadian clock genes in the biofuel plant Pongamia (*Millettia pinnata*). *Bioenergy Research*, 8(2): **760–774.**

Wishart DS. (2011). Advances in metabolite identification. *Bioanalysis*, 3(15): **1769–1782.**

Wobus U and Weber H. (1999). Sugars as signal molecules in plant seed development. *Biological chemistry*, 380(7–8): **937–944.**

Wright D and Lenssen A. (2013). Staging soybean development.

Wu P, Liu A, Li L. (2021). Metabolomics and transcriptome analysis of the biosynthesis mechanism of flavonoids in the seeds of *Euryale ferox* Salisb at different developmental stages. *Molecular Genetics and Genomics*, 296(4): **953–70.**

Xiong B, Zhang Z and Dong S. (2018). Biodiesel from *Lindera glauca* oil, a potential non–food feedstock in Southern China. *Industrial Crops and Products*, 122: **107–113.**

Yoon J, Cho LH, Tun W, Jeon JS and An G. (2021). Sucrose signaling in higher plants. *Plant Science*, 302: **110703.**

Yu B, Xiang D, Mahfuz H, Patterson N and Bing D. (2021). Understanding starch metabolism in pea seeds towards tailoring functionality for value–added utilization. *International Journal of Molecular Sciences*, 22(16): **8972.**

Yu X, Shen T, Kang X, Cui Y, Chen Q, Shoaib M ... and Liang Y. (2021). Long–term phytoremediation using the symbiotic *Pongamia pinnata* reshaped soil micro–ecological environment. *Science of the Total Environment*, 774: **145112.**

Zhang J, Wu Z, Hu F, Liu L, Huang X, Zhao J and Wang H. (2018). Aberrant seed development in *Litchi chinensis* is associated with the impaired expression of cell wall invertase genes. *Horticulture research* **5.**

Zhao L, Pan T, Cai C, Wang J and Wei C. (2016). Application of whole sections of mature cereal seeds to visualize the morphology of endosperm cell and starch and the distribution of storage protein. *Journal of Cereal Science*, 71: **19–27.**

PUBLICATIONS

Industrial Crops & Products 140 (2019) 111621

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Dynamics of metabolites and key regulatory proteins in the developing seeds of *Pongamia pinnata*, a potential biofuel tree species

Kambam Tamna Singha[a], Rachapudi Venkata Sreeharsha[a], Sureshbabu Mariboina[a], Attipalli Ramachandra Reddy[a,b,*]

[a] Dept. of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, 500046, India
[b] Yogi Vemana University, Kadapa, Andhra Pradesh, India

ARTICLE INFO

Keywords:
Pongamia pinnata
Seed development
Metabolome
Proteome
Triacylglycerols
Oil body

ABSTRACT

The present study analyzed the dynamic changes in metabolites and key proteins during the seed development of *Pongamia pinnata* L. (Family: Fabaceae) with a particular focus on lipid biosynthesis and oil accumulation. The developing seeds were collected at four different stages: 120 (stage 1), 180 (stage 2), 240 (stage 3) and 300 (stage 4) days after flowering (DAF), representing S1, S2, S3 and S4 respectively. The analysis of seed pigments and mRNA expression patterns of key photosynthetic genes confirmed the photo-autotrophic behavior of *P. pinnata* seed during the initial stages of development. The metabolite profiling of developing *P. pinnata* seeds also revealed differentially expressed sugars, amino acids, free fatty acids and organic acids. Proteins related to development, energy metabolism, lipid accumulation as well as stress responses were documented through MALDI-TOF-MS/MS analysis. The structure and pattern of oil body accumulation at each stage of seed development were determined by electron and confocal microscopy of the cotyledonary sections. The thin layer chromatogram of *P. pinnata* oil revealed higher amount of Triacylglycerides and the fatty acid profile of extracted triacylglycerides showed a rapid increase in oleic acid (C18:1) at S3 and S4. The outcomes reveal new insights into the complex oleogenic metabolism during *P. pinnata* seed development at macro level.

1. Introduction

Plant oils have a potential to replace petroleum based fuels by acting as feedstock for the oleochemical and biofuel industries. Understanding the regulatory networks of seed oil biosynthesis has been a target of agronomists and metabolic engineers for decades in order to enhance the oil production. The current knowledge on regulation of seed development and oil accumulation is limited to food crops and other model plants. While, very limited reports are available for biofuel plants including *Pongamia pinnata, Jaropha curcus, Camelina sativa* (Mudalkar et al., 2014; Chaitanya et al., 2015; Kumar et al., 2017). In general, the seed development bears a complex regulatory network of metabolic and developmental machinery which include signaling elements associated with accumulation of storage compounds (Baud et al., 2008; Atabani et al., 2013). Further, the seed storage-compound synthesis overlays the developmental progression of embryogenesis and is to an extent governed by the metabolite and hormonal signals inside the embryo (Borisjuk et al., 2003). The proportional distribution of seed storage products into carbohydrates, oils and proteins will depend on the influx of metabolic pathways during the development and can vary between species (Schiltz et al., 2005; Ekman et al., 2008; Chaitanya et al., 2015). For oil development, carbon is delivered to fatty acid synthesis in seed plastids via glycolysis with hexose and triose as the predominant carbohydrates entering the plastid. Previous reports signified the role of sucrose import into embryo and its metabolism in the cytosol and plastids in the formation of starch and oil (Luthra et al., 1991; Eastmond and Rawsthorne, 2000). It is well established in several model legumes that the dynamic changes of hexose/sucrose ratios play a key role in the commencement of storage product synthesis as the seed development progressed. On the

Abbreviations: ACP, acyl carrier proteins; BSA, bovine serum albumin; DAF, days after flowering; DAGs, diacylglycerides; DGAT, acyl-coA: diglyceride acyltransferase; DTT, dithiothreitol; DW, dry weight; EGTA, ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid; FAMEs, fatty acid methyl esters; FFA, Free Fatty acids; FW, fresh weight; NADGA, N-acetyl- D-glucosamine; PBS, phosphate buffer saline; PCA, Principle component analysis; PDAT, phospholipid:diacylglycerol acyltransferase; PL, Phospholipids; PMSF, phenylmethylsulfonyl fluoride; TAGs, triacylglycerides; TCA, tricarboxylic acid; TLC, thin layer chromatography

* Corresponding author.
E-mail address: arrsl@uohyd.ernet.in (A.R. Reddy).

https://doi.org/10.1016/j.indcrop.2019.111621
Received 10 December 2018; Received in revised form 15 June 2019; Accepted 27 July 2019

other hand, the primary metabolism inside the seed which includes major pathways of glycolysis and TCA link the biosynthesis of carbohydrates, amino acids and fatty acids which also provide energy balance inside the developing seeds (Rolletschek et al., 2005; Schwender et al., 2015). Hence, the precise understanding of changes in these regulatory metabolites will provide insights into the seed development and oil composition in non-model legume tree species, specifically *P. pinnata*.

The oilseeds can be photo-autotrophic or heterotrophic depending on the presence of chloroplast and light plays an important role in regulating the green oil seed development. It was reported in certain crop plants that the seed carries out photosynthesis during embryogenesis which provides energy for FA biosynthesis and also helps in refixing the respiratory CO_2 release (Schwender et al., 2004). Though the absence of seed photosynthesis will not affect the overall FA biosynthesis, it will be considered as an important factor in promoting carbon storage, energy flux and lipid biosynthesis in oilseeds. The dynamic shift in the expression patterns of photosystem related proteins along with FA biosynthesis enzymes during the seed development of *Arabidopsis* and *Glycine max* support the active role of seed photosynthesis. Hence, studying the pigments and associated protein expression patterns in large seeds with long developmental cycles will provide more insights into the role of seed photosystems in synthesizing storage compounds.

The cultivation of *P. pinnata* is suitable in diverse tropical and subtropical environments. The oil content of *P. pinnata* seeds ranges from 35 to 40% of seed dry weight and 55–65 % of the total lipid has been reported as oleic acid which is the ideal fatty acid for biodiesel production (Sreeharsha et al., 2016; Xiong et al., 2018). Due to the accumulation of high levels of poly unsaturated fatty acids in the *P. pinnata* seed oil, it is considered as a potential biofuel feedstock with an optimized efficiency to use as a biodiesel for diesel engines (Singh et al., 2018; Jain et al., 2018). Moreover, not only the *P. pinnata* oil, but also the seed waste after oil extraction have recently been used to produce bio-ethanol (Muktham et al., 2016). The biochemical characteristics and protein profiling during seed development has been investigated in *P. pinnata* (Kesari and Rangan, 2011; Pavithra et al., 2014). The proximal chemical composition of mature *P. pinnata* seed along with fatty acid composition were also reported earlier (Bala et al., 2011; Sharma et al., 2011; Pavithra et al., 2012). However, in depth analysis of metabolite and proteomic profiles throughout the lengthy developmental period of *P. pinnata* seed have not yet been reported. It is crucial to study the metabolome and proteome of the developing seed to understand the complex association between various metabolites and FA biosynthesis to oil body formation. The objective of the current study is to decipher the metabolome and identification of key regulatory proteins in different developing stages of *P. pinnata* seed and their role in FA accumulation. In addition, the photosynthetic behavior of the *P. pinnata* seed also have been identified which will help in understanding the carbon influx to FA biosynthesis. Our results will deepen the knowledge on metabolic regulation of *P. pinnata* seed development and oil accumulation which is important for further metabolic engineering of this potential biofuel feedstock for different industrial and agricultural applications.

2. Materials and methods

2.1. Plant material and study site description

Ten year old *P. pinnata* plantation, established in the experimental farm of Tree Oils India Limited (TOIL) Zaheerabad, Medak district, Telangana state, India (latitude 17°36′; longitude 77°31′E; 622 m MSL) was selected for the present study. The study site has a tropical, hot-steppe agroclimate with the summer months between March to May having maximum temperature up to ~42 °C and an average temperature of 22–23 °C in the winter months of September to February, monsoon starts during June to October with average rainfall ranging from 700 to 1500 mm. Plants were maintained under natural photoperiod with uniform 2 m spacing between as well as within the rows. The recorded range of light intensity in the region on a normal sunny day ranged between 1200–2000 μmol/m^2/s during 10.00–14.00 h solar time. The plants were regularly watered during hot summer months with alternate watering during winter months and no watering during monsoon season. Nitrogen was provided by applying cow dung mixed with vermicompost as farmyard manure at the rate of 12 kg/year twice in equal splits during growth. Trees had a circular canopy with 15–20 tertiary branches. The flowering and fruiting seasonal events were recorded periodically for further experimental analyses. Flowering initiated in the month of March with visible inflorescence which continued until April. Natural pollination resulted in formation of pods with first visible appearance recorded at 80–90 days after the flowering started. Seed development and maturation was completed by nearly 300 days (~10 months) from the first visible appearance of flowers. The appearance of flower buds in each tertiary branches was considered as 0 days after flowering (DAF). Individual flowers were tagged randomly and the development of pod and seed was monitored regularly in the tagged trees. The tagged pods were collected at 120, 180, 240 and 300 DAF to study the seed morphology and stored at –80 °C for further biochemical and molecular analyses. The pod contained very tiny seeds with negligible dry mass during first initial stages of development (30 to 100 DAF). Therefore the experiments were carried out when the seeds reached 120 DAF till maturation. Seeds inside 120, 180, 240 and 300 DAF old pods were considered as stage1 (S1), stage2 (S2), stage3 (S3) and stage4 (S4) respectively. The study was performed for two successive flowering seasons of *P. pinnata* (for two consecutive years, March 2016 to February 2017 and March 2017 to February 2018).

2.2. Seed morphology and biochemical analysis

Fresh seeds were initially weighed and dried for 12 h at 65–70 °C in a ventilated hot air oven. The weight of seed coat and cotyledon were measured gravimetrically and percentage was calculated using the formula: (weight of cotyledon / weight of total seed) ×100; (weight of seed coat / weight of total seed) ×100, respectively. The seed dry mass was recorded and seed moisture content (%) was calculated according to the formula: (FW–DW/FW) ×100.

Chlorophyll content in the seed samples was estimated according to the method reported by Hiscox and Israelstam (1979). Fresh seeds (100 mg) were ground and refluxed for 1 h in 10 mL of DMSO, extracted liquid was collected and absorbance was measured at 645 and 663 nm using a UV–vis spectrophotometer (Eppendorf, Germany) and chlorophyll contents were calculated according to Arnon (1949) using the formula: chlorophyll *a* (g/L) = 0.0127 × A_{663}– 0.00269 × A_{645}; chlorophyll *b* (g/L) = 0.0229 × A_{645} – 0.00468 × A_{663}; the results were expressed as mg/g FW (where A_{663}, A_{645} represents the absorbance measured at 663 and 645 nm respectively).

Total carbohydrates and starch content in the seeds were estimated according to Anthrone method as described by Hedge and Hofreiter (1962), where 100 mg of fresh seed tissue was refluxed with 5 mL of 2.5 N HCl followed by neutralizing with NaOH. The mixture was centrifuged for 10 min at 6000 × g. An aliquot of supernatant (1 mL) was taken and added with 4 mL of Anthrone reagent. After heating at 80–90 °C for 8 min in a water bath, the solution was allowed to cool rapidly and the absorbance was measured at 630 nm. Spectrometric readings were quantified using glucose standards and represented in mg/g FW of the seed.

For starch estimation, 100 mg of fresh seed tissue was first ground with hot 80% ethanol. The supernatant of this mixture was removed after centrifugation. The residue was washed repeatedly with 80% hot ethanol till the washings did not give green color with Anthrone reagent. The dried residue was mixed with 5 mL of water and 6.5 mL of 52% perchloric acid. The mixture was kept at 0 °C for 20 min and

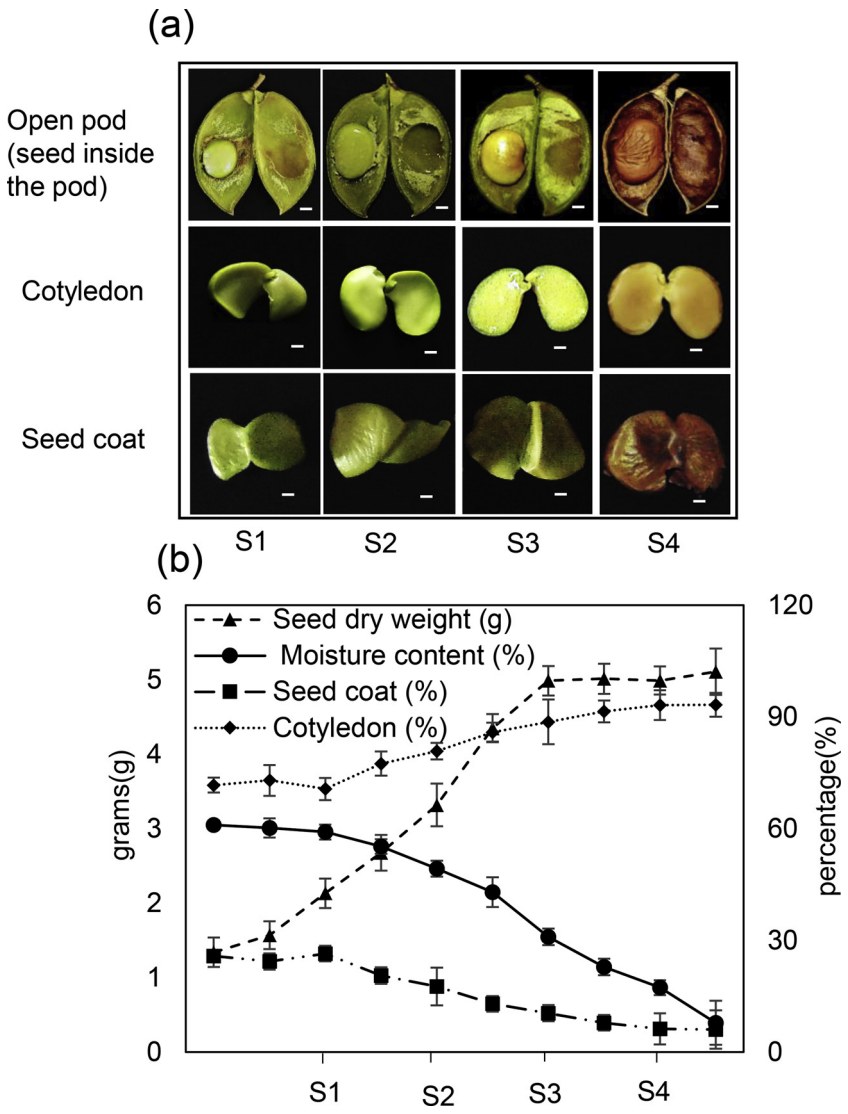


Fig. 1. Position of seed inside the pod and cotyledons separated from seed coat at each developmental stage **(a)**. Morphological and biochemical analysis of developing *P. pinnata* seeds. Changes in seed dry weight, moisture content, % cotyledon and seed coat throughout development **(b)**. Values are mean ± SD, (n = 40); the scale is presented by white bar with the length of 0.3 cm.

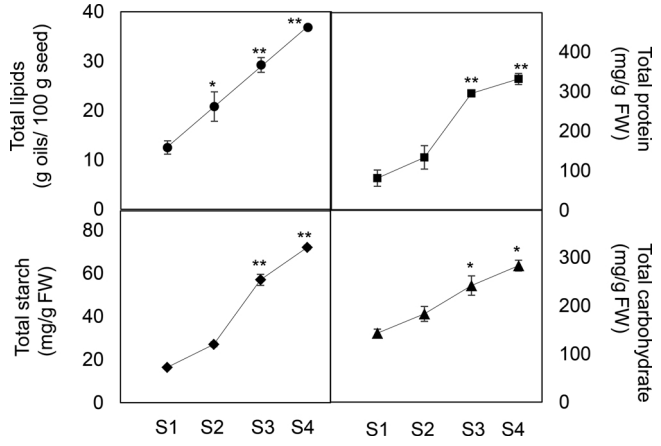


Fig. 2. Pattern of increasing storage products (lipid, protein, total carbohydrate and starch) with development. Values are mean ± SD (n = 3). Data are given as means ± SD; (n = 3). *, ** represents significant differences within the stages analyzed by one-way ANOVA, $P < 0.05$; $P < 0.01$ respectively.

centrifuged at 8000 × g for 5 min, supernatant was collected and the extraction was repeated twice. The collected supernatant (1 mL) was mixed with 4 mL of Anthrone reagent and starch was estimated by the method used in estimation of carbohydrates.

For total protein quantification, fresh tissue (100 mg) was ground with liquid nitrogen and mixed with 2 mL of protein extraction buffer consisting of 25 mM Tris − HCl (pH 7.2), 15 mM $MgCl_2$, 15 mM EGTA, 75 mM NaCl, 2 mM DTT, 1 mM NaF, 1 mM PMSF with 0.1% Nonidet, followed by centrifugation at 12,000 × g for 5 min at 4 °C. Supernatant was collected and protein was estimated according to Bradford (1976). Quantification of total protein was achieved by using BSA as the protein standard and represented as mg/g FW.

Total lipids were extracted by using the method described by Bligh and Dyer (1959) with slight modifications. Fresh seeds were ground with $CHCl_3$:MeOH:H_2O (2:2:1.8) v/v and the mixture was centrifuged at 10,000 × g for 15 min. The lower $CHCl_3$ layer was collected and the extraction was repeated three times. The chloroform was separated by a Rotary evaporator (Heidolph, Germany) and total lipid was estimated gravimetrically and represented as g oils/100 *g* seed. All of the chemical reagents used for biochemical analysis were purchased from Sigma Aldrich, USA.

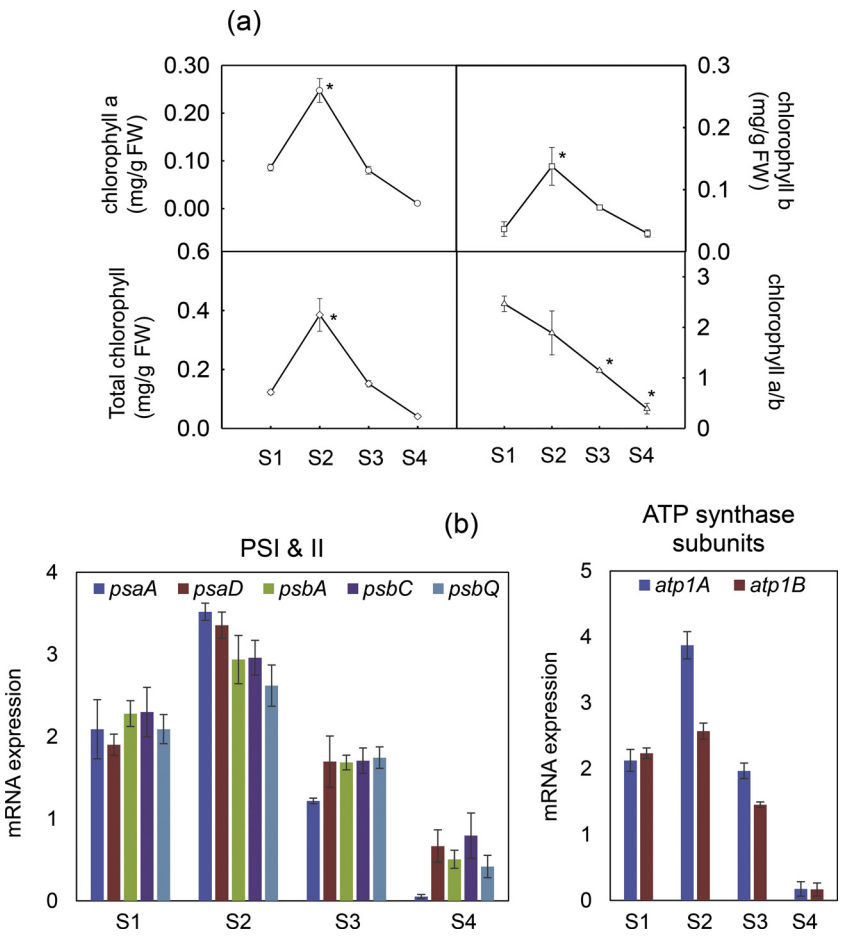


Fig. 3. Changes in the chlorophyll *a*, b, total chlorophyll and chlorophyll *a*/b **(a).** Expression patterns of key regulatory genes represented in fold change related to photosynthesis **(b).** Fold change > 2 is considered as upregulation. Values are mean ± SD (n = 3). Data are given as means ± SD; (n = 3). *, ** represents significant differences within the stages analyzed by one-way ANOVA, P < 0.05; P < 0.01 respectively.

2.3. Gene expression analysis by real time PCR

Total RNA was isolated from developing fresh seeds using Plant total RNA extraction kit by following manufacturer's instructions (Sigma Aldrich, USA). Primers for the required genes were designed from available *P. pinnata* transcriptome sequences (Sreeharsha et al., 2016). First strand cDNA was synthesized with 1 µg of RNA using Revert aid first strand cDNA synthesis kit (Thermo-Fischer Scientific, USA). Expression analysis of selected genes were carried out on Realplex thermal cycler (Eppendorf, Germany) using SYBR FAST qPCR universal master mix (2X) (KAPA Biosystems, USA) with 50 ng of cDNA as template and following the program: 2 min at 95 °C, followed by 40 cycles of 15 s at 95 °C, 30 s at 55 °C annealing temperature and 20 s at 72 °C, followed by the dissociation (melting) curve. The mRNA expression level was calculated according to the $2^{-\Delta\Delta Ct}$ formula (Livak and Schmittgen, 2001). For the internal control, 18 s ribosomal RNA gene was used.

2.4. Metabolite profiling

P. pinnata seeds at different developing stages were ground into fine powder in liquid nitrogen. 100 mg of the seed powder was homogenized with 1.4 mL precooled methanol by vortexing for 10 s and 60 µL of ribitol (0.2 mg/mL) was added to the mixture as internal standard followed by 10 s of vortexing. This solution was ultra-sonicated for 10 min, followed by centrifugation at 11,000 × g for 10 min. The supernatant was transferred to a new tube and mixed with 750 µL of precooled chloroform and 1.5 mL of precooled water. The mixture was then centrifuged at 2200 × g for 15 min. 150 µL of extraction solution from upper phase was dried under vacuum and stored at –80 °C until derivatization. The extract was methoxyaminated, silylated and the dried extract was dissolved in 20 µL of methoxy amine hydrochloride pyridine solution (40 mg/mL) which was incubated at 30 °C with vigorous shaking for 90 min. Then 80 µL of N-methyl-N-(trimethylsilyl) trifluoro-acetamide-solution was added to the sample followed by 30 min incubation at 37 °C with vigorous shaking. The derivatized sample was centrifuged at 20,000 × g for 8 min. The supernatant was then transferred to vials for measurement. Samples were measured with gas chromatography coupled with LECO Pegasus R 4D GC GC–TOF spectrometry (GC-TOF-MS) (Agilent 6890, USA). Each sample was injected under both split less and split 25 times mode for better quantification of candidates with a wide capacity range. Candidates were manually annotated by comparing their retention times (RTs) and mass spectra to those of standards in GMD database (Kopka et al., 2005) with a minimum match factor of 700. The peak areas of the same metabolite with different derived groups were merged and normalized to internal standard of ribitol. The concentration of the metabolites was calculated with respect to the known concentration of ribitol and represented as µg/g FW. The differential analysis at developing stages were analyzed using MetaboAnalyst Version 4.0 [Software] (available from http://www.metaboanalyst.ca/MetaboAnalyst/faces/home.xhtml).

2.5. 2-D protein profiling and identification using MALDI TOF MS/MS analysis

Gel based 2D proteome was carried out according to the method described by Sengupta et al. (2011). Protein extracted earlier (200 µL) was precipitated using 800 µL of 0.1% ammonium acetate in methanol and kept at –20 °C for 8–12 h. The mixture was centrifuged at 10,000 × g for 10 min at 4 °C and the precipitate was washed with fresh methanol for 3 times followed by 2 times wash with acetone. The

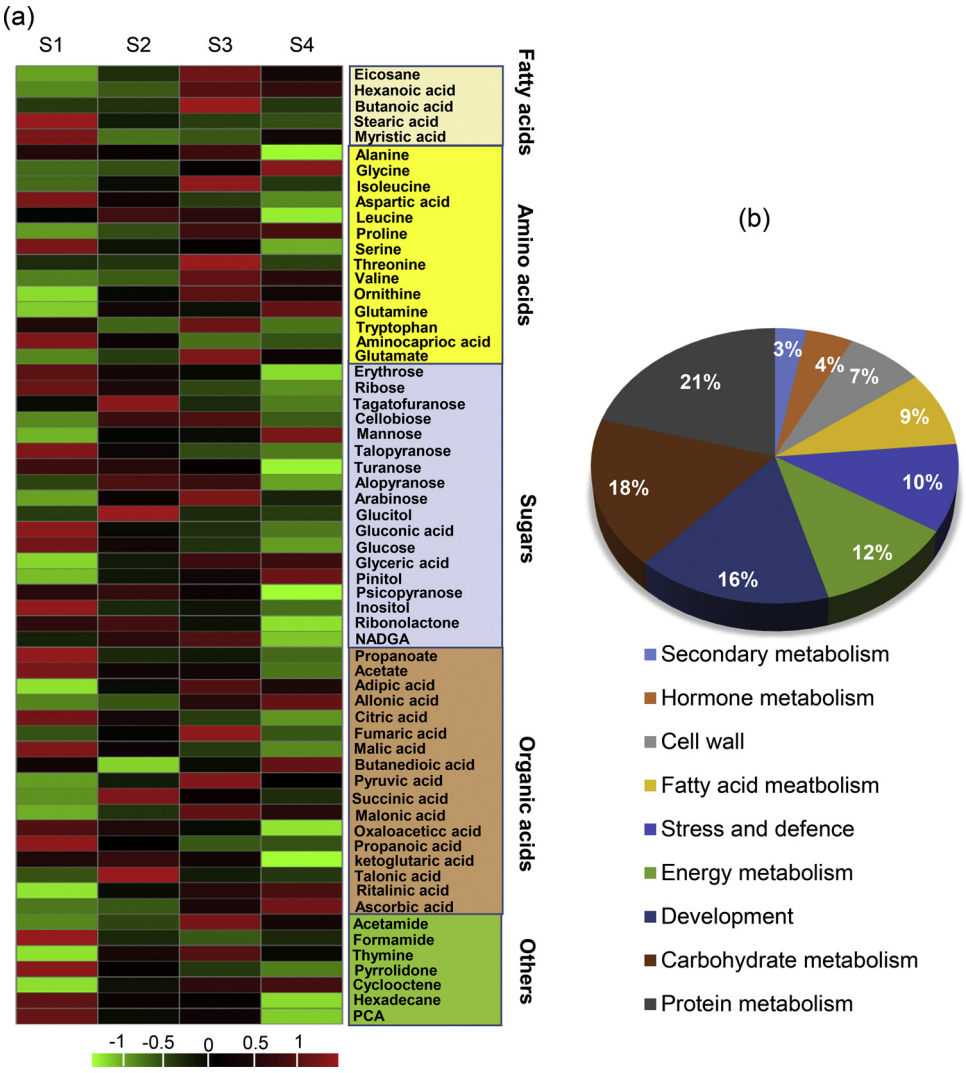


Fig. 4. Heat map based on the fold change in concentration of metabolites (μg/g FW) at S1, S2, S3 and S4 **(a)**. Pie chart representing functional classification of metabolites detected in all the developing stages **(b)**.

isolated protein sample (900 μg) was then solubilized in rehydration buffer followed by isoelectric focusing using pH 4–7 isoelectric focusing strips (18 cm, 4–7 pH linear gradient; GE Healthcare, USA) for 12 h at 50 V for first separation. Rehydration and focusing was carried out in Ettan IPGphor II (GE Healthcare, USA) at 20 °C, using the following program: 30 min at 500 V, 3 h to increase from 500 to 10,000 V and 6 h at 10,000 V (a total of 60,000 Vh). The IEF strips were then placed over 12% acrylamide gel for the second separation through SDS PAGE at 300–500 V using standard protocol. After staining, the gels were scanned and the gel image obtained was analyzed though Image Master 2-D Platinum version 6 image analysis software (GE Healthcare, USA).

For the identification of proteins, MALDI TOF MS/MS was carried out through in-gel trypsin digestion and database searches (PMF and MS/MS) using MASCOT program [Software] (available from http://www.matrixscience.com). The similarity search for mass values was done with existing digests and sequence information from NCBInr and Swiss Prot database. The taxonomic category was set to Viridiplantae (green plants). The other search parameters were: fixed modification of carbamidomethyl (C), variable modification of oxidation (M), enzyme trypsin, peptide charge of 1+ and monoisotopic. Out of top ten most significantly identified proteins, results having highest score, peptide match, and similarity of molecular weights were considered.

2.6. Microscopic studies of P. Pinnata seeds

Fresh seeds were made into thin sections and fixed in 2.5% glutaraldehyde for 8–12 h followed by re-fixing again in 0.2% osmium tetra oxide for 2–3 h. After washing with 0.1% PBS and series of ethanol dehydration, thin sections were embedded in epoxy resin (Araldite 502). Ultrathin sections were made using ultra microtome (LEICA EM UC6, Germany) and observed under Scanning electron microscope (TESCAN S8000, Czech Republic). For TEM, ultrathin sections were fixed in grids and stained with uranyl acetate and lead citrate which was observed later in TEM (FEI Model, Tecnai G2S Twin, Spain) (200 kV). Nile red staining was performed by following the method reported by Greenspan et al. (1985) with slight modifications. Prefixed thin sections of the fresh tissues were infiltrated with the working concentration of 2 μg/mL Nile red solution in 0.1% HEPES buffer for 1–2 h under dark. After washing for 4–5 times with HEPES buffer, the sections were visualized under confocal microscope (LIECA TCS SP2 AOBS, Heidelberg, Germany) under excitation of 488 nm and emission range of 530–650 nm. All images were obtained with 10% laser power.

2.7. Thin layer chromatography for lipid profiles

Total lipid (5 mg) extracted from all the samples were dissolved in chloroform and spotted on the TLC Silica gel 60 F254 (Merck,

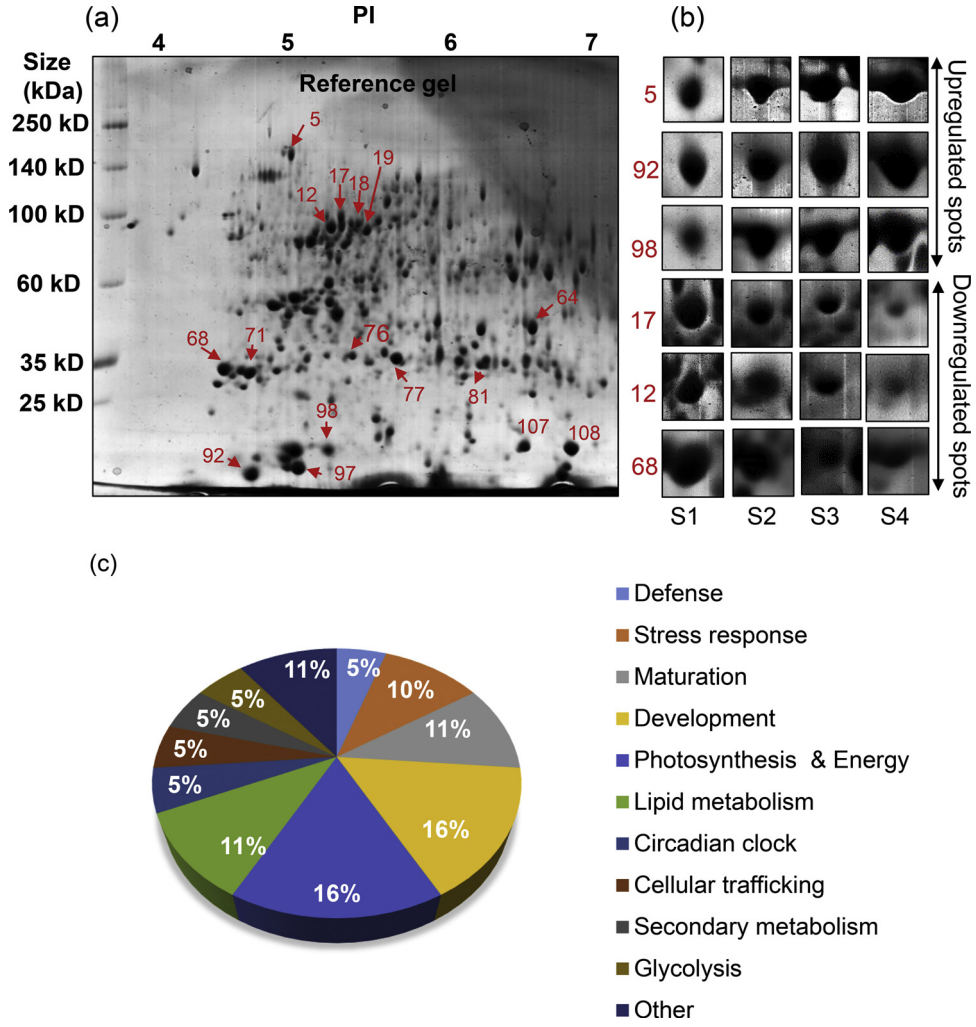


Fig. 5. 2D reference gel image where the identified spots were indicated by spot ids in red mark **(a)**, some of the upregulated and downregulated spots were presented as separated images at each stage **(b)**. Functional classification of identified proteins in pie chart **(c).**

Germany). The separation was carried out with hexane and ethyl acetate (9:1 v/v) as the mobile phase and silica gel bonded with aluminum sheet as the stationary phase. The TLC plates were air dried and exposed to iodine vapor. Iodine stained spots were compared with the standards of phospholipids, DAGs and TAGs separated in TLC plates with the same mobile phase.

The TAGs from the multiple TLC plates were scrapped carefully and extracted using chloroform. Further, chloroform was separated from the extracted TAGs with the help of rotary evaporator (Heidolph, Germany). The fatty acid profiling and quantification was carried by following the methods reported by Coetzee et al. (2008) and Sun et al. (2017) with slight modifications. FAMEs were prepared by refluxing 50 mg of the TAGs extracted with 5% H_2SO_4 in methanol (w/v) for 6 h on a hot plate. Prepared esters were analyzed through gas chromatography (GC-TOF-MS) (Agilent 6890, USA) with DB225 column (inner diameter = 0.25 mm, length = 37 m, thickness = 0.25 µm; Agilent, USA). The injector and flame ionization detectors were set at 250 °C and 270 °C, respectively. The oven temperature was set at 160 °C for 2 min and then increased to 230 °C at a rate of 5 °C/min. Nitrogen was used as the carrier gas at a flow rate of 1 mL/min. Fatty acids were confirmed by comparing with standard FAMEs mix (C14 – C22, Supelco, Sigma Aldrich, USA) analyzed at different concentrations. Quantification of fatty acids was carried out using area normalization method with percentage area of each peak corresponding to the identified fatty acid.

2.8. Statistical analysis

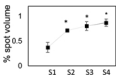
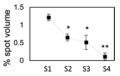
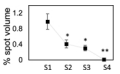
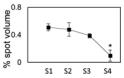
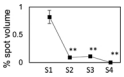
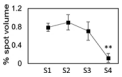
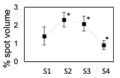
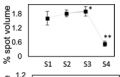
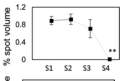
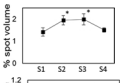
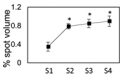
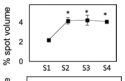
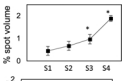
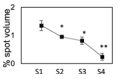
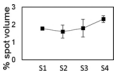
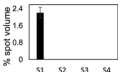
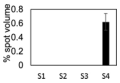
Each seed collected at various stages was considered as one unit (n = 40), where n is the number of seeds for the verification of difference in its morphological and physiological measurements. All the experiments were carried out in triplicates for the developing stages. PCA and heat map analysis was carried out for metabolome data using MetaboAnalyst online tool version 4.0 [Software] (available from http://www.metaboanalyst.ca/MetaboAnalyst/faces/home.xhtml). The mean values were compared and analysis of significance ($P < 0.05$) was determined by student's *t*-test and one way ANOVA using Sigma plot 11.0.

3. Results

3.1. Seed morphological characteristics and storage biochemistry during development

A single pod of *P. pinnata* generally houses either one or two seeds with 3–4 seeds occasionally/rarely observed. All the four different stages of *P. pinnata* seed with developmental progressing of pod, cotyledon and seed coat are shown in Fig. 1. The young seeds are green in color initially, which turned completely brown upon maturation (Fig. 1a). There was a gradual decrease in the seed moisture content from S1 to S4 whereas, the seed dry weight increased till S3 and got stabilized by maturation (Fig. 1b). Seed coat percentage decreased

Table 1
The list of protein spots identified with their % spot volume and MASCOT search results at four stages of development. Values are mean ± S.D, (n = 3), level of significant difference was analyzed by one-way-ANOVA with reference to S1. Note - *, P < 0.05; **, P < 0.01.

Sp. ID	Spot intensities	Protein identified	MW observed in gel	MW from literature	Known Function	Organism related to	Accession no.	MS/MS Score	Peptide sequences matched
5		Acetyl coA carboxylase	200kD	226 kD	Fatty acid biosynthesis	*Sesamum indicum*	XP011083400.1	98	DEGRGPMR EDAFFQAVTEVACAQK ASQLLEQTK(Total matched peptides-29)
12		Midasin	95kD	116kD	Female gamete and seed development	*Arabidopsis thaliana*	MDN1_ARANTH	94	WMYLESIFVGSDDIRHQLPAEAK KSFEMVSLAVSQK (Total matched peptides - 34)
17		AP2 Complex subunit alpha 1	91kD	112kD	Vesicular transport and development	*Arabidopsis thaliana*	AP2A1_ARATH	94	FAPDLSWYVDVILQLIDK LVLFMGWK (Total matched peptides - 12)
18		Phosphoenol pyruvate carboxykinase	95kD	73.7kD	Photosynthesis carbon fixation and energy metabolism	*Zea maize*	PCKA_MAIZE	64	GLFGVMHYLMPK AQTIDELHSL QR(Total matched peptides - 12)
19		Probable disease resistance protein	100kD	104.5kD	Defence and disease resistance	*Arabidopsis thaliana*	PX24L_ARATH	108	VLGGLLAAKYTLHDWKR ERKDEIQNMK(Total matched peptides - 31)
64		Glyceraldehyde-3-phosphate dehydrogenase	45kD	36.5kD	Glycolysis and energy metabolism	*Oryza sativa*	G3PC1_ORYSJ	278	AASFNIIPSSTGAAK LKGIIGYVEEDLVSTOFVGD SR(Total matched peptides - 7)
68		ATP synthase subunit beta	35kD	53.3kD	Photosynthesis electron transport and energy	*Brimeura amethystina*	ATPB_BRIAM	68	AVAMSATDGLTR EGNDLYMEKESGVINEK(Total matched peptides - 8)
71		Ent-copalyl diphosphate synthase 1	35kD	99kD	Secondary metabolism	*Oryza sativa*	CPS1_ORYSJ	88	ARNFSYEFLR EIEQNMDYVNR(Total matched peptides - 8)
76		1-cys peroxiredoxin B	35kD	24.4kD	Vitamins	*Oryza sativa*	REHYB_ORYSI	67	DTAGGELPNR VVIPPGVSDEEAK(Total matched peptides - 8)
77		GTP Binding protein SARB1	30 kD	22.02kD	Vesicular transport of storage bodies	*Arabidopsis thaliana*	SAR1B_ARATH	74	VWKDYYAK YHLGLTNFTTGKG (Total matched peptides - 9)
81		Heat stress transcription factor c-1b	30kD	27.4 kD	Abiotic stress regulator	*Oryza sativa*	HFC1B_ORYSJ	97	NFASFVR QLNTYGFR(Total matched peptides - 17)
92		Lipid transfer like protein VAS	10kD	17kD	Lipid transfer metabolism	*Arabidopsis thaliana*	VAS_ARATH	90	WSSQAER EVPQVCCNPLK(Total matched peptides - 8)
98		Integrin linked protein kinase 1	25kD	52.4kD	Signal transduction (Plant defence response)	*Cucurbita maxima*	XP_022970695.1	101	TNPGSRSFSK WGSTPLADAIYYK(Total matched peptides - 6)
107		Class I heat shock protein	10.5kD	17.7kD	Heat shock and stress	*Solanum pennelli*	XP_015078018.1	88	FRLPENAK VEVEEDRVLQISGER(Total matched peptides - 7)
108		Class I heat shock protein	10.3kD	17.8kD	Heat shock and stress	*Solanum lycopersicum*	HSP11_SOLLC	66	FRLPENAK VEVEEDRVLQISGER(Total matched peptides - 6)
97		Protein ELF 4 Like	15kD	14.07kD	Circadian clock	*Arabidopsis thaliana*	EF4LT_ARATH	58	DGDTTTTTTGSS NVGLINEINISQVMEIYSDLSLNFAK (Total matched peptides - 2)
N17		Allergen LEN	90kD	47.7kD	Allergen protein in non-edible plants	*Lens culineris*	CAD87730.1	112	IFENLQNYR KSVSSESESEPFNLR(Total matched peptides - 8)

gradually while the percentage of cotyledon increased as the seed matured (Fig. 1b). Accumulation of all the storage products including starch, protein and lipids increased gradually, reaching the highest levels during S4 (Fig. 2). Protein and lipid contents were similar at younger stages (S1–S2), while the lipid content reached up to 37% of total 100 g of seeds (~36.8 g oil/100 g seed), whereas protein content reached to 33% (~331 mg/g FW) by the end of the development. Similarly, the total carbohydrate content in the mature seeds (S4) was ~298 mg/g FW (1.9 folds higher when compared to S1), out of which the content of starch was only ~80 mg/g FW (4.3 folds higher when compared to S1) (Fig. 2). Chlorophyll *a*, chlorophyll b and total chlorophyll concentrations showed a distinct pattern when the seeds were green where their contents increased from S1 to S2 with a further decrease thereon at S3 and S4, during which the seeds started to develop brown color (Fig. 3a). Similarly, chlorophyll *a*/b ratio also consistently decreased with the progression of seed development (Fig. 3a). Relative gene expression of some of the key photosynthetic genes were analyzed at different stages of development by considering S1 as reference (Fig. 3b). All genes related to Photosystem I and II major subunits (*psaA, psaD psbA, psbC, psbQ*) and ATP synthase subunit (*atp1A* and *atp1B*) were upregulated at S2 which subsequently decreased from S3 to S4 (Fig. 3b).

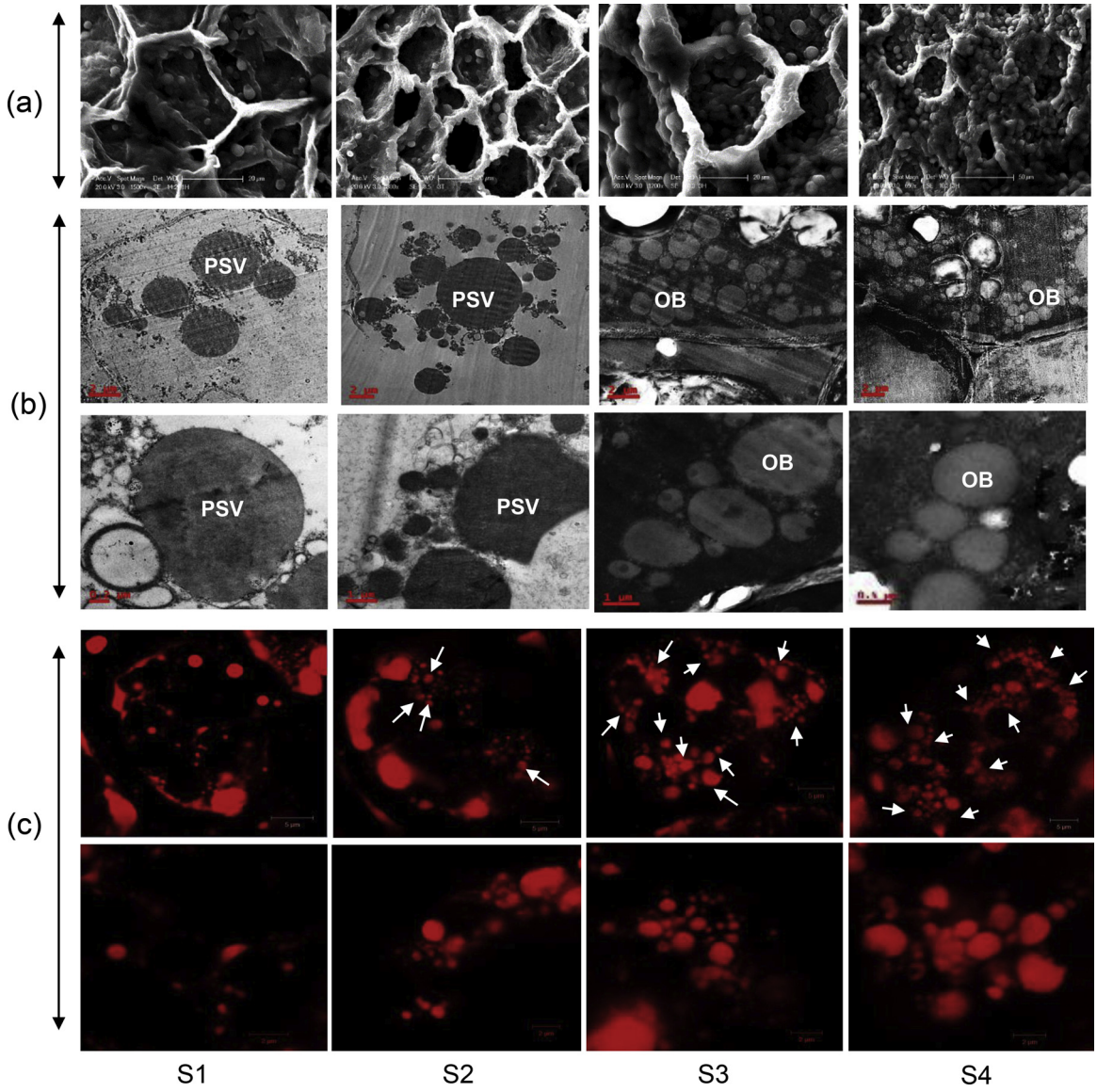


Fig. 6. *P. pinnata* seed cross section under SEM **(a)** (20–50 μ) and TEM **(b)** (2–0.2 μ). Nile red staining of *P. pinnata* seed sections observed under confocal microscope, oil bodies are marked by white arrows **(c)** (5–2 μ). PSV- protein storage vacuole; OB- oil body.

3.2. Complete metabolome analysis of developmental stages of P. Pinnata seed

In all the four stages, 61 metabolites were identified with the help of GCMS analysis (Supplementary data 1). Metabolites were classified based on their chemical structure (Fig. 4a) and biological functions (Fig. 4b). The variation among distinctive profile of metabolites was observed through PCA analysis where there was overall 71.7% separation among the developing stages (Supplementary data 3). Separation or variation of seed metabolites among S2 and S3 was less when compared to S1 and S4. Majority of the metabolites detected were functionally related to protein metabolism (21%), carbohydrate metabolism (18%) and development (16%). Rest of the metabolites belonged to energy (12%), stress and defense (10%), fatty acids (9%), cell wall formation (7%), hormonal responses (4%) and secondary metabolism (3%) (Fig. 4b). The dynamics of the metabolites were determined by their concentration during the 4 developing stages analyzed through heat map using Pearson's test in MetaboAnalyst online software. Metabolites were classified as fatty acids, amino acids, sugars and organic acids depending on their respective chemical characteristics (Fig. 4a). Among the sugars, ribose, glucose, erythrose, gluconic acid, inositol and ribonolactone had higher concentrations at S1 which gradually decreased with the maturation. Glyceric acid, pinitol and mannose were abundant at S4 whereas sugars like cellobiose, arabinose and NADGA were higher during S1 to S2. In contrast to the sugars, most of the amino acids and organic acids were abundant during S3 and S4 (Fig. 4a). Leucine, threonine, alanine, tryptophan, valine, glutamate and some organic acids including keto-glutaric acid, fumaric acid, succinic acid, pyruvic acid and malonic acid were abundant in S3 and S4. The concentrations of glycine, proline and glutamine were higher in S4. Similarly, adipic acid and allonic acids were higher during S3 and S4 respectively (Fig. 4a). In contrast, few organic acids including propionic acid, acetate, malic acid, citric acid, oxalic acid and certain amino acid such as aspartic acid were more at S1, which gradually decreased to the lowest levels at S4. Fatty acids which were abundant in S1 were stearic acid and myristic acid, hexanoic acid was abundant in S3, while others including butanoic acid, and eicosane were more in S4 (Fig. 4a).

3.3. 2D- proteomic profile of developing P. Pinnata seed

The seed protein expression patterns were quantified at different

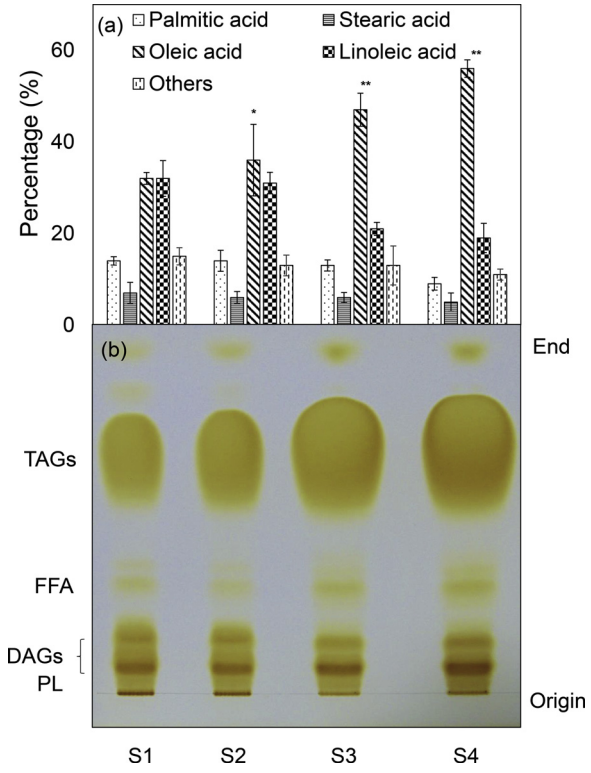


Fig. 7. Percentage of fatty acids accumulated in the TAGs at different developing stages **(a)**. Total lipid profiling through TLC **(b)**. Values are mean ± SD (n = 3). *, ** represents significance difference (P < 0.05, P < 0.01) between the fatty acid content with respect to 120 DAA, analyzed by *t*-test.

developing stages of *P. pinnata* seed. More than 300 spots were reproducibly detected in all the gels with Image Master 2D Platinum software and 125 spots matched in all the four stages (Fig. 5a). Based on the percent spot volume in the S1 stage, we have determined the significantly upregulated, unchanged or downregulated spots throughout developing stages. Some of the spots were selected and successfully identified using MALDI MSMS (Table 1). Rest of the spots which did not give any significant match in the MASCOT search were not considered. All of the differentially regulated spots were represented as gel picture in the supplementary data 2, and the changes in some of the spots among the different stages were represented along with the reference gel in Fig. 5a,b. The spot distribution and intensity at each stage were analyzed through PCA analysis which illustrated the variation in the form of percentage separation (Supplementary data 4). There was overall 64.4% separation among the stages and the graph clearly showed the protein profile of S4 was different compared to the younger stages (120–240 DAF). The profiles of S2 and S3 were more similar with least separation (Supplementary data 4). The functions of the identified spots were classified based on the available literature (Fig. 5c). The upregulated spots from S1 to S4 were identified as Accase, lipid transfer like protein VAS, ATP synthase subunit β and GTP binding protein SAR1b. The unchanged spots were identified as PEP carboxykinase and G3PD. Two of the downregulated spots included Midasin and ap2 complex subunit alpha1 (Table 1). The protein spot identified as ELF 4 was observed only in S1 whereas, Allergen Len was observed only in S4 (Table 1).

3.4. Visualization of storage granules and oil bodies in developing P. Pinnata seeds

The surface morphology of *P. pinnata* cotyledonary cross sections analyzed by SEM at various stages showed an increase in storage granules from S1 to S4 (Fig. 6a). Further the biochemical nature of these storage granules were studied through TEM analysis and identified large protein vacuoles at S1 which further increased during S2. The appearance of a large amount of oil body accumulation was observed in S3 and S4, where the lipid droplets were located at cellular periphery (Fig. 6b). For more clarification of the accumulation of oil bodies in developing stages, Nile red fluorescent staining was performed which specifically stains only the lipid containing substances. Oil body accumulation was very low at S1 and S2, but showed rapid increase at S3 to S4 (Fig. 6b, c). Moreover, the oil bodies were localized near the periphery of the cell with sizes varying from 1 to 2 μ while the protein storage vacuoles have unusually large diameter (~2–5 μ) and situated mostly in the central region (Fig. 6).

3.5. Fatty acid profiling of the TAGs

The separation of phospholipids, DAGs and TAGs during developing stages were compared by the intensities on the TLC plates which have been loaded with equal quantity and separated with the same mobile phase. There was a visible change in the TAGs accumulation where it progressively increased as the seeds reached maturity (Fig. 7b). The fatty acid profile of the accumulated TAGs was examined with the help of GCMS analysis and compared with the FAMEs standards, based on which the percentage was also determined. The major fatty acids during initial stages of development (S1 and S2) were oleic and linoleic acid which were present in equal amounts (Fig. 7a). In later stages of development, oleic acid gradually increased with a significant ($P < 0.05$) decrease in linoleic acid (Fig. 7a). Furthermore, palmitic acid also decreased gradually from S1 to S4 (Fig. 7a) while there was no change in stearic acid content in all the stages.

4. Discussion

The basic seed morphology and biochemistry of developing *P. pinnata* seed share certain similar developmental regulatory processes with other legumes where, the pods were nearly flat in younger stages with a tiny developing seed inside and the pod thickness increased as the seed developed, finally filling the pod (Wright and Lenssen, 2013). The duration of seed development in *P. pinnata* is a prolonged process making it complex to understand the seed filling and oil accumulation patterns in this imminent biofuel tree species. Thus, knowledge on physiological and molecular dynamics of this lengthy seed development will boost the attempts to improve the seed productivity towards potential biofuel feedstock. The current study thoroughly examined the seed developmental processes of *P. pinnata* and deduced the physiological, biochemical and molecular changes through metabolomics and proteomic approaches. Our present studies also provide insights into the photo-autotrophic nature of *P. pinnata* seed.

Legume seed development and differentiation is characterized by three rapid growth phases separated by two lag phases (Weber et al., 2005). Majorly, the growth phases are marked by the development of endosperm, seed coat and embryo associated with maturation and cell expansion. The differentiation of processes during seed development is a sequential process which involves active mitotic cell division, sucrose uptake, cell expansion, greening and gaining of photosynthetic activity as well as accumulation of storage products (Baud et al., 2002). For oil seeds, the differentiation phase is of special interest because during this stage a regulatory network initiates the accumulation of storage products. The developing seeds of *P. pinnata* are green in color at younger stages (S1-S2) which turned brown during maturation with a gradual increase in seed weight, storage products and lipid content but decrease in moisture content as well as the weight of seed coat. The importance of maternal seed coat in providing the necessary sources of filial cotyledons for development and controlling germination is well known in legumes (Weber et al., 2005). Seed coat percentage in terms of its structural weight is high during younger stages of development which indicated its active role in maintaining seed development. During

natural seed maturation process, there is a gradual increase in seed desiccation, wherein the water content declined significantly, (Angelovici et al., 2010). In general, chlorophyll content and the ratio of chlorophyll *a*/b ratio play a significant role in developing green seeds and are directly related to the photosynthetic efficiency of the green seeds (Eastmond and Kolacna, 1996). In the current study, initial stages of *P. pinnata* seed showed higher chlorophyll *a*/b ratio (> 1), which declined gradually with maturation. This clearly indicated the active photosynthesis in *P. pinnata* developing seeds during S1 and S2. To support our results, we have also analyzed the gene expression of some of the key photosynthetic electron transport genes. PS I and PS II related gene expression was reported in many green and oil yielding seeds and it is also considered to be an important factor for embryo development and fatty acid biosynthesis (Niu et al., 2009; Allorent et al., 2015). The expression of photosynthetic genes in *P. pinnata* was higher till S2 after which there was a steady decline in expression levels indicating very active photosynthetic process during the initial stages of seed development. The temporal induction of ATP synthase subunit gene (atp1A) could have also contributed for substantial production of ATP which is needed for the energy productions and fatty acid biosynthesis. The seed photosynthesis may provide energy for the fatty acid biosynthesis during subsequent stages and also helps in re-fixing the respiratory CO_2 release. The release of O_2 through the photosynthesis can aid in reducing NO:O_2 ratio, thereby preventing anoxia in seeds during development (Borisjuk and Rolletschek, 2009). Active photosynthesis might also contribute for significant accumulation of storage products including protein, starch and other carbohydrates which was evident in this study with *P. pinnata* seeds.

Metabolites are considered as very robust and sensitive as they maintain the seed physiology and storage biochemistry (Weselake et al., 2009). Understanding the dynamics of metabolic profile among various stages of development might give new insights into the mechanism of seed oil biosynthesis with respect to quantity and quality of lipids in this potential biofuel crop. The stage dependent metabolite changes of various crops including *Oryza sativa*, *Glycine max* and *Brassica napus* were recently reported which had enlightened the role of metabolites in regulating oil biosynthesis, cellular morphology, seed filling and maturity (Tan et al., 2015; Hu et al., 2016; Gupta et al., 2017). We have detected sugars, organic acids, free fatty acids and amino acids through our metabolite analysis in developing *P. pinnata* seed. Sugars and organic acids take part in energy metabolism thus providing reducing compounds as well as precursors for fatty acid biosynthesis, cell division and differentiation. Among the detected sugars, glucose is utilized as a signal for cellular multiplication and differentiation while other sugars like ribose and erythrose take part in nucleotide formation and energy metabolism such as Calvin cycle. Interestingly, sugars related to cell wall formation and cell division were higher in S2-S3, for example NADGA, cellobiose, pyranose and arabinose indicating that the active stages of cellular division and differentiation are from S1 to S3. Sugars such as mannose were most abundant in the mature stage S4 and it may contribute to the production of bioethanol from remaining seed residue after the extraction of oil. The major organic acid components of TCA cycle were also detected in the metabolites where, the stage wise quantitative changes were observed in *P. pinnata* seeds. The presence of a flux mode for respiratory energy synthesis through TCA may be the reason of expression of these metabolites in higher concentration during initial stages (S1-S3), which in turn can aid the fatty acid biosynthesis. Amino acids in the developing seeds contribute in various metabolic activities including storage protein accumulation, germination, stress responses and in respiratory catabolic processes to maintain optimum energy status inside the developing seeds (Miranda et al., 2001; Galili et al., 2014). Most of the amino acids were abundant in S2-S3, while glycine and proline which are having stress responsive characteristics were abundant during S4 providing clues for stress tolerance nature of *P. pinnata* seed as the development progressed (Hayat et al., 2012; Czolpinska and Rurek, 2018). The other important amino acids such as glutamate, aspartate, glutamine were also detected which play a central role in the synthesis of other amino acids during seed development (Brian et al., 2007; Jander and Joshi, 2009).

Total seed storage protein profile analyzed through 2D gel electrophoresis gave the overall picture of stage wise variation in protein expression and regulation, where some key regulatory proteins have been successfully identified. The proteins, related specifically to seed development are Midasin and ap2 complex subunit alpha which take part in the development of female gametophyte as well as the cotyledonary development (Chantha et al., 2010). On the other hand, glyceraldehyde-3-phosphate dehydrogenase, a glycolytic enzyme showed higher expression during younger stages which decreased significantly at S4 indicating an active cellular mechanism such as glycolysis at younger stages S1-S2 which gradually decreased with maturation. ATPase subunit beta of photosynthetic electron transport and PEP carboxykinase of carbon assimilation metabolism were upregulated at S2-S3. Our data infers that photosynthesis in *P. pinnata* seed plays a critical role in carbon recycling and energy metabolism which is crucial for fatty acid biosynthesis. Significantly upregulated spots from S1 to S4 included acetyl CoA carboxylase (Accase), the first catalyzing enzyme of fatty acid biosynthesis and lipid transfer like proteins (LTPs). The LTPs are low molecular weight (9–10 kD) proteins and have a significant role in plant development such as transporting lipids for membrane biosynthesis, defense responses and vesicular transports (Liu et al., 2015). A subcellular localization study in seeds also showed the involvement of LTPs with protein as well as lipid bodies (De O. Carvalho et al., 2004). In addition, GTP binding protein SAR1b, which is well known for taking part in vesicular protein traffic from Endoplasmic reticulum to Golgi leading to storage body formation was detected in *P. pinnata* seeds (Memon, 2004). Moreover, recently in liver cells, phosphorylation of SAR1b had been proven to be an important step for release of lipid chylomicrons (Siddiqi and Mansbach., 2012). However, in plants particularly the high oil yielding species like *P. pinnata*, it still remains a question to whether this protein carries a role in the synthesis of lipid bodies.

The accelerating accumulation of oil bodies in developing seeds could help in determining the best harvest time of *P. pinnata* seeds. However, the size and pattern of oil body accumulation during seed development is important for understanding the dynamics of molecular regulation behind this process. It was evident from our data that TAG levels were increased gradually as the seed development progressed and the fate of most of the TAGs accumulated is to ultimately form oil bodies (Hills, 2004). The increasing number of oil bodies with maturation suggest that majority of storage product observed were lipids in *P. pinnata* seed. The FAMEs composition of biofuel and its blend can determine its efficiency for fuel capacity (Sbihi et al., 2018; Xiong et al., 2018). In *P. pinnata*, the fatty acid composition of accumulated TAGs at each stage had shown that oleic acid was predominantly incorporated only after S3 making the seed oil of *P. pinnata* rich with unsaturated fatty acids and making it a valuable source of biodiesel which meets all the standard criteria for biodiesel production (Karmee and Chadha, 2005).

5. Conclusion

Present study highlights the active role of metabolites and some key proteins in the stage specific regulation of *P. pinnata* seed development. The pattern of gene expressions along with metabolites and the proteome data can contribute to possible regulatory networks of cellular metabolism responsible for synthesis and accumulation of oil in Pongamia seeds during different developmental stages. Major metabolites and proteins of each stage belonging to various pathways are identified and their profiles reveal that energy metabolism and cell division are major processes in S1 and S2 which is followed by rapid increase in storage products along with oil bodies from S3 to S4. The TAGs are the major form of lipids in Pongamia oil and we confirmed

that in mature seeds they contribute to the highest content of oleic acid, which also concur with high content of unsaturated fatty acid in *P. pinnata* making it desirable for biofuel production. These results will possibly lay the foundation for metabolic engineering of oilseed crops for improving the quality and quantity of seed oil as a potential feedstock in the biofuel industry.

Declaration of Competing Interest

Authors declare to have no conflict of interest.

Acknowledgements

The present study was supported by a grant from Department of Biotechnology (DBT), Govt. of India (BT/PR12024/BCE/8/1097/2014). Authors like to thank Tree Oils India Limited for providing access to the farm for sample collection and experimental setup. KTS thanks DBT, New Delhi and RVS, SM thanks UGC, New Delhi for fellowship.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.indcrop.2019.111621.

References

Allorent, G., Osorio, S., et al., 2015. Adjustments of embryonic photosynthetic activity modulate seed fitness in Arabidopsis thaliana. New Phytol. 205, 707–719.
Angelovici, R., Galili, G., et al., 2010. Seed desiccation: a bridge between maturation and germination. Trends Plant Sci. 15, 211–218.
Arnon, D., 1949. Copper enzymes isolated chloroplasts, polyphenoloxidase in Beta vulgaris. Plant Physiol. 24, 1–15.
Atabani, A.E., Silitonga, A.S., et al., 2013. Non-edible vegetable oils: a critical evaluation of oil extraction, fatty acid compositions, biodiesel production, characteristics, engine performance and emissions. Renew Sust Energ Rev. 18, 211–245.
Bala, M., Nag, T.N., et al., 2011. Proximate composition and fatty acid profile of *Pongamia pinnata*, a potential biodiesel crop. J. Am. Oil Chem. Soc. 88, 559–562.
Baud, S., Boutin, J., et al., 2002. An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. Plant Physiol. Biochem. 40 (2), 151–160.
Baud, S., Dubreucq, B., et al., 2008. Storage reserve accumulation in *Arabidopsis*: metabolic and developmental control of seed filling. Arabidopsis. Book. 6, e0113.
Bligh, E.G., Dyer, W.J., 1959. A rapid method of total lipid extraction and purification. Can. J. Biochem. Physiol. 37, 911–917.
Borisjuk, L., Rolletschek, H., 2009. The oxygen status of the developing seed. New Phytol. 182, 17–30.
Borisjuk, L., Rolletschek, H., et al., 2003. Differentiation of legume cotyledons as related to metabolic gradients and assimilate transport into seeds. J. Exp. Bot. 54 (382), 503–512.
Bradford, M.M., 1976. A Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.
Chaitanya, B.S.K., Kumar, S., et al., 2015. Pivotal role of sugar fluxes between the inner integument and endosperm in lipid synthesis during seed ontogeny in *Jatropha curcas* L. Ind. Crops Prod. 76, 1106–1113.
Chantha, S.C., Gray-Mitsumune, M., et al., 2010. The MIDASIN and NOTCHLESS genes are essential for female gametophyte development in Arabidopsis thaliana. Physiol. Mol. Biol. Plants 16 (1), 3–18.
Coetzee, R., Labuschagne, M.T., et al., 2008. Fatty acid and oil variation in seed from kenaf (Hibiscus cannabinus L.). Ind. Crops Prod. 27 (1), 104–109.
Czolpinska, M., Rurek, M., 2018. Plant glycine-rich proteins in stress response: an emerging, still prospective story. Front. Plant Sci. 9, 302.
De O. Carvalho, A., De S. Teodoro, C.E., et al., 2004. Intracellular localization of a lipid transfer protein in *Vigna unguiculata* seeds. Physiol. Plant. 122, 328–336.
Eastmond, P.J., Kolacna, L., 1996. Photosynthesis by developing embryos of oilseed rape Brassica napus. J. Exp. Bot. 47, 1763–1769.
Eastmond, P.J., Rawsthorne, S., 2000. Coordinate changes in carbon partitioning and plastidial metabolism during the development of oilseed rape embryos. Plant Physiol. 122 (3), 767–774.
Ekman, A., Hayden, D.M., et al., 2008. Carbon partitioning between oil and carbohydrates in developing oat (*Avena sativa* L.) seeds. J. Exp. Bot. 59 (15), 4247–4257.
Galili, G., Avin-Wittenberg, T., et al., 2014. The role of photosynthesis and amino acid metabolism in the energy status during seed development. Front. Plant Sci. 5, 447.
Greenspan, P., Mayer, E.P., et al., 1985. Nile red: a selective fluorescent stain for intracellular lipid droplets. J. Cell Biol. 100 (3), 965–973.
Gupta, M., Bhaskar, P.B., et al., 2017. Integration of omics approaches to understand oil/protein content during seed development in oilseed crops. Plant Cell Rep. 36 (5), 637–652.
Hayat, S., Hayat, Q., et al., 2012. Role of proline under changing environments: a review. Plant Signal. Behav. 7 (11), 1456–1466.
Hedge, J.E., Hofreiter, B.T., 1962. In: Whistler, R.L., Be Miller, J.N. (Eds.), Methods in Carbohydrate Chemistry 17. Academic Press, New York.
Hills, M.J., 2004. Control of storage-product synthesis in seeds. Curr. Opin. Plant Biol. 7, 302–308.
Hiscox, J.D., Israelstam, G.F., 1979. A method for the extraction of chlorophyll from leaf tissue without maceration. Can. J. Bot. 57, 1332–1334.
Hu, C., Tohge, T., et al., 2016. Identification of conserved and diverse metabolic shifts during rice grain development. Sci. Rep. 6, 20942. https://doi.org/10.1038/srep20942.
Jander, G., Joshi, V., 2009. Aspartate-derived amino acid biosynthesis in Arabidopsis thaliana. Arabidopsis Book 7, e0121.
Karmee, S.K., Chadha, A., 2005. Preparation of biodiesel from crude oil of Pongamia pinnata. Bioresour. Technol. 13, 1425–1429.
Kesari, V., Rangan, L., 2011. Coordinated changes in storage proteins during development and germination of elite seeds of *Pongamia pinnata* a, a versatile biodiesel legume. AoB Plants 2011 (2011), plr026.
Kopka, J., Schauer, N., et al., 2005. GMD @ CSB. DB: the golm metabolome database. Bioinformatics 21, 1635–1638.
Kumar, S., Sreeharsha, R.V., et al., 2017. Molecular insights into photosynthesis and carbohydrate metabolism in Jatropha curcas grown under elevated CO2 using transcriptome sequencing and assembly. Sci. Rep. 7, 11066.
Liu, F., Zhang, X., et al., 2015. Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis, J. Exp. Bot. 66 (19), 5663–5681.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real time quantitative PCR and the $2-\Delta\Delta CT$ method. Methods 25, 402–408.
Luthra, R., Munshi, S.K., et al., 1991. Elationship of carbohydrate metabolism with lipid biosynthesis in developing sunflower (*Helianthus annuus* L.) seeds. J. Plant Physiol. 137 (3), 312–318.
Memon, A.R., 2004. The role of ADP-ribosylation factor and SAR1 in vesicular trafficking in plants. Biochim. Biophys. Acta 1664 (1), 9–30.
Miranda, M., Borisjuk, L., et al., 2001. Amino acid permeases in developing seeds of *Vicia faba* L.: expression precedes storage protein synthesis and is regulated by amino acid supply. Plant J. 28, 61–71.
Mudalkar, S., Golla, R., et al., 2014. De novo transcriptome analysis of an imminent biofuel crop, Camelina sativa L. Using Illumina GAIIX sequencing platform and identification of SSR markers. Plant Mol. Biol. 84, 159–171.
Muktham, R., Ball, A.S., et al., 2016. Bioethanol production from non-edible de-oiled *Pongamia pinnata* seed residue-optimization of acid hydrolysis followed by fermentation. Ind. Crops Prod. 94, 490–497.
Niu, Y., Wu, G.Z., et al., 2009. Global analysis of gene expression profiles in *Brassica napus* developing seeds reveals a conserved lipid metabolism regulation with *Arabidopsis thaliana*. Mol. Plant 2, 1107–1122.
Pavithra, H.R., Gowda, B., et al., 2012. Oil, fatty acid profile and karanjin contentin developing *Pongamia pinnata* (L.) Pierre seeds. J. Am. Oil Chem. Soc. 89, 2237–2244.
Pavithra, H.R., Gowda, B., et al., 2014. Biochemical changes in the composition of developing seeds of *Pongamia pinnata* (L.) Pierre. Ind Crops Pro6d 53, 199–208.
Rolletschek, H., Radchuk, R., et al., 2005. Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. New Phytol. 167, 777–786.
Sbihi, H.M., Nehdi, I.A., et al., 2018. Study of oxidative stability and cold flow properties of Citrillus colocynthis oil and *Camelus dromedaries* fat biodiesel blends. Ind. Crops Prod. 122, 133–141.
Schiltz, S., Munier-Jolain, N., et al., 2005. Dynamics of exogenous nitrogen partitioning and nitrogen remobilization from vegetative organs in pea revealed by 15N in Vivo labeling throughout seed filling. Plant Physiol. 137 (4), 1463–1473.
Schwender, J., Goffman, F., et al., 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature 432, 779–782.
Schwender, J., Hebbelmann, I., et al., 2015. Quantitative multilevel analysis of central metabolism in developing oilseeds of oilseed rape during in vitro culture. Plant Physiol. 168 (3), 828–848.
Sengupta, D., Kannan, M., et al., 2011. A root proteomics-based insight reveals dynamic regulation of root proteins under progressive drought stress and recovery in *Vigna radiata* (L.) Wilczek. Planta 233 (6), 1111–1127.
Sharma, S., Guleria, S., et al., 2011. Lipid accumulation in developing soybean: influence of seed position on stem axis. Genet. Plant Physiol. 1, 56–67.
Siddiqi, S., Mansbach, C.M., 2012. Phosphorylation of Sar1b protein releases liver fatty acid-binding protein from multiprotein complex in intestinal cytosol enabling it to bind to endoplasmic reticulum (er) and bud the pre-chylomicron transport vesicle. J. Biol. Chem. 287 (13), 10178–10188.
Singh, Y., Sharma, A., et al., 2018. Optimization of performance and emission parameters of direct injection diesel engine fuelled with Pongamia methyl esters-response surface methodology approach. Ind. Crops Prod. 126, 218–226.
Sreeharsha, R.V., Mudalkar, S., et al., 2016. Unravelling molecular mechanisms from floral initiation to lipid biosynthesis in a promising biofuel tree species, *Pongamia pinnata* using transcriptome analysis. Sci. Rep. 6, 34315.
Sun, C., Jia, L., et al., 2017. Natural variation in fatty acid composition of *Sapindus spp*. Seed oils. Ind. Crops Prod. 102, 97–104.
Tan, H., Xie, Q., et al., 2015. Dynamic metabolic profiles and tissue-specific source effects on the metabolome of developing seeds of Brassica napus. PLoS One 10 (4), e0124794.
Weber, H., Borisjuk, L., et al., 2005. Molecular physiology of legume seed development. Annu. Rev. Plant Biol. 56, 253–279.
Weselake, R.J., Taylor, D.C., et al., 2009. Increasing the flow of carbon into seed oil. Biotechnol. Adv. 27 (6), 866–878.
Wright, D., Lenssen, A.W., 2013. Staging soybean development. Agri.Environ.Exten. Pub. 191.
Xiong, B., Zhang, Z., et al., 2018. Biodiesel from *Lindera glauca* oil, a potential non-food feedstock in Southern China. Ind. Crops Prod. 122, 107–113.

SCIENTIFIC REPORTS

OPEN

Received: 10 May 2016
Accepted: 12 September 2016
Published: 28 September 2016

Unravelling molecular mechanisms from floral initiation to lipid biosynthesis in a promising biofuel tree species, *Pongamia pinnata* using transcriptome analysis

Rachapudi V. Sreeharsha*, Shalini Mudalkar*, Kambam T. Singha & Attipalli R. Reddy

Pongamia pinnata **(L.) (Fabaceae) is a promising biofuel tree species which is underexploited in the areas of both fundamental and applied research, due to the lack of information either on transcriptome or genomic data. To investigate the possible metabolic pathways, we performed whole transcriptome analysis of Pongamia through Illumina NextSeq platform and generated 2.8 GB of paired end sequence reads. The *de novo* assembly of raw reads generated 40,000 contigs and 35,000 transcripts, representing leaf, flower and seed unigenes. Spatial and temporal expression profiles of photoperiod and floral homeotic genes in Pongamia, identified GIGANTEA (GI) - CONSTANS (CO) - FLOWERING LOCUS T (FT) as active signal cascade for floral initiation. Four prominent stages of seed development were selected in a high yielding Pongamia accession (TOIL 1) to follow the temporal expression patterns of key fatty acid biosynthetic genes involved in lipid biosynthesis and accumulation. Our results provide insights into an array of molecular events from flowering to seed maturity in Pongamia which will provide substantial basis for modulation of fatty acid composition and enhancing oil yields which should serve as a potential feedstock for biofuel production.**

Decreasing the fossil fuel consumption and reconciling the worsening global environmental conditions are fundamental concerns of the society in this industrial era. The development and use of alternative fuels, including bioethanol and biodiesel are predicted to significantly alleviate the problems caused by the usage of fossil fuels. An imminent biofuel tree *Pongamia pinnata* (L.) (Family: Fabaceae), is a native species of India which can be grown in diverse tropical and subtropical marginal lands of the world. It is a drought and salinity tolerant, semi-deciduous, nitrogen fixing tree which grows up to 15–20 meters in height with a large canopy[1]. The oil content of Pongamia seeds ranges from 35 to 40% of seed dry weight and 55% of it is oleic acid which is the ideal fatty acid for good quality biodiesel production. Upon trans-esterification, the Pongamia oil, with a blend of diesel, can be applied in automobiles without any further modification of engines. Pongamia has a long life cycle, usually sets flowering after 4–5 years of plantation and takes 9–11 months to form a mature pod after anthesis. There are many positive attributes that could be potentially achieved by understanding the genetics and genomics of Pongamia. For instance, the oil content of Pongamia which is around 35% could be increased to about 50% with higher oleic acid content and also the fruit maturity time could be reduced. But, limited genetic resources and long production cycles have constrained the molecular breeding programs aimed at better oil quality and seed yields in this potential biofuel tree species.

Flowering time, fertilization and seed development are inter-related and determine the yield of certain promising biofuel tree species, including Pongamia. Understanding the molecular mechanisms that control the onset of reproductive events and development of the seeds is crucial for improving the biofuel feedstock. Photoperiod and vernalization pathways have been reported as major control mechanisms to synchronize environmental cues with the internal rhythm[2–4]. The former promotes flowering in response to increasing day length and the latter

Department of Plant Sciences, University of Hyderabad, Hyderabad, 500046, India. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.R.R. (email: arrsl@uohyd.ernet.in)

enables induction of flowering following a prolonged exposure to cold. In order to trigger flower initiation at the precise time and in the right conditions, circadian clocks perceive and integrate both environmental and endogenous signals, which in turn activate mobile florigen and other related proteins to induce flowering. Besides the floral promotion pathways, mutation studies in *Arabidopsis* have also revealed the existence of genes that repress the floral transition[5]. Exploiting these genes through genetic engineering by rendering knowledge obtained from the model plants like *Arabidopsis* and sequence information obtained from transcriptome will have substantial impact on floral transitions in Pongamia. However, lack of publicly available sequence information is a hindrance to the efforts for improving seed oil quality and floral transition period in non-model and non-edible biofuel crops like Pongamia.

In many seeds, triacylglycerols (TAGs) that accumulate during the maturation phase of embryo and/or endosperm act as major storage reserves of carbon and in the due course of germination, they support the establishment of the seedling[6]. Fatty acid (FA) metabolism and the enzymes involved in it play an important role in plant morphology, growth, seed development and stress responses. Earlier reports have shown that FATB mutant of *Arabidopsis* resulted in reduced concentrations of palmitate and stearate and affect both plant growth and seed development[7]. Similarly, enoyl-CoA reductase and stearoyl – ACP desaturase had essential roles in endocytic membrane trafficking and defence responses respectively by regulating the oleic acid contents[8,9]. Also, owing to the commercial importance of plant lipids, several attempts have been made to alter seed TAG composition through genetic engineering. For instance, *Arabidopsis* and *Camelina* transgenic plants were engineered to produce oil with high omega-3/omega-6 ratio and high DHA[10]. Engineering of key FA biosynthetic enzymes, including fatty acid desaturase 2 (FAD2) and fatty acid elongase 1 (FAE1) has resulted in high oleic acid containing lines of *Camelina* and soybean[11,12]. Oil accumulation depends on the seed development patterns which show great diversity in duration of maturity and oil formation stages. Knowledge about temporal expression patterns of oil biosynthetic enzymes is crucial to understand reprogramming strategies of oil biosynthesis as well as modification in fatty acid composition.

High throughput deep sequencing of transcriptome is a promising and powerful tool to identify the key genes associated with species-specific exotic FA biosynthetic enzymes and molecular marker development. Of late, several non-model organisms, whose reference genome sequence was absent, were sequenced and annotated using platforms such as Roche/454, AB SOLiD and Illumina[13–17]. Recently, salt responsive genes were identified upon transcriptome sequencing of leaf and root tissues of *Pongamia pinnata* using Illumina platform[18]. In the same year, another group had claimed the chloroplastic and mitochondrial genome of Pongamia through second generation DNA sequencing[19]. Further, Pavithra *et al.*[20], reported the FA profiles of seeds at different developmental stages of Pongamia. Very recently, parallel to the current study, Huang *et al.*[21], reported the seed transcriptome of Pongamia which provided valuable information about lipid biosynthetic genes and SSR markers. But the whole transcriptome information, together with gene expression patterns of lipid biosynthetic enzymes during seed development and flowering pathway genes, are far from being characterised in Pongamia. In the present study, we constructed the paired end cDNA library from pooled RNA isolated from leaf, flower, pod and seed tissues of mature Pongamia tree and sequenced using Illumina TrueSeq protocol on Illumina NextSeq 500 platform. We also data mined circadian clock genes and lipid biosynthetic genes in Pongamia and reported their homology with other model organisms. Our data provide a comprehensive information on Pongamia transcriptome and temporal expression of candidate genes involved in lipid accumulation and flowering which can be applied to molecular breeding programs for improving the seed oil content in Pongamia.

Results

Sequencing and *de novo* assembly of Pongamia transcriptome. We generated a total of 24 004 632 paired end sequence reads, each 76 bp in length, encompassing about 2.8 GB of sequence data in fastQ format. After stringent filtering of sequence data for low-quality reads and reads containing primer/adaptor sequences, we obtained a total of 22 158 278 high quality sequence reads (with phred quality score of <20). The final data set comprising ~22 million very high-quality reads were used for optimization of *de novo* assembly and analysis of Pongamia transcriptome. All processed reads were assembled into contigs without any reference (*de-novo*) using velvet – 1.2.10 software. Assembly was tried on various hash lengths (k-mers) and 41 was selected as the best hash-length. Best k-mer is decided on various parameters including: number of contigs, total number of reads used, total contig length and number of non-ATGC characters. A total of 42,724 contigs were generated with maximum and minimum lengths of 11,665 and 200 bp respectively with an average read length of 639.7 bp. Contigs were then processed into transcripts (spliced isoforms) using oases – 0.2.8 software. The majority of high quality reads (81.77%) were assembled to generate a total of 36,047 transcripts that ranged from 200–30, 656 bp length with an average length of 937.586 bp. The number of transcripts were decreased as the length increased with maximum number of transcripts falling in the range of 0–100 bp followed by 100–500 bp. The N50, N90 and rpkm values of Pongamia transcriptome along with other related parameters were presented in Supplementary Table S1. The average GC content of Pongamia transcripts was 48% and has a higher proportion of transcripts in the range of 40–45%, followed by 45–50%, resulting much broader GC content range (Supplementary Fig. S1).

Functional annotation and characterisation of Pongamia transcripts. The assembled transcripts were annotated against NCBI-BLAST 2.2.29 using GeneMark software. A total of 47,461 genes/proteins were predicted of which 25,112 proteins were annotated with either Swiss-Prot (108) or TrEmbL (25, 004) databases. Most of the predicted proteins in Pongamia showed homology in UniProt with *G. max* (44%) and *G. soja* (22%) followed by *P. vulgaris* (21%) and other Papilinoideae members (Supplementary Data 1). The transcripts that show significant homology to the genes against UniProt database were selected for GO annotation. A total of 16,146 (45%) transcripts were assigned with at least one GO term, in which 9508 (26.37%) were assigned in biological process category, 4420 (12.26%) were assigned in cellular component category and 2218 (6.15%) were assigned to molecular function category (Fig. 1a). Among the various biological processes, ignoring unknown

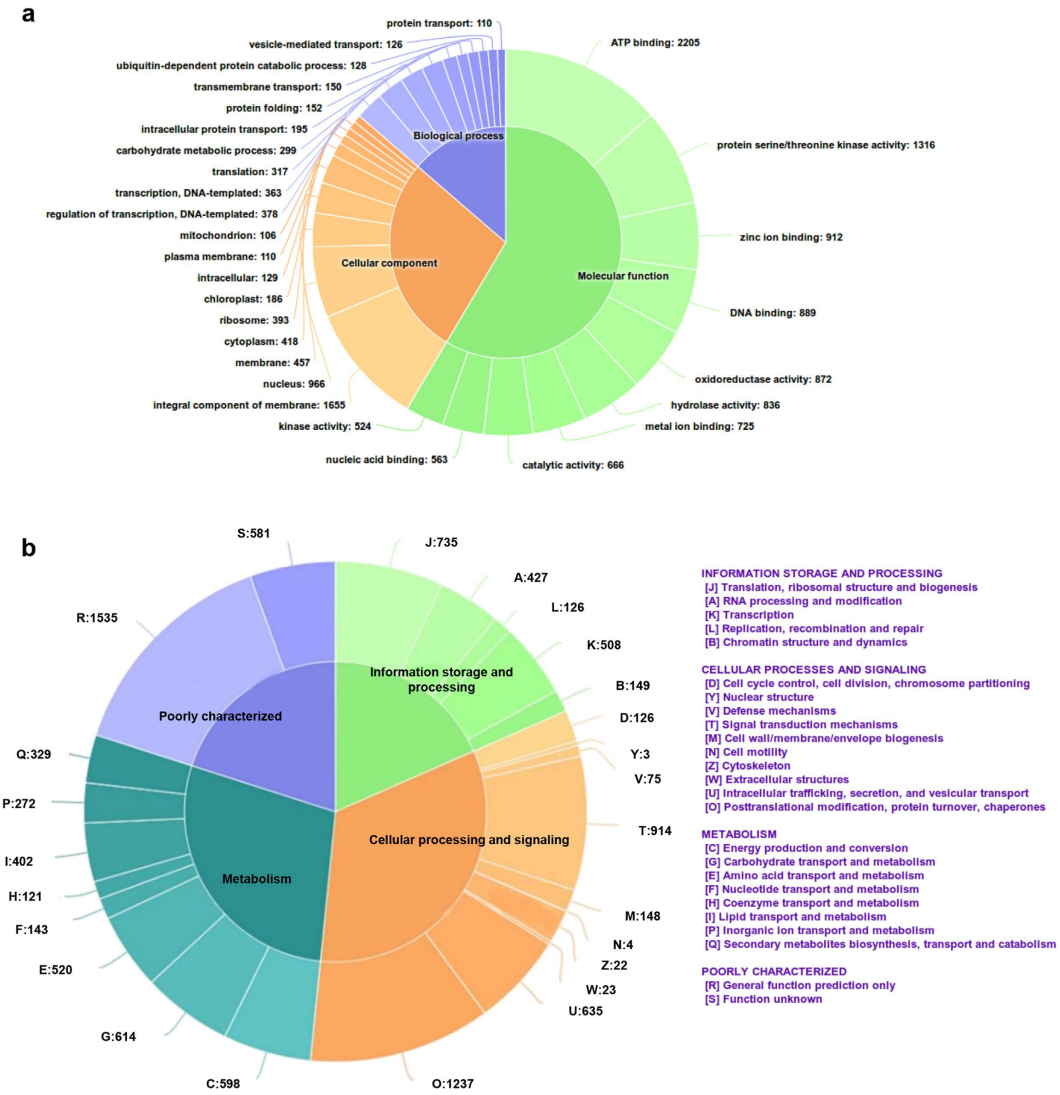


Figure 1. GO and KOG classification. (**a**) Gene ontology distribution of the transcripts into biological process, molecular function and cellular component. The number of transcripts encoded for each category is represented. (**b**) Comparison of transcripts with the KOG database and classification into groups such as metabolism, information storage and processing, cellular processes and signalling resulting in 25 different categories.

and other biological process categories, ATP binding (2205) and protein serine/threonine kinase activity (1316) were highly represented. The genes involved in other biological processes such as zinc ion binding, DNA binding, oxidoreductase activity, hydrolase activity and metal ion binding and those having catalytic activity were also identified through GO annotations (Fig. 1a). Similarly, genes involved in transcription and transcription regulation were mostly represented in molecular function category, followed by carbohydrate, protein metabolism and transport (Fig. 1a). Integral membrane components and nucleus were most represented among the cellular components followed by membrane and cytoplasm. Also, we annotated 13,764 Pongamia genes to KOG (Eukaryotic Orthologous Groups) database which aids in identification and phyletic classification of the orthologous proteins, coded in whole genome of almost 21 organisms including bacteria, algae and eukaryotes[22]. The resulting KOG annotation grouped the transcripts into three functional categories: cellular processing and signalling (3214; 31%); metabolism (2999; 30%) and information storage and processing (1945; 19%) and rest of the genes resulted in poorly characterized annotations (Fig. 1b) (Supplementary Data 2). We identified a total of 4148 SSRs in which mono-nucleotide SSRs represented the largest fraction (36.4%) followed by tri-nucleotide (31.3%) and di-nucleotide (28.8%) SSRs (Supplementary Fig. S2). Pongamia transcripts also contained a quite significant number of tetra- (79), penta- (31) and hexa – nucleotide (28) SSRs though their representation is small in total SSR pool.

Sequence similarity of Pongamia transcripts with other plants. The transcripts of Pongamia were analysed for similarity against the unigene datasets of legume crops, biofuel plants and other oil bearing plants

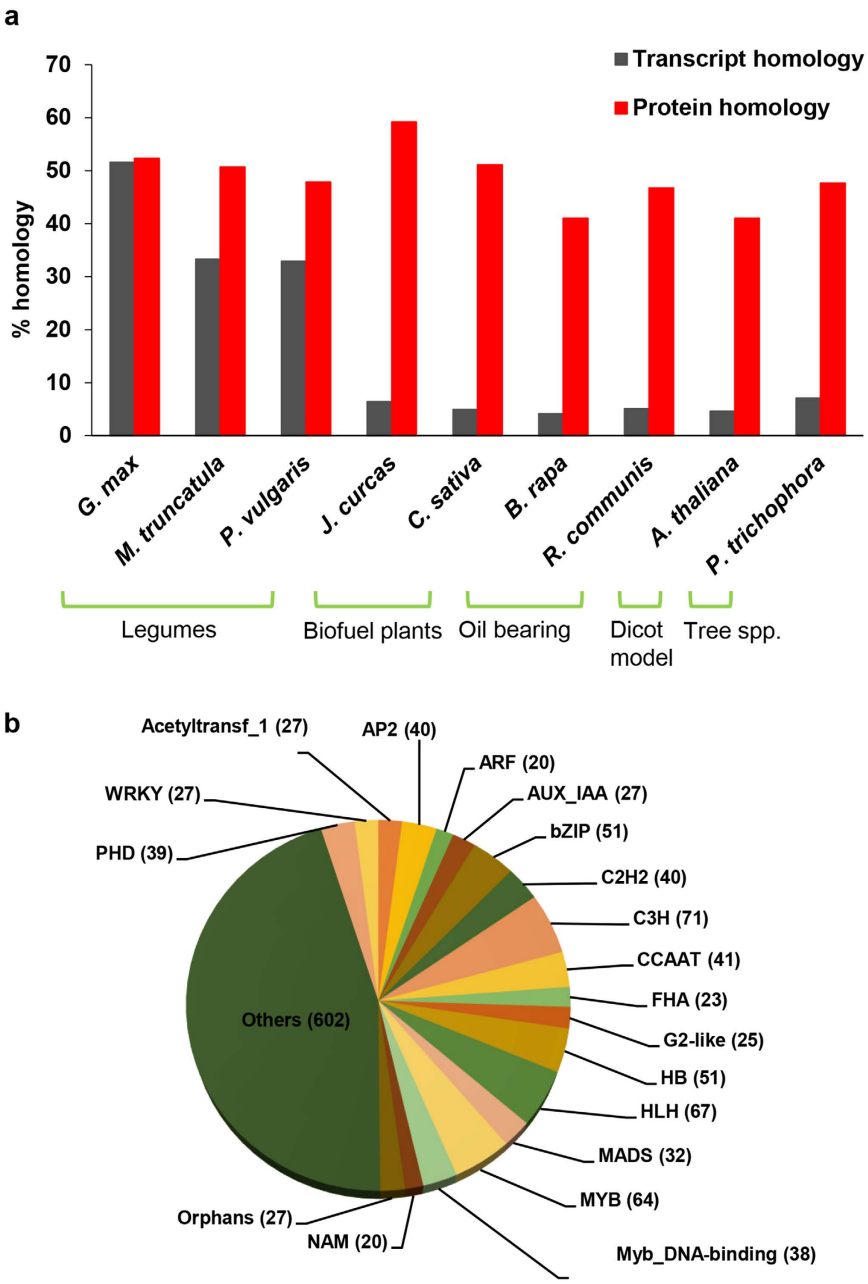


Figure 2. (**a**) Sequence similarity and percentage homology of Pongamia transcripts and putative proteins with other related organisms. (**b**) Classification of transcripts into transcription factor families. Number of genes represented in each family were indicated in parenthesis.

belonging to different families using TBLASTX search. An E-value cut-off threshold of 1E – 05 was considered to define a significant hit. The largest number of Pongamia transcripts showed significant similarity with soybean putative mRNA sequences (53%) followed by *M. truncatula* (35%). While, Pongamia showed little conservation with oil bearing trees of other families (Fig. 2a). We also analysed the sequence conservation of translated Pongamia transcripts with proteomes of selected plant species. Putative Pongamia proteins showed maximum homology with the biofuel plant *J. curcas* (59.2%) followed by *C. sativa* (52%) and *G. max* (53%) (Fig. 2a). Our analysis also showed that 53% of Pongamia transcripts are having homology with legumes indicating these genes are legume specific (Supplementary Data 3).

Identification of transcription factor families. We identified the transcription factor encoding transcripts by sequence comparison to known transcription factor gene families in Plant TFDB (Transcription factor data base). In total, 1332 putative transcription factor genes distributed in at least 18 families were identified representing 3.7% of Pongamia transcripts (Supplementary Data 4). Genes encoding for C3H, HLH, MYB, bZIP and HB transcription factor families were abundantly expressed while minimum number of transcripts were observed

for NAM, ARF, FHA, MADS and WRKY families (Fig. 2b). Further, we annotated and analysed transcription factors exclusive to lipid biosynthetic pathway which resulted in identification of genes belonging to MYB, PLATZ, GRAS, MYB-related, bHLH8, CCAAT, G2-like and PHD transcription factor families.

Pathway mapping of transcripts by KEGG. Ortholog assignment and mapping of the contigs to the biological pathways were performed using KEGG automatic annotation server (KAAS). All the contigs were compared against the KEGG database using BLASTX with threshold bi-score value of 60 (default). It assigned Enzyme Commission numbers for 2784 contigs, and they were mapped to respective pathways (Supplementary Data 5). Among the mapped contigs, 1709 were identified as genes involved in metabolic pathways of major biomolecules such as carbohydrates (306, 11%), amino acids (252, 9%), lipids (177, 6.3%), nucleotides (150, 5.3%), cofactors, vitamins (130, 4.6%), glycans (71, 2.5%), terpenoids (74, 2.5%). The KEGG pathway analysis also showed that 186 and 69 contigs represent the energy and secondary metabolites metabolisms respectively. A total of 247 transcripts that represent enzymes involved in carbon metabolism, fatty acid metabolism, degradation of aromatic compounds, biosynthesis of amino acids as well as 2-oxocarboxylic acid metabolism were also identified. Further, the mapped contigs also represented the genes involved in genetic information processing that include, translation (12.3%), folding, sorting and degradation (9.5%), transcription (5.7%) as well as replication and repair (4.5%). Cellular processes (transport and catabolism, cell motility, cell growth and death, cell communication) and environmental information processing (membrane transport, signal transduction, signalling molecules and interaction) are other minor groups represented in the KEGG annotation of Pongamia. The KAAS analysis also represented genes involved in biosynthesis of karanjin, ansamycin and siderophore thus substantiating the insecticidal, medicinal and anti-bacterial properties of Pongamia respectively (Supplementary Data 5). Further, we studied genes involved in the following major metabolic events which had a prominent role in improving the yield and oil quality related traits in Pongamia.

Genes involved in circadian rhythms. Our transcriptome data represented 25 key genes including TFs that are involved in the regulation of floral meristem identity, photoperiod as well as vernalization pathways (Table 1). Majority of the genes were having sequence homology with *G. max* followed by *M. truncatula* which shows the evolutionarily conserved relationship between legumes. The phylogenetic relationship of Pongamia flowering genes with other related organisms was deduced (Fig. 3a). ORF sequence analysis unfold the complete protein coding sequence of all the genes and the polypeptide information was presented (Table 1). Genes like Flowering locus T (FT), GIGANTEA (GI), Chalcone synthase (CS), PRR1/TOC1, PRR5, PRR7 represented in this study were known to promote flowering through photoperiod regulation under long day conditions. The Pseudo-receiver (PR) domain at the N-terminal region and CCT domain at the C-terminal region which are characteristic conserved domains of PRR protein family were identified in Pongamia compared with other related organisms (Fig. 3b). Other genes like PST, SPT, APT, CAU, AGA which encode TFs play a key role in defining floral meristem identity. The transcriptome also represented COP1, SPA1, PHYA, PHYB and CK2A genes which inhibit flowering process by repressing CONSTANS (CO) gene during prolonged cold periods through vernalization responses.

Spatial and temporal profiling revealed the diurnal nature of clock genes. To understand the diurnal behaviour of circadian clock genes and expression trends of floral homeotic genes, we quantified the expression of key flowering pathway genes in leaf tissues collected at four different time points in a day (6, 12, 18, 24 hrs) as well as in inflorescence of four different stages (10, 20, 30, 40 days after flowering - DAF) (Stage 1, 2, 3 and 4 respectively) in a field grown Pongamia plant (Fig. 4a). We selected photoperiod pathway genes, vernalisation genes and certain crucial transcription factors which act as floral homeotic proteins and a heatmap was constructed based on their expressin profiles (Fig. 4b). The genes were divided into three clusters based on their expression profiles (i) circadian clock genes that function in initiation of flowering through photoperiod pathway: genes like PRR1, PRR5, PRR7 showed time dependent regulation in leaves wherein, the three genes showed significant up regulation in the morning and decreased expression during dusk (Fig. 4b). Contrastingly, ELF3 which is an evening gene showed peak expression during 18–24 hrs. GI which acts upstream to FT, is a major component of leaf generated mobile florigen, showed coordinated expression with FT wherein, peak expression at 6 and 12 h of day was observed. At the same time, they showed minimal expression in inflorescence at all stages. (ii) Circadian clock genes that repress flowering process: COP1-SPA1 complex which operates in dark conditions to degrade CO protein by polyubiquitination showed downregulation as the day light progressed. PHYA, PHYB and CK2A were only expressed during midday in leaves but showed significant up regulation in earlier stages of inflorescence (Fig. 4b). (iii) TFs which operate in the process of flower development and photoperiod: APT, PTL, AGA, CAU and SPT showed significant and constitutive expression in all stages of inflorescence. The expression of PTL, SPT and APT was significantly peaked at stage 2 of inflorescence whereas, CAU, AGA showed higher expression at stage 1. MYB75 TF expression was significantly high in flowers and roots compared to leaf tissue. LHY which is a MYB – related TF that binds to the promoter region of PRR1/TOC1 thereby repressing their expression to postpone the flowering process, showed basal level of expression in leaves and significant up regulation in flowers (Fig. 4b).

Genes involved in membrane and storage lipid metabolism. A total of 203 transcripts corresponding to 136 unigenes were identified and grouped into 14 categories of various lipid metabolic pathways (Supplementary Table S3). The KEGG annotation, sorted out 264 unigenes representing all essential enzymes involved in fatty acid biosynthesis, elongation, degradation as well as glycerolipid and phospholipid metabolisms. From our data, it is evident that 41 genes have mapped to transcripts at two different loci indicating the diploid nature in Pongamia and 20 genes have duplicate contigs at the same locus presuming the gene duplication and

Enzyme name	Symbol	U	CDS (bp)	Homology (%)	PP (aa)	Mass (KDa)	pI
Pseudo-response regulator 1	PRR1	1	1710	85 (*G. max*)	269	30.20	5.89
Pseudo-response regulator 5	PRR5	4	2671	82 (*P. trichocarpa*)	687	76.20	6.48
Pseudo-response regulator 7	PRR7	3	2984	83 (*M. truncatula*)	758	83.00	6.27
Phytochrome-interacting factor3	PIF3	3	2005	79 (*G. max*)	288	32.47	8.99
Casein kinase II alpha	CK2A	2	1447	93 (*G. max*)	333	39.30	8.14
Casein kinase II beta	CK2B	9	1597	86 (*M. truncatula*)	284	31.96	5.06
MYB75	MYB75	1	485	88 (*G. max*)	136	15.20	7.80
LHY – MYB related	LHY	3	3305	88 (*P. vulgaris*)	750	82.16	6.07
Chalcone synthase	CS	4	1634	78 (*P. frutescus*)	391	42.96	6.24
Constitutive Photomorphogenic 1	COP1	1	2392	89 (*P. sativum*)	672	75.56	6.94
Clock-associated PAS protein	ZTL	4	2669	91 (*G. max*)	613	66.70	5.44
Phytochrome A	PHYA	1	1966	91 (*G. max*)	332	36.57	6.34
Phytochrome B	PHYB	1	2040	91 (*G. max*)	268	30.75	5.45
GIGANTEA	GI	2	3807	85 (*G. max*)	1159	127.0	6.30
Protein suppressor of PHYA	SPA	1	2267	76 (*A. thaliana*)	529	59.33	5.48
EARLY FLOWERING 3	ELF3	1	1138	81 (*P. sativum*)	316	34.38	8.95
Cryptochrome 1	CRY1	1	1733	88 (*G. max*)	368	41.82	4.96
Cryptochrome 2	CRY2	1	2477	89(*G. max*)	634	72.20	6.16
FLOWERING LOCUS T	FT	1	768	—	178	19.70	5.57
PISTILLATA	PST	1	1089	91 (*G. max*)	226	26.16	9.15
SEPALLATA	SPT	1	1104	86 (M. truncatula)	243	28.10	8.95
APETALA 2	APT2	1	2161	84 (C. cajan)	536	59.54	5.91
APETALA 3	APT	1	1206	90 (G. max)	247	28.4	8.94
CAULIFLOWER/APETALA 1	CAU	1	1132	94 (V. ungiculata)	236	27.40	8.91
FRIGIDA	FRI	1	2000	88 (C. cajan)	549	61.37	6.28
AGAMOUS	AGA	1	971	94 (C. cajan)	243	28.0	9.41

Table 1. Flowering and circadian rhythms related genes. The number of unigenes (U) and the length of the cDNA sequence (CDS) as well as the percentage homology of Pongamia transcripts with other organisms having maximum sequence coverage are represented. The protein parameters like polypeptide (PP) length, mass and pI were deduced using Expasy ProtParam tool.

the presence of paralogs. The overall lipid metabolism is an interplay between carbohydrate and fatty acid stoichiometric profiles and can be viewed as three major events: (i) pyruvate to Acetyl CoA synthesis (ii) FA synthesis from Acetyl CoA (iii) TAG assembly and degradation. Our transcriptome data represented all the major rate limiting enzymes involved in these three categories (Table 2). We also analysed the homology of Pongamia transcripts with respective organisms having maximum sequence coverage. ORF analysis indicated that the putative Pongamia transcripts were having full length sequences and the polypeptide information (mass and pI) of translated putative transcripts suggested that the proteins involved in the oil biosynthetic pathway are functional mostly at basic pH (Table 2). We further searched the orthologs for PpFAD8, PpFAD2, PpFAD6 and PpSAD to establish the phylogenetic relationship of Pongamia with other legumes, oil bearing trees and crop plants. Our data demonstrated PpFAD6 and PpSAD are part of a small clades consisting of respective genes from *Glycine max* and *Glycine soja* (Fig. 5a). Intriguingly, PpFAD8 was grouped with neither legumes nor biofuel trees and formed a separate clade indicating its distinct evolutionary conservation (Fig. 5a). Also, phylogeny of certain key flowering related genes of Pongamia was established. Considerable number of transcripts were also found for arachidonic acid (4%) and linoleic acid (2%) metabolism in our transcriptome data. Besides lipids for oil biosynthesis, enzymes involved in membrane lipid biosynthesis, including glycerolipid metabolism, glycerophospholipid metabolism, sphingolipid metabolism, steroid biosynthesis, ether lipid metabolism and cutin as well as suberin and wax synthesis were also identified through our data which corroborated that the transcriptome has covered all the genes involved in lipid biosynthesis signifying the depth of the sequencing (Supplementary Table S3).

Profiling of lipid biosynthetic genes revealed stage specific FA synthesis during seed development. Our data on seed oil accumulation showed that, Pongamia accumulate oil actively after 210 DAF and progressed with the pod development till maturity (300 DAF) (Fig. 5b). The oil red staining of total lipids in the seed sections also showed a gradual increase in the lipid content of seeds during the four stages (Fig. 5c). Further, mRNA expression levels of genes coding for oil biosynthetic enzymes were studied during four stages of seed development which include: mature green pod stages (210 DAF, 240 DAF and 270 DAF) (Stage 1, 2, 3 respectively) and late dark brown pod stage (300 DAF) (stage 4) that were normalised with immature green pod stage (150 DAF) (Fig. 5d). Expression profiles of all the genes were shown in a schematic representation of lipid metabolism (Fig. 6). Based on their expression levels at four different stages, the lipid biosynthetic genes were categorised into 4 groups: (a) those which expressed in a bell shaped manner which include, MAT, KASII, KASIII,

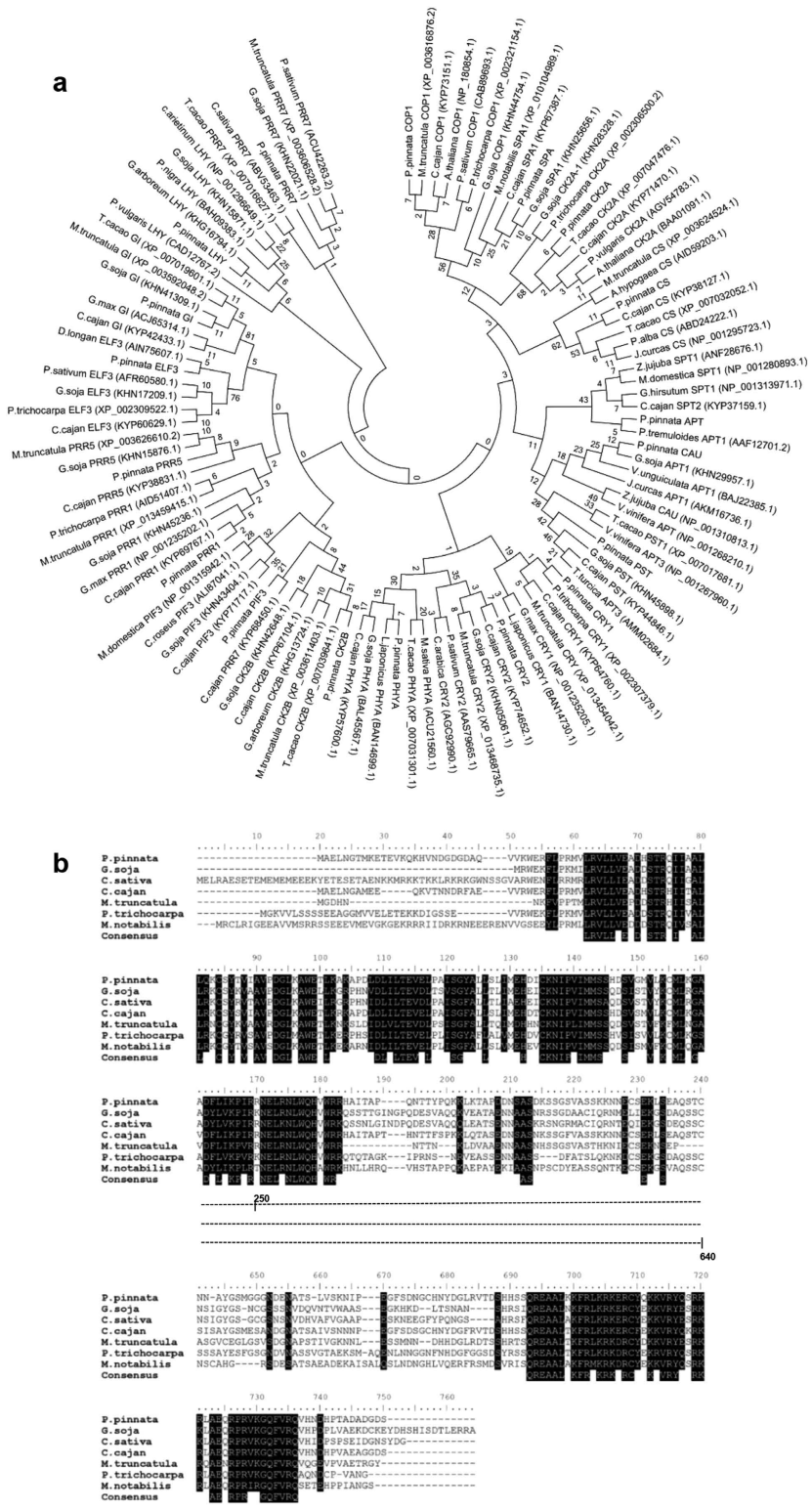


Figure 3. Sequence homology and phylogenetic analysis. (**a**) Pongamia flowering related genes and their phylogenetic relationship with ortholog proteins from other related organisms. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7. The accession numbers of all the genes were given in parenthesis. (**b**) multiple sequence alignment of the PpPRR5 with orthologs from other related organisms depicting the conserved C - terminal and N - terminal regions (highlighted black). BioEdit software was used to align the sequences.

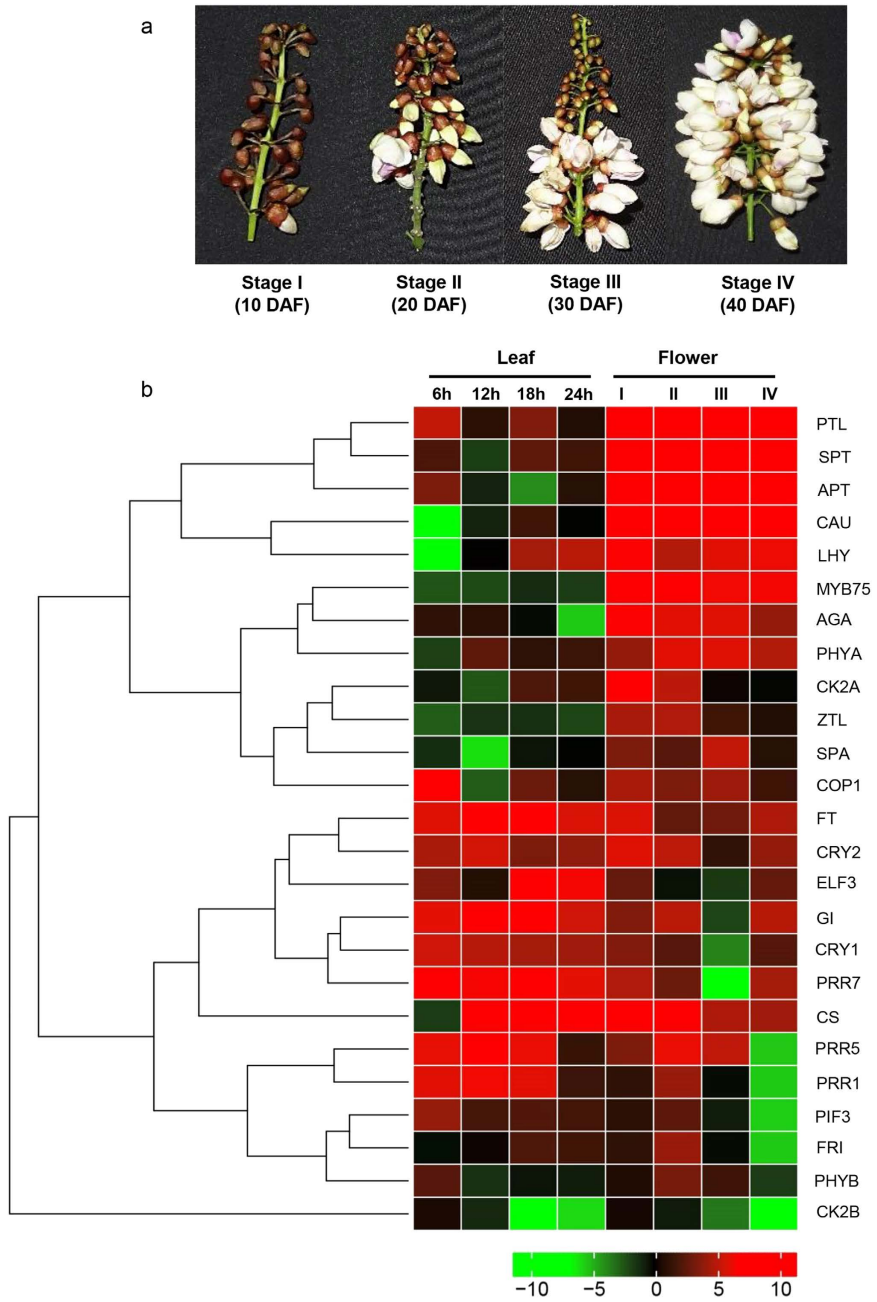


Figure 4. (**a**) Infloresence of Pongamia at four different developmental stages used in gene quantification studies. (**b**) Heatmap representing expression profiles of flowering related genes and transcription factors in Pongamia leaves and flowers. Heatmap was constructed using gplot in R-package. For gene abbreviations see Table 1.

KAR, HAD, EAR, LPAT, PAP. These genes showed peak expression during 2nd and 3rd stages and decreased thereafter. (b) Those which showed a gradually decreased expression as the development progressed: FATB and PDAT. (c) Those which showed increased expression towards the development of pod: Thiolase, HDH, ECH, ACD which majorly involved in β-oxidation and (d) those which expressed constantly throughout the pod development like ACC, LACS, DGAT (Fig. 6).

Discussion

The next generation sequencing technologies and bioinformatics tools enable assembly and annotation of short reads into expressed sequence data, particularly for non-model organisms without a known reference. In this study, using the Illumina NextSeq platform, we characterized whole transcriptome of a non-model legume biofuel tree Pongamia, for which the sequence data are limited so far in the public databases. Genes related to lipid biosynthesis, flowering cycle and flavonoid biosynthesis were emphasised and the transcript information was further used to understand the temporal expression patterns of oil biosynthetic and flowering related

Enzyme name	Symbol	U	CDS (bp)	Homology (%)	PP (aa)	Mass (KDa)	pI
Acetyl CoA synthesis from pyruvate							
Acetyl CoA synthesis							
PDHC - E2 component	DLAT	7	2350	88 (*M.truncatula*)	625	67.72	6.05
PDHC - E1 component α	PDHA	4	1749	82 (*M.truncatula*)	429	47.57	6.52
PDHC - E1 component β	PDHB	11	1910	91 (*G.max*)	402	43.86	5.75
ATP – Citrate lyase	CL	5	2082	93 (*L. albus*)	615	66.81	7.57
Acetyl CoA synthetase	ACSS	2	2395	88 (*M.truncatula*)	754	84.01	5.98
Acetyl CoA acetyltransferase	AAT	2	1612	83 (*P. trichocarpa*)	414	42.86	8.58
FA synthesis from Acetyl CoA							
Fatty acid biosynthesis							
ACCase Carboxyl transferase α	ACC-CT	1	2990	85 (*G.max*)	727	80.0	6.80
ACCase Carboxyl transferase β	ACC-CT	1	1437	81 (G. max)	278	29.45	8.67
ACCase Biotin carboxylase	ACC-BC	1	2050	95 (*C. cajan*)	540	59.17	6.97
ACCase homomeric protein	ACAC	1	3894	95 (*G. max*)	1297	145.0	6.02
ACP – Malonyl transferase	MAT	1	1614	90 (*A.hypogea*)	385	40.0	8.53
Oxoacyl-ACP reductase	KAR	1	1213	79 (*M.truncatula*)	260	27.0	7.70
Oxoacyl-ACP synthase II	KAS-II	3	1816	93 (*G.max*)	469	49.7	8.31
Oxoacyl-ACP synthase III	KAS-III	1	1840	89 (*G.max*)	399	41.0	6.71
Enoyl-[ACP] reductase I	EAR	2	1406	91 (*G.max*)	393	40.0	8.64
3-hydroxyacyl-[ACP] dehydratase	HAD	2	932	82 (*M.truncatula*)	213	23.0	8.95
Fatty acyl-ACP thioesterase B	FATB	3	1552	91 (*G.max*)	417	46.0	6.53
Fatty acyl-ACP thioesterase A	FATA	1	1589	86(*M.truncatula*)	350	39.7	5.68
Long-chain acyl-CoA synthetase	ACSL	2	2414	89 (*M.sativa*)	662	74.1	6.60
Fatty acid elongation							
Ketoacyl-CoA synthase	KCS	1	2051	85 (*M.truncatula*)	521	58.0	9.09
Very-long-chain enoyl-CoA reductase	TER	1	1368	90 (*G.max*)	310	36.0	9.63
Palmitoyl-protein thioesterase	PPT	1	1508	85 (*P.vulgaris*)	321	36.0	6.46
Very-long-chain hydroxyacyl-CoA dehydratase	PHS1	1	916	88 (*V.radiata*)	218	24.7	9.23
Mitochondrial trans-2-enoyl-CoA reductase	MECR	1	1176	90 (*G.max*)	318	34.8	7.61
Very-long-chain 3-oxoacyl-CoA reductase	HSDB	2	1178	82 (*M.truncatula*)	320	35.8	9.48
Acyl-coenzyme A thioesterase 1/2/4	ACOT	1	1420	85 (*M.truncatula*)	429	48.4	8.45
Biosynthesis of unsaturated fatty acids							
Acyl-ACP desaturase	DESA1	1	1245	92 (*P.vulgaris*)	385	42.0	5.64
Stearoyl-CoA desaturase (Δ9 desaturase)	SAD	3	1400	80 (*M.truncatula*)	378	43.5	9.60
ω6 fatty acid desaturase (Δ12 desaturase)	FAD2/6	3	1600	88 (*G.max*)	363	42.1	8.80
ω3 fatty acid desaturase (Δ15 desaturase)	FAD8	5	1433	85 (*G.max*)	382	44.5	9.20
TAG assembly and degradation							
TAG biosynthesis							
Glycerol-3-phosphate acyltransferase	GPAT	2	1971	91 (*M.truncatula*)	376	43.5	9.15
Diacylglycerol O-acyltransferase	DGAT	1	821	90 (*G.max*)	155	18.8	9.18
LPA O-acyltransferase	LPAT	1	1293	91 (*M.truncatula*)	376	43.5	9.15
Phospholipid: diacylglycerol acyltransferase	PDAT	1	2723	87 (*M.truncatula*)	673	75.0	6.28
Phosphatidate phosphatase	PAP	2	3417	77 (*M. truncatula*)	751	83.1	4.86
Fatty acid degradation							
Alcohol dehydrogenase	ADH	1	1680	83 (*M.truncatula*)	425	46.2	6.41
Acyl-CoA oxidase	AOX	2	2089	90 (*M.truncatula*)	616	69.2	9.01
S- glutathione dehydrogenase	GDH	3	1314	87 (*G.max*)	376	40.2	5.42
Aldehyde dehydrogenase	ADH7	1	1930	92 (*G.max*)	508	54.1	5.71
Acetyl-CoA acyl transferase	ACAA	2	1731	91 (*G.max*)	461	48.5	7.06
3-hydroxyacyl-CoA dehydrogenase	HDH	2	2580	92 (*G.max*)	722	78.7	9.11
Acetyl-CoA acetyl transferase	AAT	2	1612	81 (*H. brasiliensis*)	414	42.8	8.58
Acyl-CoA dehydrogenase	ACD	1	2957	76 (*T. cocao*)	466	51.9	8.90
TAG lipase/phospholipase A2	TGL4	2	225	80 (*M.truncatula*)	189	20.7	4.49

Table 2. Genes involved in lipid biosynthesis. The number of unigenes (U) and the length of the cDNA sequence (CDS) as well as the percentage homology of Pongamia transcripts with other organisms having maximum sequence coverage are represented. The protein parameters like polypeptide (PP) length, mass and pI were deduced using Expasy ProtParam tool.

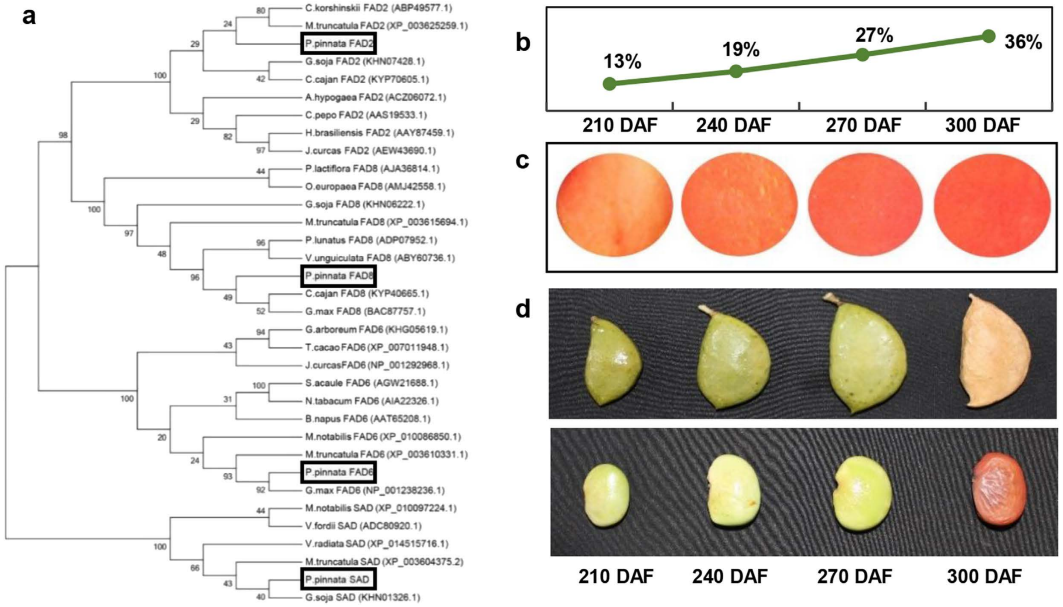


Figure 5. (**a**) Phylogenetic analysis: PpFAD2, PpFAD6, PpFAD8, PpSAD and their phylogenetic relationship with ortholog proteins from other related organisms. (**b**) Percentage of oil content during four developmental stages of Pongamia seed. (**c**) oil-red staining of seed endosperm showing oil accumulation patterns in Pongamia seed. (**d**) four developmental stages of Pongamia seed used in quantification studies of lipid biosynthetic genes.

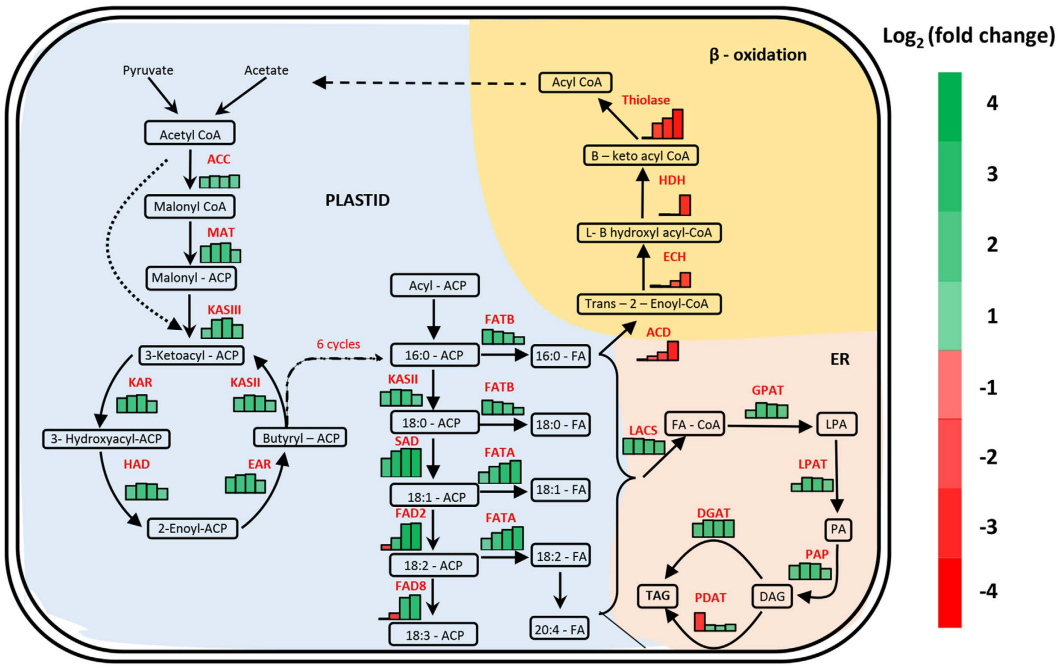


Figure 6. Schematic representation of oil biosynthetic pathway. The relative expression levels of genes involved in fatty acid biosynthesis, elongation, TAG assembly and degradation were analysed through qRT PCR. The fold change was $\log_2$ transformed and represented as a bar diagram with colour coding against each enzyme. LPA: lysophosphatidic acid; PA: phosphatidic acid; DAG: diacylglycerol; TAG: triacylglycerol. For gene abbreviations see Table 2.

genes in Pongamia. The total RNA from the four tissues were pooled, and normalized cDNA was synthesised which remarkably reduces the frequency of abundant transcripts and increases the rate recovery of unique transcripts[23]. Upon sequencing and *de novo* assembly, we selected a total of 92% of sequenced raw reads by stringent filtering to annotate into functional transcripts belonging to crucial metabolic pathways of leaf, flower and seed.

The average length of Pongamia unigenes (937.58 bp) was more than those of reported in other related species like chickpea (523 bp), peanut (619 bp), alfalfa (803 bp), as well as in a recent report on Pongamia (787 bp) but shorter than those of *Camelina* (1198 bp)[15,24–26]. The GC content (ratio of guanine and cytosine) which ranges from 20 to 72% among different organisms, is an important criterion for establishing the phylogenetic and evolutionary relationships among various species. Our analysis revealed that the average GC content of Pongamia transcripts (48%) was little lower than the *C. sativa* (49%) and higher than *J. curcas* (43%) and in other Pongamia report (44.77%) which explains the complexity and diversity of the transcriptome sequencing[15,21]. The parameters like mean length of unigenes, GC content, N50 value of present data show the increased coverage and depth of the sequencing. Annotation of transcripts to UniProt resulted in identification of putative proteins which corresponded to various metabolic pathways. In the current study, we outlined SSR markers identified in Pongamia, which act as important resource in gene mapping and marker assisted molecular breeding. More emphasis on Pongamia EST-SSR markers development, characterization and validation was given in a recent study[21]. Certain transcription factors identified in the current study were reported to play an important role in regulation of gene expression in various metabolic and signalling pathways like fatty acid biosynthesis (MYB, PLATZ), elongation (MYB-related, Bhlh), palmitoleate biosynthesis (MYB), oleate biosynthesis (bHLH, GRAS) stearate biosynthesis (G2-like) and fatty acid degradation (PHD, CCAAT)[15]. Interestingly, MYB and MYB related transcription factors which were deciphered in this study are involved in regulation of circadian rhythms and flowering.

Flowering, which is regulated by circadian rhythms, determines production of seeds and yield of the plant. Circadian clock genes integrate the environmental signals required for flowering and also help in adaptation of plants to different geographical locations[4]. It is of great significance to know the sequence information of genes involved in circadian rhythms and floral transitions to understand and alter the flowering cycle in Pongamia. Recently, Winarto *et al.*[4], reported four circadian clock genes (ELF4, LCL1, PRR7, AND TOC1) in Pongamia which are key regulators of central oscillator and showed that they were under diurnal regulation. Here, we reported 25 other crucial genes including TFs whose sequence information is not available for Pongamia in the public databases. The clock gene ELF3 is known to form an evening complex (EC) with ELF4 and LUX thus generating circadian rhythms and hence regulate output pathways such as flowering[27–30]. The peak expression of ELF3 during dawn in leaves of Pongamia is in accordance with the ELF4 expression observed in a previous study[4]. PRR1, PRR3, PRR5, PRR7 and PRR9 are members of PRR gene family and have important roles in the central oscillator[31]. The presence of highly conserved PR and CCT domains in the putative Pongamia PRR proteins implies a similar role to that of *G. soja* and *Arabidopsis* by repressing LCL1 expression in the central oscillator[32,33]. LHY and MYB75 are MYB-like transcription factors that play pivotal roles in the morning loop of the central oscillator. These transcription factors belong to the REVEILLE (RVE) family which consists of 11 proteins with conserved MYB-like domain[34]. Since the MYB domain is known for DNA-binding, these transcription factors could play an important role in the DNA-binding activity of Pongamia LCL1. In Pongamia, MYB was actively expressed in all stages of inflorescence development which could be attributed to the anthocyanin metabolism that gives the characteristic colour to the flowers. GI-CO-FT-APT model of signal cascade for floral initiation under long day conditions was well established in *Arabidopsis*[35]. GI protein represses the CYCLING DOF FACTOR 1 (CDFs) thereby allowing the expression of CO protein during late day which eventually activates FT expression[5]. The peak expression of Pongamia GI and FT in the evening as observed in Arabidopsis corroborated the fact that the photoperiod was an ancient and conserved pathway for controlling flowering. Implication of CO-FT module in the control of photoperiodic flowering has also been described in garden pea, sugar beet and woody species such as poplar where this regulatory module has been proposed to mediate other photoperiodic responses such as growth cessation and bud set[36–38].

Positional cloning and mutation studies on clock genes provided substantial evidence for the role of transcripts showing circadian rhythms in regulating the grain yield, grain weight, number of grains per panicle and flowering time in many cereals[39–41]. Also, in legumes, significant number of transcripts including genes involved in protein, fatty acid synthesis, lipid metabolism and photosynthesis are showing circadian rhythms suggesting the potential roles of circadian clock in flower opening, nectar secretion, seed composition and development[42,43]. Pongamia, which is an outcrossing species, through insect - mediated pollination, starts flowering after 3 to 4 years and seed maturation takes about 10 months after flowering. The information provided in this study about circadian clock genes will provide substantial basis for the studies related to modulation/manipulation of flowering time to get shorter vegetative period and prolonged reproductive stage that leads to the extended period of seed production.

Pongamia is believed to contribute to biodiesel production through its ability to biosynthesize and accumulate considerable amounts of unsaturated triacylglycerols (TAGs) in seeds. In this study, the transcripts involved in lipid metabolism were annotated and further analysed to understand oil accumulation and degradation in the seeds of Pongamia which are of great interest for biofuel production. Pongamia takes 9–10 months to form a mature pod after fertilization of the flower. The initial pod and seed development are at low pace with negligible oil content and poor development of the cotyledons. At 175 DAF, the cotyledon development and oil biosynthesis go at constant pace till maturity (300 DAF). Many other Pongamia accessions belonging to different geographical locations had shown similar patterns of oil accumulation during seed development[20,44]. During FA biosynthesis, plastidial acetyl CoA and malonyl CoA are converted into long-chain acyl-ACP by a series of reactions involving certain enzymes with ACP as a cofactor. Carboxylation of acetyl-CoA to malonyl-CoA is the first committed step in FA synthesis which is catalysed by a multi-subunit acetyl-CoA carboxylase (ACCase) complex and in turn limits the oil accumulation in the seeds. Our data represented all four subunits of ACCase: alpha carboxyltransferase (CTA), beta carboxyl transferase (CTB) and biotin carboxylase (BC) and also a homomeric isoform. The transcript for homomeric isoform was absent in previous reports on Pongamia, Jatropha and peanut. qPCR analysis showed that these three genes exhibited a coordinated and stable expression pattern throughout the

seed development which is in consistent with previous reports on *Arabidopsis*, *B.napus and R.communis*[45,46]. The subsequent formation of plastidial malonyl ACP from malonyl CoA that is catalysed by malonyl-ACP-transferase (MAT) showed maximum expression at stage 2 and stage 3 and decreased thereafter towards the end of the seed development. The activity of ketoacyl-ACP reductase (KAR), which is a component of fatty acyl synthase (FAS) multiprotein complex, is essential for FA biosynthesis and catalyses an NADPH-dependent reduction of 3-ketoacyl-ACP to the 3-hydroxyacyl isomer. Another key enzyme, enoyl-ACP-reductase (EAR) plays a determinant role in establishing the rate of FA biosynthesis[47]. KAR, EAR together with HAD and KAS-II showed a coordinated expression pattern wherein the genes were up regulated at all stages of seed development but showed a downtrend during maturation of the seed (Fig. 6). Similar type of bell shaped expression pattern of FAS genes was also observed in Jatropha seeds[48]. The enzymes SAD, FAD6 and FAD8 biosynthesise oleic acid, linoleic acid and linolenic acid respectively and are crucial for an ideal biofuel feedstock. PpFAD8 is the most abundant transcript represented in our transcriptome data followed by SAD when compared to other lipid biosynthetic enzymes, suggesting the unsaturated FA synthesis potency of Pongamia. However, the expression levels for SAD during seed development were higher than any other enzyme involved in FA synthesis which could be attributed to the low catalytic efficiency of SAD associated with high oleic acid content in Pongamia[45]. Further studies are needed to understand the gene regulation at promoter level and functional characterization of the PpFAD8 protein which provide important clues about oil accumulation patterns in Pongamia seeds. In addition to PpFAD8, other enzymes involved in biosynthesis of unsaturated fatty acids including PpFAD2, PpFAD6 and SAD could be the potential targets for gene engineering to improve oil quality and quantity in Pongamia, where the sequence information can be deduced through our transcriptome data.

The transcripts that encode two acyl-ACP thioesterases that terminate plastid FA synthesis, FATA (responsible for unsaturated FA production) and FATB (for saturated FA production) showed varied expression patterns during the four stages (Fig. 6). The expression of FATA increased significantly from stage 2 to 4 while, FATB decreased after stage 2. This is in agreement with greater plastid production of unsaturated than saturated FAs in Pongamia seeds. However, in Jatropha seeds, FATA expression was at its peak during late developmental stages. Palmitic acid and stearic acid, which are major constituents of cell membrane, also play important role in development of cotyledons which are usually active during 120–210 DAF in case of Pongamia. Our data clearly indicated that the expression of FA biosynthetic genes for saturated FAs followed a typical bell shaped pattern where it increased during stage 1 and was stable during 2nd and 3rd stages which decreased slightly at stage 4 of seed development. Towards the maturity, more unsaturated FAs were synthesized as evidenced from our data on FATB and FATA expression levels and also supported by previous findings on oil content in Pongamia at various seed developmental stages[20]. The free FAs generated by thioesterases in the plastid are esterified to CoA by long-chain acyl-CoA synthetases (LACS) at the plastid envelope. PpLACS showed consistently high level of expression during all the four stages when compared to 150 DAF, suggesting that the oil accumulation was accelerated at 150 DAF. After FA synthesis, a series of membrane-associated reactions assemble the acyl chains into TAG. Glycerol-3-phosphate acyltransferases (GPAT) catalyse sn-1 acylation of glycerol-3-phosphate to yield lysophosphatidic acid (LPA). The second acylation in *de novo* TAG assembly is catalysed by LPA acyltransferase. In Pongamia, the genes involved in TAG assembly including GPAT, LPAT and PAP were mostly expressed during stage 2–4, wherein the maximum oil accumulation has been recorded in our study. However, considerable expression was also noticed during stage I which should account for the membrane lipid biosynthesis during cotyledon development. The final step in TAG biosynthesis is the acylation of diacylglycerol (DAG) to form TAG. Depending on the acyl donor to DAG, two classes of enzymes, diacylglycerol acyltransferases (DGAT) and phospholipid:diacylglycerol acyltransferases (PDAT), can catalyse this crucial step of TAG synthesis[49]. Our results on DGAT and PDAT expression patterns also demonstrate the active involvement of DGAT in Pongamia TAG assembly. Further, the expression patterns of genes involved in β- oxidation revealed that the fatty acid degradation in Pongamia seeds was active during early stages of seed development which could presumptively be responsible for cotyledon development (Fig. 6).

In conclusion, the transcriptome of *Pongamia pinnata* seed along with leaf, pod as well as flower tissues was sequenced and assembled, to maximize the gene representation associated with flowering and lipid biosynthesis. Our data have led to the identification of transcripts, transcription factors involved in various physiological processes and metabolic pathways, which will provide ample information to the database on Pongamia and also aid in the functional and comparative genomic studies to improve oil and seed yield related traits. GI-CO-FT signalling cascade was found to be active in regulating photoperiod control of flowering in Pongamia. The expression patterns of lipid biosynthetic genes at different developmental stages revealed ACCase, SAD and FAD8 as candidate genes during seed maturity and clearly showed that 270–300 DAF was optimum time for seed harvesting. In summary, our results provide an insight into the complex metabolic pathways and regulatory networks involved in different tissues of the Pongamia.

Methods

Plant material. *P. pinnata* plantation was established in the experimental farm of Tree Oils India Limited (TOIL), Zaheerabad, Medak district, Andhra Pradesh (latitude 17°36′; longitude 77°31′E; 622 m MSL). High quality, disease free seeds of nearly 600 accessions were collected from various regions of India and planted in the farm. After attaining reproductive phase, the plants which did not give flowering for two years were removed from the farm and remaining plants were assessed for their yield potential for three consecutive years. The highest yielding variety (TOIL 1) was selected as the experimental plant for the current study. Leaves, flowers, pods and seeds at 210 DAF were collected and snap froze in aseptic conditions for transcriptome sequencing.

For gene quantification studies, five year old Pongamia accession (TOIL 1) those were actively flowering were selected. Leaf samples at four time points 00:00, 06:00, 12:00, 18:00 hr of a day and flowers at four developmental phases of infloresence were collected and stored at −80 °C until further use. Seeds of four developmental stages were collected from same accession of Pongamia to quantify lipid biosynthetic genes.

Transcriptome sequencing. Total RNA was isolated from leaves, flowers, pods and seeds of Pongamia using Agilent plant RNA isolation kit (Agilent Technologies, USA). The concentration, intactness and purity of RNA were checked with Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Samples having RNA integrity number (RIN) value greater than 8 were used for library preparation. Paired-end cDNA library preparation was done according to Illumina TruSeq RNA library protocol outlined in "TruSeq RNA Sample Preparation Guide". Briefly, 1 μg of total RNA was subjected to Poly A purification of mRNA. Purified mRNA was fragmented for 4 minutes at 94 °C in the presence of divalent cations and reverse transcribed with Superscript III Reverse transcriptase by priming with Random Hexamers (Invitrogen, USA). Second strand cDNA was synthesized in the presence of RNA Polymerase I and RnaseH. The cDNA was cleaned up using Agencount Ampure XP SPRI beads (Beckman Coulter, USA). Illumina adapters were ligated to the cDNA molecules after end repair and the addition of A base. SPRI clean – up was performed after ligation. The library was amplified using 8 cycles of PCR for the enrichment of adapter - ligated fragments. The prepared library was quantified using Nanodrop and validated for quality by running an aliquot on High Sensitivity Bioanalyzer Chip (Agilent Technologies, USA). Sequencing of constructed cDNA library was performed on Illumina NextSeq 500 sequencer. RNA-Seq data were generated in FastQ format.

Transcriptome assembly, annotation and analysis. Sequencing resulted in the generation of 76 nucleotide raw reads having attached adapter sequences. These raw reads were subjected to filtering through the standard Illumina pipeline. The filtered reads were further subjected for quality control using NGS QC tool kit V 2.3.1 to remove adapters, B-block and low quality bases towards 3′ ends[50]. The high quality filtered reads were *de novo* assembled by Velvet 1.2.10 and Oases 0.2.08 was used for transcript generation[51,52]. Genes/proteins were predicted from assembled transcripts using GeneMark software[53].

MEGA7 was used for the construction of phylogenetic tree using Clustal W and neighbour-joining analysis by taking the known amino acid sequences of all targeted genes and deduced amino acid sequence of Pongamia[54].

Pathway analysis and identification of transcription factors. After assembly and clustering, transcript annotation was done by performing BLASTX analysis at an e-value cut-off of 10^{-5} against UniProt-Papilionoideae database[55]. Blast2GO was used to assign GO (Gene Ontology) terms to transcripts on the basis of best significant match with proteins of members of Papilionoideae to impart a broad overview of their functions and categorized into biological process, molecular function and cellular component. Also, KOG (Eukaryotic Orthologous Groups) was used to identify the transcript homologues from other organisms and thus assigning a probable function to transcripts. KAAS (KEGG (Kyoto Encyclopedia of Genes and Genomes) Automatic Annotation Server) was used for metabolic pathway analysis using *Arabidopsis thaliana*, *Arabidopsis lyrata* and *Glycine max* as reference organisms to identify the enriched metabolic pathways in various gene sets[56]. The transcripts were categorized into various transcription factors (TFs) using Transcription factor Family Data Base (TFDB)[57].

SSR marker identification. The percentage compositions of the nucleotides A, T, G and C were calculated for each sequence and across the entire distribution of transcripts. Simple Sequence Repeats (SSRs) were detected using MIcroSAtellite tool. SSRs were detected by considering 100 bp flanking sequences on upstream and downstream of SSRs.

Real-time PCR analysis. Seeds, leaves, roots and flower tissues were collected in triplicates and total RNA was isolated using Spectrum Plant Total RNA isolation kit (Sigma, USA). 1 μg of RNA was used for cDNA synthesis by Revert aid first strand cDNA synthesis kit (Thermo-Fisher Scientific, USA). qRT – PCR was performed on Eppendorf thermal cycler using SYBR FAST qPCR universal master mix (2X) (KAPA Biosystems, USA). Each reaction contained 1 μl of the first-strand cDNA as template in a total volume of 10 μl reaction mixture. List of genes, primer sequences and melting temperatures used in this study were given in Supplementary Table S4 and S5. The amplification program was performed at 95 °C for 30 s followed by 95 °C for 5 s and 55 °C for 30 s (35 cycles). The relative expression was calculated using the formula, $2^{-\Delta\Delta Ct}$, with actin as housekeeping gene for normalisation of data[58]. The fold change values were log transformed with base 2 so that 1.5 fold which corresponds to 0.58 was used to identify differentially expressed genes.

Quantification of oil. Oil was extracted from four growing stages of Pongamia seeds by soxhlet extraction method using hexane as a solvent as described in Kumar *et al.*[59]. Briefly, seeds (5 g) were ground in a coffee grinder to make powder. Oil was extracted with 150 ml of hexane at distillation temperature for 2 to 3 hours in the Soxhlet extractor using a heating mantle. Hexane was removed from the extracted oil using a rotary evaporator (Heidolph 514-01002-06-0, Germany) at 55 °C under reduced pressure for 30 min.

Statistics. For qRT PCR analysis, three independent biological replicates with three technical replicates of each were used and the mean ± standard deviation (SD) values were calculated for each sample. The significance of the difference was tested by using Analysis of Variance (ANOVA) and the comparisons were tested with Holm-Sidak method, the level of significance was set to 0.05. Microsoft excel 2013 was used for data processing. Statistical analysis was performed using software, Sigma plat 11.0.

References

1. Arote, S. & Yeole, P. Pongamia pinnata L: a comprehensive review. *Int. J. Pharm. Tech. Res.* **2,** 2283–2290 (2010).
2. Abou-Elwafa, S. F. *et al.* Conservation and divergence of autonomous pathway genes in the flowering regulatory network of Beta vulgaris. *J. Exp. Bot.* **62,** 3359–3374 (2011).
3. Coupland, G. Regulation of flowering by photoperiod in Arabidopsis. *Plant, Cell Environ.* **20,** 785–789 (1997).

4. Winarto, H. P. *et al.* Isolation and Characterization of Circadian Clock Genes in the Biofuel Plant Pongamia (Millettia pinnata). *Bioenergy Res*. **8,** 760–774 (2015).
5. Jarillo, J. A. & Piñeiro, M. Timing is everything in plant development. The central role of floral repressors. *Plant Sci*. **181,** 364–378 (2011).
6. Baud, S., Dubreucq, B., Miquel, M., Rochat, C. & Lepiniec, L. Storage reserve accumulation in Arabidopsis: metabolic and developmental control of seed filling. *The Arabidopsis Book* **6,** e0113 (2008).
7. Bonaventure, G., Salas, J. J., Pollard, M. R. & Ohlrogge, J. B. Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell* **15,** 1020–1033 (2003).
8. Kachroo, A. *et al.* Oleic acid levels regulated by glycerolipid metabolism modulate defense gene expression in Arabidopsis. *Proc.Natl. Acad. Sci. USA* **101,** 5152–5157 (2004).
9. Zheng, H., Rowland, O. & Kunst, L. Disruptions of the Arabidopsis enoyl-CoA reductase gene reveal an essential role for very-long-chain fatty acid synthesis in cell expansion during plant morphogenesis. *Plant Cell* **17,** 1467–1481 (2005).
10. Petrie, J. R. *et al.* Metabolic engineering plant seeds with fish oil-like levels of DHA. *PloS One* **7,** e49165 (2012).
11. Kang, J., Snapp, A. R. & Lu, C. Identification of three genes encoding microsomal oleate desaturases (FAD2) from the oilseed crop Camelina sativa. *Plant Physiol. Biochem*. **49,** 223–229 (2011).
12. Kinney, A. & Clemente, T. Modifying soybean oil for enhanced performance in biodiesel blends. *Fuel Process. Technol*. **86,** 1137–1147 (2005).
13. Meyer, E., Logan, T. L. & Juenger, T. E. Transcriptome analysis and gene expression atlas for Panicum hallii var. filipes, a diploid model for biofuel research. *Plant J*. **70,** 879–890 (2012).
14. Fan, Z. *et al.* Genome-wide transcriptome profiling provides insights into floral bud development of summer-flowering Camellia azalea. *Sci Rep* **5,** 9729 (2015).
15. Mudalkar, S., Golla, R., Ghatty, S. & Reddy, A. R. De novo transcriptome analysis of an imminent biofuel crop, Camelina sativa L. using Illumina GAIIX sequencing platform and identification of SSR markers. *Plant Mol. Biol*. **84,** 159–171 (2014).
16. Wang, F. *et al.* Mining and identification of polyunsaturated fatty acid synthesis genes active during camelina seed development using 454 pyrosequencing. *BMC Plant Biol*. **15,** 147 (2015).
17. Xu, C. *et al.* De novo and comparative transcriptome analysis of cultivated and wild spinach. *Sci Rep*. **5,** 17706 (2015).
18. Huang, J. *et al.* Transcriptome characterization and sequencing-based identification of salt-responsive genes in Millettia pinnata, a semi-mangrove plant. *DNA Res*. **19,** 195–207 (2012).
19. Kazakoff, S. H. *et al.* Capturing the biofuel wellhead and powerhouse: the chloroplast and mitochondrial genomes of the leguminous feedstock tree Pongamia pinnata. *PloS One* **7,** e51687 (2012).
20. Pavithra, H., Gowda, B., Kumar, K. R., Prasanna, K. & Shivanna, M. Oil, fatty acid profile and karanjin content in developing Pongamia pinnata (L.) Pierre seeds. *J. Am. Oil Chem. Soc*. **89,** 2237–2244 (2012).
21. Huang, J. *et al.* De novo sequencing and characterization of seed transcriptome of the tree legume Millettia pinnata for gene discovery and SSR marker development. *Mol. Breed*. **36,** 1–15 (2016).
22. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC bioinformatics* **4,** 1 (2003).
23. Natarajan, P. & Parani, M. De novo assembly and transcriptome analysis of five major tissues of Jatropha curcas L. using GS FLX titanium platform of 454 pyrosequencing. *BMC genomics* **12,** 1 (2011).
24. Garg, R., Patel, R. K., Tyagi, A. K. & Jain, M. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res*. **18,** 53–63 (2011).
25. Liu, Z. *et al.* Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. *PloS One* **8,** e83549 (2013).
26. Zhang, J. *et al.* De novo assembly and Characterisation of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (Arachis hypogaea L.). *BMC Genomics* **13,** 90 (2012).
27. Doyle, M. R. *et al.* The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. *Nature* **419,** 74–77 (2002).
28. Khanna, R., Kikis, E. A. & Quail, P. H. EARLY FLOWERING 4 functions in phytochrome B-regulated seedling de-etiolation. *Plant Physiol*. **133,** 1530–1538 (2003).
29. Nozue, K. *et al.* Rhythmic growth explained by coincidence between internal and external cues. *Nature* **448,** 358–361 (2007).
30. Nusinow, D. A. *et al.* The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature* **475,** 398–402 (2011).
31. Matsushika, A., Makino, S., Kojima, M. & Mizuno, T. Circadian waves of expression of the APRR1/TOC1 family of pseudo-response regulators in Arabidopsis thaliana: insight into the plant circadian clock. *Plant Cell Physiol*. **41,** 1002–1012 (2000).
32. Makino, S. *et al.* Genes encoding pseudo-response regulators: insight into His-to-Asp phosphorelay and circadian rhythm in Arabidopsis thaliana. *Plant Cell Physiol*. **41,** 791–803 (2000).
33. Strayer, C. *et al.* Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science* **289,** 768–771 (2000).
34. Carré, I. A. & Kim, J. Y. MYB transcription factors in the Arabidopsis circadian clock. *J. Exp. Bot*. **53,** 1551–1557 (2002).
35. Roux, F., Touzet, P., Cuguen, J. & Le Corre, V. How to be early flowering: an evolutionary perspective. *Trends Plant Sci*. **11,** 375–381 (2006).
36. Chia, T., Müller, A., Jung, C. & Mutasa-Göttgens, E. Sugar beet contains a large CONSTANS-LIKE gene family including a CO homologue that is independent of the early-bolting (B) gene locus. *J. Exp. Bot*. **59,** 2735–2748 (2008).
37. Horvath, D. Common mechanisms regulate flowering and dormancy. *Plant sci*. **177,** 523–531 (2009).
38. Weller, J. L. *et al.* Update on the genetic control of flowering in garden pea. *J. Exp. Bot*. **60,** 2493–2499 (2009).
39. Beales, J., Turner, A., Griffiths, S., Snape, J. W. & Laurie, D. A. A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). *Theor. Appl. Genet*. **115,** 721–733 (2007).
40. Shaw, L. M., Turner, A. S. & Laurie, D. A. The impact of photoperiod insensitive Ppd-1a mutations on the photoperiod pathway across the three genomes of hexaploid wheat (Triticum aestivum). *Plant J*. **71,** 71–84 (2012).
41. Turner, A., Beales, J., Faure, S., Dunford, R. P. & Laurie, D. A. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* **310,** 1031–1034 (2005).
42. Hudson, K. A. The circadian clock-controlled transcriptome of developing soybean seeds. *The Plant Genome* **3,** 3–13 (2010).
43. Preuss, S. B. *et al.* Expression of the Arabidopsis thaliana BBX32 gene in soybean increases grain yield. *PloS One* **7,** e30717 (2012).
44. Pavithra, H., Sagar, B. C., Prasanna, K., Shivanna, M. & Gowda, B. Localisation of Storage Reserves in Developing Seeds of Pongamia pinnata (L.) Pierre, a Potential Agroforestry Tree. *J. Am. Oil Chem. Soc*. **90,** 1927–1935 (2013).
45. Troncoso-Ponce, M. A. *et al.* Comparative deep transcriptional profiling of four developing oilseeds. *Plant J*. **68,** 1014–1027 (2011).
46. Baud, S. & Lepiniec, L. Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. *Plant Physiol.Biochem*. **47,** 448–455 (2009).
47. Niu, Y. *et al.* Global analysis of gene expression profiles in Brassica napus developing seeds reveals a conserved lipid metabolism regulation with Arabidopsis thaliana. *Molecular plant* **2,** 1107–1122 (2009).
48. Gu, K. *et al.* Expression of fatty acid and lipid biosynthetic genes in developing endosperm of Jatropha curcas. *Biotechnol. Biofuels* **5,** 47 (2012).
49. Chaitanya, B. S. K. *et al.* Stage-Specific Fatty Acid Fluxes Play a Regulatory Role in Glycerolipid Metabolism during Seed Development in Jatropha curcas L. *J. Agric. Food Chem*. **63,** 10811–10821 (2015).

50. Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS One* **7,** e30619 (2012).
51. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28,** 1086–1092 (2012).
52. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. **18,** 821–829 (2008).
53. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res*. **42,** e119–e119 (2014).
54. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol Evol*. **33,** 1870–1874 (2016).
55. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10,** 421 (2009).
56. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. **35,** W182–W185 (2007).
57. Jin, J., Zhang, H., Kong, L., Gao, G. & Luo, J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res*. **42** D1182–D1187 (2013).
58. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods* **25,** 402–408 (2001).
59. Kumar, S., Chaitanya, B. S., Ghatty, S. & Reddy, A. R. Growth, reproductive phenology and yield responses of a potential biofuel plant, Jatropha curcas grown under projected 2050 levels of elevated CO2. *Physiol. Plant*. **152,** 501–519 (2014).

Acknowledgements

The work was funded by DBT grant (BT/PR-12024/BCE/8/1097/2014) from Department of Biotechnology, Government of India. We greatly acknowledge Tree Oils India Limited (TOIL), Hyderabad, India for generously providing plant material for all our experiments. Thanks are due to Sandor Technologies, Hyderabad, India for library construction, sequencing and assembly. RVS and SM are thankful to UGC, New Delhi, India, for fellowship. KTS was supported by BBL fellowship, University of Hyderabad, India.

Author Contributions

R.V.S., S.M. and A.R.R. are involved in designing experiments. R.V.S., S.M. and K.T.S. performed experiments. R.V.S., S.M., K.T.S. and A.R.R. analysed data and discussed results. R.V.S., S.M. and A.R.R. wrote the manuscript. All authors revised the manuscript.

Additional Information

Accession codes: The data has been submitted to NCBI Sequence Read Archive (SRA) and BioSample databases with BioSample accession number SAMN04212410 and BioProject Id PRJNA299718. The SRA project Id is SRP065225.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sreeharsha, R. V. *et al.* Unravelling molecular mechanisms from floral initiation to lipid biosynthesis in a promising biofuel tree species, *Pongamia pinnata* using transcriptome analysis. *Sci. Rep.* **6**, 34315; doi: 10.1038/srep34315 (2016).

Oleochemical changes during seed development in oil yielding legume tree, Pongamia pinnata L. Pierre

by K Tamna Singha

Submission date: 27-Jul-2023 04:01PM (UTC+0530)
Submission ID: 2137520668
File name: Tamna_thesis_final_plagiarism_check.pdf (384.34K)
Word count: 18662
Character count: 98141

Oleochemical changes during seed development in oil yielding legume tree, Pongamia pinnata L. Pierre

ORIGINALITY REPORT

36%	35%	35%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	coek.info Internet Source	21%
2	Kambam Tamna Singha, Rachapudi Venkata Sreeharsha, Sureshbabu Mariboina, Attipalli Ramachandra Reddy. "Dynamics of metabolites and key regulatory proteins in the developing seeds of Pongamia pinnata, a potential biofuel tree species", Industrial Crops and Products, 2019 Publication	10%
3	link.springer.com Internet Source	1%
4	Submitted to University of Hyderabad, Hyderabad Student Paper	1%
5	Xiaoguang Shang, Yulei Zhu, Xiang Chen, Xin-Ding Wang, Ray J. Rose, Youhong Song. "Seed oil storage in three contrasted legume species: implications for oil improvement", Acta Physiologiae Plantarum, 2020	<1%

	Publication	
6	plantbiologyconference.com Internet Source	<1%
7	Lionel M. Hill, Edward R. Morley-Smith, Stephen Rawsthorne. "Metabolism of Sugars in the Endosperm of Developing Seeds of Oilseed Rape", Plant Physiology, 2003 Publication	<1%
8	repositorio.unipampa.edu.br Internet Source	<1%
9	www.frontiersin.org Internet Source	<1%
10	J. Kreuzwieser. "Impact of waterlogging on the N-metabolism of flood tolerant and non-tolerant tree species", Plant Cell and Environment, 8/2002 Publication	<1%
11	kodomo.cmm.msu.ru Internet Source	<1%
12	hdl.handle.net Internet Source	<1%
13	worldwidescience.org Internet Source	<1%
14	www.ebi.ac.uk Internet Source	<1%

Prof. S. RAJAGOPAL
Dept. of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

15	www.researchgate.net Internet Source	<1 %
16	P.K. Mohapatra, R.K. Sarkar, S.R. Kuanar. "Starch synthesizing enzymes and sink strength of grains of contrasting rice cultivars", Plant Science, 2009 Publication	<1 %
17	Sumit Kumar, Bharatula S.K. Chaitanya, Sreenivas Ghatty, Attipalli R. Reddy. " Growth, reproductive phenology and yield responses of a potential biofuel plant, grown under projected 2050 levels of elevated CO ", Physiologia Plantarum, 2014 Publication	<1 %
18	edepot.wur.nl Internet Source	<1 %
19	Zsófia Bánfalvi, Attila Molnár, Lóránt Lakatos, Holger Hesse, Rainer Höfgen. "Differences in sucrose-to-starch metabolism of Solanum tuberosum and Solanum brevidens", Plant Science, 1999 Publication	<1 %
20	www.mdpi.com Internet Source	<1 %
21	Submitted to University of Sheffield Student Paper	<1 %

Prof. S. RAJAGOPAL
Dept. of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

22	www.agri.gov.il Internet Source	<1%
23	academicjournals.org Internet Source	<1%

Exclude quotes On

Exclude bibliography On

Exclude matches < 14 words

This is to certify that 31% similarity is from student's own publication which includes 21% from S.No. 1 and 10% from S.No. 2 which leaves 5% similarity overall.

Prof. S. RAJAGOPAL
Dept. of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.