COMPARATIVE PROTEOMIC STUDIES OF SPROUTED LENTIL

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Plant Sciences
University of Saskatchewan
Saskatoon

By

SUDIPTA DUTTA

Permission to use

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences
College of Agriculture and Bioresources
University of Saskatchewan
51 Campus Drive
Saskatoon, Saskatchewan S7N 5A8 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

A major portion of lentil nutritional composition consists of protein and protein-derived biomolecules. Seed germination improves the protein quality. It is important to identify at what stage of seed germination the type and quality of protein improvement occurs. This study aimed to elucidate the molecular mechanism of seed germination in terms of proteomic changes that occur during distinct stages of seed germination and post germination in seven lentil species.

In this study of six lentil species, four seed germination phases were defined, including control, for dry seed and one post-germination phase was defined Phases I, II, and III occur during the seed germination, and phase IV is the post-germination phase. The genotype IG 72815 had the smallest seed size among the genotypes studied and CDC Redberry had the largest seed size. Imbibition period was the longest for CDC Redberry, while its period of seed germination was the least. For genotypes with smaller seed size the imbibition period was longer.

The protease activity of all eight genotypes representing all lentil species began to increase during phase III, except for BGE 016880 in which protease activity started increasing in phase IV. For IG 72805 and IG 72815, the protease activity was the highest overall and then reduced in phase IV. The total essential and non-essential free amino acids released was the highest in phase IV. Among the genotypes studied, IG 110813, and IG 72623 had the highest amount of essential free amino acids (FEAAs) released during the germination phase. Among the FEAAs released, the amount of histidine was the highest followed by threonine. In phase IV, a significant increase was observed in theF EAAs isoleucine (ILE), lysine (LYS), leucine (LEU), valine (VAL), and threonine (THR) in germination phase IV.

Carbohydrate degradation was initiated from the beginning of germination as many of the enzymes were present in control seeds, and some of the enzymes were absent in phase IV, presumably due to degradation. According to this study, phase III, most of the enzymes required for the tricarboxylic acid (TCA) cycle were present, which indicates the TCA became fully functional in phase III. Protein metabolism and synthesis related to enzymes and other factors was initiated in germination phase II. Neosynthesis of lipid catabolism, lipid biosynthesis, amino acid biosynthesis, protein modification, and protein translocation related proteins were initiated in phase IV.

Overall, this study found that germination phases III and IV were superior in terms of free amino acid release and protease activity. Based on that observation, it could be concluded that phase III and phase IV were better in terms of digestibility and availability of as the protease activity and released free amino acids were found to be highest in these phases. Based on proteome study it could be concluded that carbohydrate degradation followed by protein metabolism initiates the seed germination phenomenon.

Acknowledgements

I would like to convey heartfelt gratitude to my supervisor Albert Vandenberg for allowing me to pursue the dream to study at the University of Saskatchewan and providing unwavering support throughout the program. I would also like to acknowledge the committee members: Kirstin Bett, Cristina Martínez-Villaluenga, and Haixia Zhang for providing their advice, pragmatic comments and suggestions.

Conducting research experiments and projects requires financial support, which was provided by the Saskatchewan Pulse Growers NSERC Industrial Chair in Genetic Improvement of Lentil and by BASF. In addition, the HE Martynse Postgraduate Scholarship and Rene Vandeveld Postgraduate Scholarship also encouraged me to fully focus on academic progression.

A special thanks goes to Randy Purves from the Canadian Food Inspection Agency for allowing me to work and providing suppor throughout the experiments, and to Sabine Banniza, Sean Prager, and Ravindra Chibbar from the U of S for allowing access to laboratory instruments despite unavoidable circumstances created by the pandemic. My heartfelt gratitude also extends to Cristina Martínez-Villaluenga's team at Institute of Food Science, Technology and Nutrition (ICTAN-CSIC) in Madrid, Spain, and to Haixia Zhang for their support to accomplish the major portion of her experiments. I would also like to extend my thanks to Devini De Silva, Brent Barlow, Sarita Jaiswal, Alison Sackville, Kevin Mikituk, Carmen Breitkreutz, Barry Goetz, Rozer Munro, Robert Stonehouse, and Monica Baga for providing their help with experimental setup, scheduling laboratory work, and troubleshooting.

Apart from that, I am thankful to Dr. Rajib Podder and Erin Daniels Jones for their guidance whenever it was requested. The author would also like to thank Tom Warkentin and Randy Kutcher for their help whenever required in program related activities.

I would also like to thank my parents, teachers, friends, and relatives who motivated and guided me towards the right path. I would also like to thank friends from the department – Stanley Adobor, Ishita Patel, Jeremy Marshall, Matthew Wengler, Junsheng Zhou, Tadesse Gela, and Fatma Elessawy who encouraged me throughout, and helped me in stress management with their advice.

Table of Contents

List of Tables

List of Figures

List of appendices

List of Abbreviations

11S	Large subunit 11
7S	Small subunit 7
ABA	Abscisic acid
ABC	Ammonium bicarbonate
ABR	ABA-responsive protein
ACE	Angiotensin-converting enzyme
ACN	Acetonitrile
ACP	Acyl carrier protein
ACX1	Acyl-CoA coa oxidase
ADH	Alcohol dehydrogenase
ADP	Adenosine diphosphate
AGC	Automatic Gain Control
AHL	N-Acyl homoserine lactone
AKR2	Ankyrin repeat-containing protein 2
ALA	Alanine
ANFs	Antinutritional factors
ANOVA	Analysis of variance
APC	Amino acid polyamine organocation
APT	Adenine phosphoribosyltransferase
APX	Ascorbate peroxidase
ARF	ADP-ribosylation factor
ARG	Arginine
ARP	Apurinic redox protein
ARX1	Associated with ribosomal export complex protein 1
ASN	Asparagine
ASP	Aspartic acid
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BRE	Before radicle emergence

BXL	Beta-D-xylosidase
Ca	Calcium
CAMK	Ca2+/calmodulin-dependent
CAT2	Catalase2
CCR	C-C chemokine recepto
CCV	Clathrin coated vesicle
CDC	Crop Development Centre
CDPK	Calcium-dependent protein kinases
CHS	Chalcone synthase
CIA	Cytosolic iron-sulfur cluster
COG0354m	Plastidial Clusters of Orthologous Groups 0354 protein
COP	Constitutive photomorphogenesis protein
CORVET	Class C core vacuole/endosome tethering
CPD photolyase	Cyclobutane pyrimidine dimer
CSC	Cellulose synthase complex
CSLC	Cellulose synthase-like protein
Cu	Copper
CYS	Cysteine
DAD	Diode array detection
DALL5	Triacylglycerol lipase
DAP	Aspartyl aminopeptidase
DDA	Data dependent acquisition
DFE	Dietary folate equivalents
DHAR	Dehydroascorbate reductase
DHM	Dihydromyricetin
DNA	Deoxyribonucleic acid
DPP-IV	Dipeptidyl Peptidase IV
DRRG	Disease resistance response genes
DS	Dry seed
DTT	Dithiothreitol

DW	Dry weight
ECR	Enoyl-CoA reductase
EIF2	Eukaryotic Initiation Factor 2
ESCRT	Endosomal sorting complexes required for transport
ESCRT-mediated	Endosomal sorting complex required for transport sorting
ESI-TRAP	Electrospray ion trap mass spectrometry
FA	Formic acid
FAAs	Free amino acids
FAE	Fatty acid elongase
FAO	Food and Agriculture Organization
Fe	Iron
FFEAAs	FFree essential essential amino acids
FMOC	Fluorenylmethyloxycarbonyl
FNEAAs	Free non-essential amino acids
FQR	Flavodoxin-like quinone reductase
FRA	Protein fragile fiber
FTFAAs	Total Free free total amino acids
GA	Gibberellin
GABA	Gamma-aminobutyric acid
GAPs	GTPase-activating proteins
GDP	Guanosine diphosphate
GLN	Glutamine
GLU	Glutamic acid
GLY	Glycine
GR	Glutathione reductase
GS-GOGAT	Glutamate synthase-Glutamine oxoglutarate aminotransferase
GTP	Guanosine triphosphate
HCl	Hydrochloric acid
HIS	Histidine
HLB	Hydrophilic-Lipophilic-Balanced

HOPS	homotypic fusion and protein sorting
HPPD	Hydroxyphenylpyruvate dioxygenase
HSP	Heat shock proteins
IAM	Iodoacetamide
IBA57	Iron-Sulfur Cluster Assembly Factor 57
ICARDA	International Center for Agricultural Research in the Dry Areas
ILE	Isoleucine
ILITHYIA/GCN1	HEAT repeat protein
ISA3	Starch-debranching isoamylase
ISC	Iron-sulfur cluster
IU	International unit
JA	Jasmonic acid
K	Potassium
KAT5	3-ketoacyl-CoA thiolase
KDa	Kilodaltons
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid-chromatography-Mass Spectrometry
LDH	Lactate dehydrogenase
LEU	Leucine
LSU	Large ribosomal subunit
LYS	Lysine
MAC	MOS4-associated complex
MAT	S-adenosyl methionine synthetase
MDAR	Monodehydroascorbate reductase
MDH	Malate dehydrogenase
MEP	Methylerythritol phosphate
MeV	Mega electron-volt
MFP	Multifunctional enzyme
MFW	Mean fresh weight
Mg	Magnesium

Mn	Manganese
MOS2	MODIFIER of snc1
MPP	Mitochondrial processing peptidase
MPP	M-phase phosphoprotein
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry
MSX	Methionine sulfoximine
Mt	Metric Ton
MUB	Membrane-anchored Ubiquitin-fold
MVBs	Multivesicular bodies
Na	Sodium
NADH	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NBP35	Nucleotide-Binding Protein 35 kDa
NCE	Normalized collision energy
NCS	Nucleobase cation symporter
NFU	NifU-like protein 4
NPC	Nuclear pore complex
NRT1	Nicotinamide riboside transporter 1
NSF	N-ethylmaleimide sensitive fusion proteins
NSP	Non-starch polysaccharides
NUP58	Nuclear pore complex nucleoporin
NVA	Norvaline
OPA	Ortho-phthalaldehydeOxalate-phosphate-amine
OPC	3-oxo-22′[Z]-pentenyl-cyclopentane
OPR3	Oxophytodienoate reductase 3
P	Phosphorus
PDC	Pyruvate decarboxylase
PHE	Phenylalanine
PHOS	Phosphorylated protein

PIC	Pre-initiation complex
PLAT	Polycystin-1, lipoxygenase, alpha-toxin
PLS-DA	Partial least-squares discriminant analysis
PMI2	Plastid movement impaired
PP2A	Protein phosphatase 2
PPP	Pentose phosphate pathway
PRO	Proline
PRONE	Plant-specific Rop nucleotide exchanger
PTR	Phosphonotyrosine
PUR	Phosphoribosyl-aminoimidazole synthetase
PUR5	Purine synthetase
RAB	Ras-related protein
RAB-GDI	Ras-related GTP-binding proteins-GDP dissociation inhibitor
RAB-GTPase	Ypt/Rab-type GTPase ypt71
RAM	Regulation of Ace2p and morphogenesis
RAPD	Random Amplified Polymorphic DNA
RFU	Relative fluorescence units
RISC	RNA-induced silencing complex
RN	Ribonucleotide
RNA	Ribonucleic Acid
ROP	Rho proteins of plants
RopGEF	Guanine nucleotide exchange factor
ROS	Reactive oxygen species
RP-HPLC	Reversed-phase high-performance liquid chromatography
RPT1	Proteasome regulatory ATPase subunit 1
RRF	Ribosome release/recycling factor
RuBisCo	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SAM	S-adenosyl methionine
SDC	Sodium deoxycholate
SDP1	Sugar-Dependent protein 1

SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SER	Serine
SIM	Selected ion monitoring
SLE1	N-acetylmuramoyl-L-alanine amidase
SNARE	Soluble N-ethylmaleimide-sensitive-factor attachment protein receptor
SOD	Superoxide dismutase
SUMO	Small Ubiquitin-like Modifier
TAG	Triacylglycerol
TCA	The tricarboxylic acid
THR	Threonine
Tic110	TIC translocation channel
TRIC	T-complex protein Ring Complex
tRNA	Transfer RNA
TRP	Tryptophan
TYR	Tyrosine
U of S	University of Saskatchewan
UDP	Uracil-diphosphate
VAL	Valine
VPS4/SKD1	Vacuolar protein sorting/suppressor of k+ transport growth
VPS41/VAM2	Vacuolar protein sorting/ Vacuolar protein sorting-associated protein 41
VPS4-VTA1	Vacuolar protein sorting-associated protein VTA1
WEB1-PMI2	Protein weak chloroplast movement under blue light 2
XPO1	Exportin 1
Zn	Zinc

Chapter 1

Introduction

Lentil (*Lens culinaris* L.) is gaining recognition worldwide as a nutritious staple legume, with a protein content of its seeds around 26 % of dry weight. Lentil seeds are the main source of dietary protein for over one billion people and provide opportunities for greater use in new plant-based protein foods (Jarpa-Parra, 2018; Khazaei et al., 2019). Lentils are rich in protein, fibre, vitamins, available carbohydrates as well as essential macro and micro-elements including phosphorous, iron, zinc, magnesium, potassium, sodium, copper, calcium, and manganese (Shahwar et al., 2017). The protein fraction of lentil is mainly composed of globulins and albumins that are rich in essential amino acids, although, like most legumes, with low content of sulfur-containing amino acids (methionine and cysteine) and tryptophan. Lentil is a great source of peptides that after being released from proteins during digestion by may provide health benefits (Marambe & Wanasundara, 2014; Khazaei et al., 2019). Lentil protein fractions are also known to contain antinutritional compounds, including trypsin and Bowman–Birk protease inhibitors that negatively affect protein digestibility, mineral chelation, amino acid and peptide bioavailability (Benincasa et al., 2019; Khazaei et al., 2019).

Methods are available that can improve lentil protein digestibility, amino acid composition, and bioavailability as well as nutritional and functional value of lentil - germination is one of those. Seed germination represents an opportunity to enhance nutritional quality and health benefits of lentil proteins. Through the germination process, some proteins are newly synthesized, some proteins are hydrolyzed into peptides and amino acids, and some enzymes are formed that can serve a wide range of functions from performing metabolic activities to hydrolyzing larger biomolecules (He and Yang, 2013; Galland et al., 2014). In addition, the storage proteins are partially hydrolyzed into peptides and amino acids by proteolytic enzymes, thereby decreasing protease inhibitors and increasing the content and bioavailability of amino acids and bioactive peptides (Benincasa et al., 2019). These biochemical changes lead to the improvement of lentil proteins in terms of digestibility, bioavailability, and in the overall quality.

Genotype is an important factor affecting the nutritional and functional value of sprouted grains. It may be possible to maximize the beneficial effects of germination on protein quality through the study of the genetic diversity of lentil. Lentil species are grouped into primary (*L. culinaris, L. orientalis, L. tomentosus*), secondary (*L. lamottei*, *L. odemensis*), tertiary (*L. ervoides*) and quaternary (*L. nigricans*) gene pools (Wong et al., 2015). The wild lentil species have been studied for their potential as sources of resistance to biotic and abiotic stresses, but relatively little information is available on the variation in the germination pattern and subsequent biochemical and proteome changes, with potential to tailor nutritional quality and health benefits of lentil proteins. The results of this project could be of major importance for improving nutritional quality of proteins through lentil germination.

Overall, this project focuses on comparative aspects of protein characteristics oflentil species. The specific goal is to identify proteins that are differentially expressed among different phases of germination as well as aims to find natural variation in amino acid profiles and protease activity during distinct phases of germination. Ultimately, this research is intended to compare between sprouted and un-sprouted seeds with dry seeds as control at protein level. Therefore, the hypotheses established for the project as following:

Hypothesis 1. Lentil species will show variation in the free amino acid profile and protease activity during germination and post-germination phases when compared to the intact seeds.

Hypothesis 2. Lentil species will show proteomic differences among four phases of seed germination process.

Experiments were carried out based on these hypotheses to achieve the research objectives described below:

- To characterize distinct phases of seed germination and post-germination in *Lens* spp.
- To study the influence of genotype on protease activity and free amino acid profile.
- To compare the effects of germination on the proteome profile change among the seven lentil species.

Chapter 2

Review of literature

2.1 Pulses

Pulses are edible legumes that produce nutritious seeds composed mainly of protein and starch. According to Food and Agriculture Organization (FAO), pulses are the edible seeds of leguminous plants cultivated for both food and feed. The most common species and types of temperate pulse crops include *Phaseolus vulgaris* (kidney bean, cannellini bean, navy bean, cranberry bean, black bean, pinto bean, etc.), *Vicia faba* (fava bean), *Cicer arietinum* (chickpea), *Pisum sativum* (pea) and *Lens culinaris* (lentil) (Polak et al., 2015). The worldwide yearly production of pulses is 57 million metric tonnes and of which dry bean is the largest produced crop followed by field pea and lentil (Alberta Pulse Growers, 2022). India, Turkey and Canada are the largest lentil producers, and Canada is the largest global exporter of lentils (Alberta Pulse Growers, 2022). Canada's yearly exports in 2022 were 60000, 63000, 92000, 42000, and 2600 tonnes to Europe, Mediterranean countries, South America, African countries, and the Pacific Rim market, respectively.

2.2 Lentil

Lentil is cultivated worldwide (in more than 70 countries, in three major agro-ecosystems) as a staple source of plant-based protein. In 2020, lentil is listed among the major global pulse crops with 5.41 Mt production (FAOSTAT, 2022). . Like other legume crops, lentil production is also important in sustainable agricultural systems due to its ability to fix biological nitrogen, which promotes soil health and reduces the cost of synthetic fertilizer to some extent (Erskine et al., 2009). Lentil seeds are rich in protein, starch, dietary fiber, and essential micronutrients such as iron, zinc, and many B vitamins (Zhang et al., 2021; Khazaei et al., 2017; Khazaei et al., 2019). Lentil is grown in more than 70 countries globally, and lentil consumption is increasing rapidly in comparison to other major pulses (Khazaei et al., 2019). Lentils are consumed in whole, dehulled, and split forms. Lentil has a wide range of seed coat colors (green, tan, brown, gray, white, and black) and a series of seed coat patterns, classified as marbled, dotted, spotted, complex and un-patterned, of which non-patterned seed coat is most common (Vandenberg and Slinkard, 1990). Lentil cotyledons can be red (most common), yellow, or green (least common), a major factor that

determines marketability based on consumer preference color in different regions around the world.

2.3 Phylogenetic classification of lentil

The *Lens* (2n = 14) genus is in the *Fabaceae* family, sub-family *Papilionoideae*. A study by Ferguson et al. (2000) amended earlier classifications based on crossability and phenetic relations (Van Oss et al., 1997) and corroborated evidence by morphological marker (RAPD and isozyme) approaches that the genus *Lens* consists of seven taxa, where *L. culinaris* Medik. has four subspecies: *culinaris*, *orientalis* (Boiss.) Ponert, *tomentosus* (Ladizinsky) M.E. Ferguson, N. Maxted, M.W. van Slageren and L.D. Robertson, and *odemensis* (Ladizinsky) M.E. Ferguson, N. Maxted, M.W. van Slageren and L.D. Robertson. Three other separate species include *L. ervoides* (Brign.) Grande; *L. nigricans* (M. Bieb.) Godron; and *L. lamottei* Czefr.

Based on taxonomy studies (morphology, cytogenetics, hybridization studies), the genus has seven species, namely *L. culinaris* Medik., *L. orientalis* Boiss, *L. tomentosus* Ladizinsky, *L. odemensis* Ladizinsky, *L. lamottei* Czefr, *L. ervoides* and *L. nigricans* M. Bieb. Based on genomic analyses, the lentil species are grouped into primary, secondary, tertiary and quaternary gene pools (Wong et al., 2015). *L. culinaris*, *L. orientalis* and *L. tomentosus* are grouped as the primary gene pool, where *L. orientalis* is the closest to *L. culinaris. Lens lamottei* and *L. odemensis* form the secondary gene pool; *L. ervoides* is in the tertiary gene pool and finally *L. nigricans* belonged to the quaternary pool. These genetically diverse pools of lentil species determine their variation in nutritional composition and functional properties (Khazaei et al., 2019).

2.4 Nutritional composition of lentil seed

Lentil is a legume crop and a vital source of protein, complex carbohydrates, minerals, vitamins, phenolic compounds as well as dietary fiber with low fat content (Jarpa-Parra, 2018). Total protein content in lentil seeds is on average is 26%, and may vary from 23.4% to 36.4% of total seed weight depending on genotype, environment, and cultural practices (Hulse et al., 2003; Urbano et al., 2007**;** Khazaei et al., 2019). Studies of total protein content in different lentil species revealed the presence of a wider range of protein content (from 18.1% to 32.7% of dry weight) among *L. orientalis*, *L. tomentosus*, *L. odemensis*, *L. ervoides*, and *L. nigricans* compared to the cultivated lentil (Kumar et al., 2016). An earlier study found that the total protein content of wild lentil species

(*L. orientalis*, *L. ervoides*, and *L. nigricans*) was similar to that of the cultivated lentil, ranging between 24.2% and 26.2% (Bhatty, 1986). However, the effect of the environment on the protein content of wild lentil species is not well described, presumably because of the difficulty of producing sufficient quantities of seed. Lentils contain about 60–64.44 g/100g and 30.5 g/100g of dry seed weight carbohydrate and dietary fiber, respectively (Ganesan and Xu, 2017). A study of carbohydrate fractions showed that the amount of reducing sugars, sucrose, raffinose, stachyose, verbascose and starch present in lentil seeds are 0.75, 1.79, 0.40, 1.81, 0.48, and 42.0 g/100g of dry weight, accordingly (Hefnawy, 2011). A study of mineral composition of lentils revealed concentrations (g per 100 g FM) for Fe (iron), Zn (zinc), Cu (copper), Mn (manganese), Ca (calcium), K (potassium), Mg (magnesium), P (phosphorus), and Na (sodium) are around 7.5, 4.8, 0.5, 1.3, 56, 955, 122, and 451, respectively (Al-Islam et al., 2012). This study also evaluates the amount of vitamin C (4.4 mg/100 g), thiamin (0.9 mg/100 g), riboflavin (0.2 mg/100 g), niacin (2.6 mg/100 g), pantothenic acid (2.1 mg/100 g), pyridoxine (0.5 mg/100 g), folate (479 µg, DFE), vitamin A (39 IU), beta carotene (23 µg/100 g), a-tocopherol (0.5 mg/100 g), c-tocopherol (4.2 mg/100 g), phylloquinone (5.0 µg/100 g), and total choline (96.4 mg/100 g) in whole lentils. Thavarajah et al. 2007 showed that the Se (selenium) concentration in lentil is between 2 and 9 µmol per kg. Phenolic compounds in lentil can include epigallocatechin, catechin, epicatechin, epicatechin gallate, quercetin, kaempferol, procyanidin, prodelphinidin, eriodictyo, naringenin, myricetin, acetylsagittatin, and luteolin (Liu et al., 2020).

2.5 Protein quality of lentil

Seeds of legumes contain significant amounts of protein, and therefore can be an economical alternative as meat analogue (Aggarwal and Drewnowski, 2019). A major portion of lentil seed weight is protein. The seed protein of lentil is comprised of storage proteins and metabolic proteins or housekeeping proteins (Khazaei et al., 2019). Around 80% of the total protein content is dominated by storage proteins located in the cotyledon tissue of lentil, and exhibit a low proportion of sulfur-containing amino acids (Jarpa-Parra, 2018). The storage proteins are formed and deposited in the seed cotyledons during their growth and development, and primarily function to form a supply of nitrogen, carbon, and sulfur required for seed germination and early seedling growth and development (Shewry et al., 1995). Metabolic proteins are minor protein constituents that include enzymes and structural proteins. Based on their solubility in different solvents,

proteins can be classified into four categories: salt soluble globulins, water-soluble albumins, acid-soluble glutelins, and ethanol soluble prolamins (Osborne, 1924). Distribution of Osborne protein fractions in lentil is 70% globulins, 16% albumins, 11% glutelins, and 3% prolamins (Boye et al., 2010). Storage proteins in lentils are globulins consisting of several subunits. Both the globulins and albumins of lentils are heterogeneous (Bhatty, 1988; Scippa et al., 2010). Both the globulins and albumins of lentils are heterogeneous, and the highest degree of heterogeneity is in 7S globulins of lentil seeds (Bhatty, 1988; Scippa et al., 2010). And Scippa et al. 2010 found that the 7S/11S ratio of lentil seeds is 2.78, whereas in *Medicago truncatula* and *Pisum sativum* the ratio found was 0.25 and 1, respectively. The most abundant proteins in lentils are globulins, which can be further categorized into two groups based on sedimentation coefficient: legumin (11S) and vicilin (7S). The legumin group is characterized by noncovalent bonds between six polypeptide pairs where each pair is linked by a single disulfide bond between acidic subunit (about 40 kDa) and basic subunit (about 20 kDa) making a total molecular weight of 320–380 kDa (Shewry, 1995). The vicilin 7S proteins are trimers of glycosylated subunits with molecular weight that varies between 50 and 60 kDa (Argos et al., 1985). The ratio of 7S/11S has been considered an indicator of protein nutritional quality (Bourgeois et al., 2009). Albumins consist of 13 polypeptides and are water-soluble with a molecular weight of 20 kDa. The structure of glutelin and prolamins contain 4 and 10 polypeptides and the molecular weight is 17–46 kDa and 16–64 kDa, respectively (Jarpa-Parra, 2018).

The determination of protein quality typically involves evaluating essential amino acid profile and digestibility. Legumes are rich in amino acids (around 50%) such as arginine, aspartate, glutamate, and lysine, and are comparatively lower in methionine and cysteine (Sosulski and Holt, 2020; Afify et al., 2012). Lentil protein has low levels of a few essential amino acids, including threonine, histidine, valine, isoleucine, leucine, and phenylalanine and tryptophan. Lentil is relatively low in sulfur-containing amino acids like methionine and cysteine compared to cereal grains and lysine content is always higher in lentil and in other grain legumes compared to cereals (Erskine et al., 2016; Zambrowicz et al., 2013). The nutritional quality of storage protein varies among different species of grain legumes (Sosulski and Holt, 2020). A study of seeds of different species of *Lens* showed there was variation in free protein amino acid content among species (Rozan et al., 2001).

The digestibility of a protein specifies the amount of protein that is hydrolyzed by the digestive enzymes relative to the consumed amount. This depends on the protein structure and the presence of compounds that limit digestion, such as the quantity and quality of protease inhibitors, polyphenols, fibers, non-starch polysaccharides (NSP), phytates, tannins, trypsin inhibitors, and lectins (Berno et al., 2007; López et al., 2018; Sá et al., 2019; Shi et al., 2017). The evaluation of protein quality is important because the bioavailability of essential amino acids and the digestibility of protein influences the growth and health of humans (Aguilar et al., 2015; Arribas et al., 2017; Potier and Tomé, 2008). In general, plant proteins have lower digestibility (75–80%) compared to animal proteins (90–95%) because the rigid cell walls result in lower enzyme accessibility, and seed coats and antinutritional factors interfere with biochemical processes (Annor et al., 2017; Kniskern and Johnston, 2011). Protein digestibility is influenced by exogenous factors (presence of antinutritional factors) and endogenous factors (cross-linking, hydrophobicity, and changes in protein secondary structure) (Duodu et al., 2003). Antinutritional factors inhibit protein digestion and increase endogenous nitrogen losses through interfering metabolism, and provoking deleterious effects on the gastrointestinal tract physiology (Adenekan et al., 2018; Schaafsma, 2012; Suffo Kamela et al., 2016; Tuśnio et al., 2017).

2.6 Influence of lentil seed germination on protein quality

The protein digestibility of legume seeds depends on processing conditions. To improve the plant protein quality by increasing the digestibility of free amino acids, peptides, and other forms of nutrients, several seed processing techniques, for example, cooking, autoclaving, germination, microwave, irradiation, spray and freeze-drying, fermentation, and extrusion have been studied (Sá et al., 2019).

Germination is one of the bioprocesses to improve protein digestibility, bioavailability, and biological properties as well as deactivation of antinutritional factors. Lower levels of antinutritional factors such as tannin and phytic acid were found in germinated seeds (Fouad and Rehab, 2015). For instance, polyphenols can form complexes with digestive enzymes, thereby hindering digestion and decreasing protein digestibility (Sá et al., 2019). During seed germination, the proteins are hydrolyzed and enzymes are formed in presence of water, which helps to break down indigestible proteins and make them in readily bioavailable forms (Ohanenye et al., 2022). Alongside proteins carbohydrates, lipids, and other metabolites are hydrolyzed and be readily

available to be utilized. A study of sprouted lentil sprouts reported that the amount of leucine, valine, lysine, and phenylalanine was increased, while there was a significant decrease in the content of total sulfur amino acids, threonine, tyrosine, and tryptophan (Fouad and Rehab, 2015; see **Table 2.1** below). Some studies have reported that germination increases the amount of essential amino acids (Afify et al., 2012; Elemo et al., 2011). Germination also disrupts the interactions of trypsin inhibitors and phenolic compounds with nutrients and minerals so that nutrients (attached proteins, carbohydrates) and phytochemicals become more accessible to digestive enzymes (α-amylase, pullulanase, phytase, and other glucosidases), and thus, minerals will also become bioavailable (EL-Suhaibani et al., 2020; Singh et al., 2015; Fait et al., 2011; Nkhata et al., 2018; Avilés-Gaxiola et al., 2018).

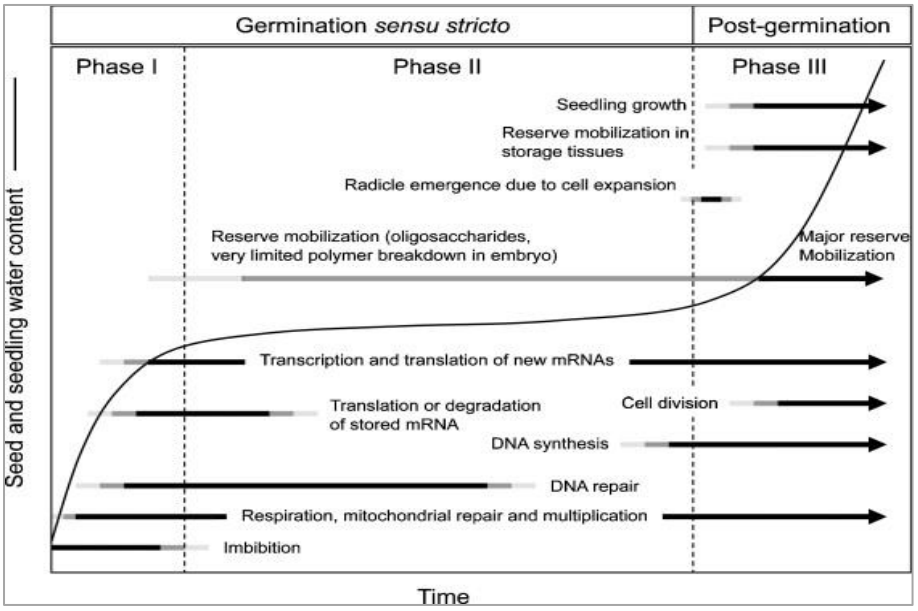


Figure 2.1. Metabolic activities during distinct phases of seed germination (adapted from Nonogaki et al., 2010).

Table 2.1. Change in carbohydrates, protein, antinutritional factors, amino acids and other bioactive compounds in raw and sprouted lentil (*Lens culinaris* Medik.) seeds (Fouad and Rehab, 2015)

Carbohydrates (g/100 g dry weight)	**Day 0**	**Day 3**	**Day 4**	**Day 5**	**Day 6**
Total sugar	5.01	8.60	10.42	12.27	14.50
Reducing sugar	1.52	2.76	3.24	3.90	4.72
Non-reducing sugar	3.49	5.84	7.18	8.37	9.78
Protein digestibility					
Free amino acids mg/g DW	1.86	13.80	14.02	14.10	15.20
Protein solubility g/100 g DW	3.62	5.10	5.90	6.00	6.10
Antinutritional factors (mg/100 g sample)					
Phytic acid	233.04	126.19	86.21	71.47	61.13
Tannins	466.10	243.00	210.17	200.01	189.20
Others					
Total polyphenols (mg gallic acid/100 g)	1341.13	1411.5	1463	1630.2	1510.1
Flavonoids (mg catechin/100 g)	398.33	429.5	461.38	483.1	496.21
DPPH radical scavenging %	40.76	49.26	54.73	61.31	62.19
Amino acid (g/100 g protein)					
Isoleucine	3.80	3.70	3.70	3.65	3.60
Leucine	7.8	8.00	8.10	8.10	8.25
Lysine	7.3	7.54	7.55	7.55	7.59
Cystine	0.7	0.64	0.63	0.62	0.62
Methionine	0.8	0.52	0.50	0.47	0.42
Total sulfur amino acids	1.5	1.16	1.13	1.09	1.04
Tyrosine	3.3	3.20	3.22	3.19	3.16
Phenylalanine	4.5	4.70	4.71	4.93	5.60
Total aromatic amino acids	7.8	7.90	7.93	8.12	8.76
Threonine	3.0	2.89	2.85	2.79	2.54
Tryptophan	1.2	1.2	1.16	1.15	1.15
Valine	4.5	4.96	5.70	5.75	5.95
Total essential amino acids	36.9	37.35	38.10	38.20	38.88
Histidine	2.5	2.56	2.31	2.45	1.96
Arginine	7.60	7.81	7.62	7.65	7.66
Aspartic acid	13.7	13.8	14.01	14.02	14.02
Glutamic acid	21.4	21.5	21.3	21.1	21.1
Serine	3.5	3.5	3.5	3.5	3.5
Proline	4.9	4.9	4.8	4.8	4.6
Glycine	3.6	3.6	3.5	3.4	3.4
Alanine	4.70	4.98	4.86	4.88	4.88

Through germination, bioactive compounds such as polyphenols, flavonoids, and DPPH radical scavenging activity are also enhanced (**Table 2.1**). This study found that the level of reducing and non-reducing sugar increased in sprouted seeds due to breakdown of starch and polysaccharides.

These breaking-down of polymers into simpler form improve both digestibility and bioavailability. Moreover, evidence was found that during the process of germination the biosynthesis rate for vitamins, minerals, amino acids, and health-promoting phytochemicals were improved (Žilić et al., 2014). A study of lentils revealed that soaking seeds in water for 8-14 h (depending on size) increased the bioavailability of essential amino acids, including methionine and cysteine (Marcela et al., 2016). In *Phaseolus vulgaris* sprouts, several bioactive peptides with ACE inhibitory, DPP-IV inhibitory, hypotensive, antidiabetic and antioxidant activities were identified (De Souza Rocha et al., 2015). While study on bioactive peptides in lentils are not well-characterized, a few studies reveal the presence of angiotensin-I converting enzyme, ACE inhibitory peptides (Bamdad et al., 2009b; Barbana and Boye, 2011). Rozan et al. (2001) showed that free protein amino acid content increased dramatically during germination, and in lentil, there is evidence for increased levels of amino acids (Fouad and Rehab, 2015).

2. 7 Phases of seed germination

Seed germination is the beginning of the first developmental phase in a plant's lifecycle. Based on cellular and metabolic activities, the stages of seed germination can be categorized into three phases (phase I, phase II, and phase III) (**Figure 2.1**).

2.7.1 Phase I

Phase I is the initiation of the rapid uptake of water by dry seed until the seed tissues are fully hydrated. As water uptake increases, the number of hydrated cells with reserve polymers of the dry seed, seed microstructure becomes larger and thus, just after the beginning of hydration, solutes like soluble carbohydrate monomers begin to leak out of the cells. Due to the lack of fully functional mitochondria and other cell components, fermentation or anaerobic respiration becomes the main source of energy during this period. Metabolic activities such as synthesis of new mRNAs, protein synthesis from existing mRNAs, repairing of DNA and mitochondria, and starch degradation takes place within the cell (He and Yang, 2013; Galland et al., 2014). In this phase, synthesis of seed storage proteins, heat shock proteins, dehydrin, homeobox-leucine zipper protein, and late embryogenesis abundant proteins is prominant. This is the phase after maturation and desiccation, and cell remodeling occurs in this phase (Galland et al., 2014). The rapid uptake of

water, as well as substrate and energy starvation, triggers the signaling cascade to produce phytohormones in the embryo, especially GA (gibberellic acid) towards the end of Phase I.

2.7.2 Phase II

Phase II starts when there is more limited water uptake with various cellular and biochemical events and is completed just before radical protrusion. This is the most crucial phase of seed germination as the activities in this phase determine whether the embryo cells will re-enter or halt in the cell cycle. This phase is signified by increased metabolic and cellular activity including DNA and mitochondria repair mechanisms, and protein synthesis from existing mRNAs (Galland et al., 2014). Gradually, glycolysis, fermentation, starch hydrolysis, pentose phosphate pathway, cell wall biosynthesis, non-function protein degradation, mitochondria assembly, amino acid biosynthesis processes start taking place (He and Yang, 2013). Then the TCA cycle, stocking protein degradation, starch biosynthesis, protein synthesis from newly synthesized mRNAs, synthesis of new mitochondria occurs before testa ruptures. Synthesis of gibberellic acid (GA) signals for the synthesis of amylase and hydrolytic enzymes such as protease, and lipase. Moreover, antioxidant or detoxification-related enzymes, for instance, monodehydroascorbate reductase-antioxidant, dehydroascorbate reductase, peroxidase, peroxiredoxin, superoxide dismutase, catalase, and mercaptopyruvate sulfurtransferase were found to be abundantly expressed during this phase in some species (Galland et al., 2014).

2.7.3 Phase III

Phase III is the post-germination phase that commences with radicle emergence until the radicle is fully elongated. Aerobic respiration (He and Yang, 2013), cell division and DNA synthesis, protein synthesis from new mRNAs radicle cell elongation, protein fate, and N/S remobilization are the significant metabolic events during phase III (Bewley, 1997; Galland et al., 2014). Proteins like actin-2 and tubulin-3 are abundantly expressed during this phase (Galland et al., 2014**)**. A study on the proteome during germination of *Arabidopsis* revealed the presence of glutamine synthetase, peptidase, tripeptidyl peptidase, proteasome subunit- proteasome protein degradation, cobalamin-independent methionine synthase, formyltetrahydrofolate synthetase (Galland et al., 2014**)**. At this point, seedling establishment begins.

The period of germination, however, depends on water intake capabilities and the seed microstructure that varies among species (Feng et al., 2018).

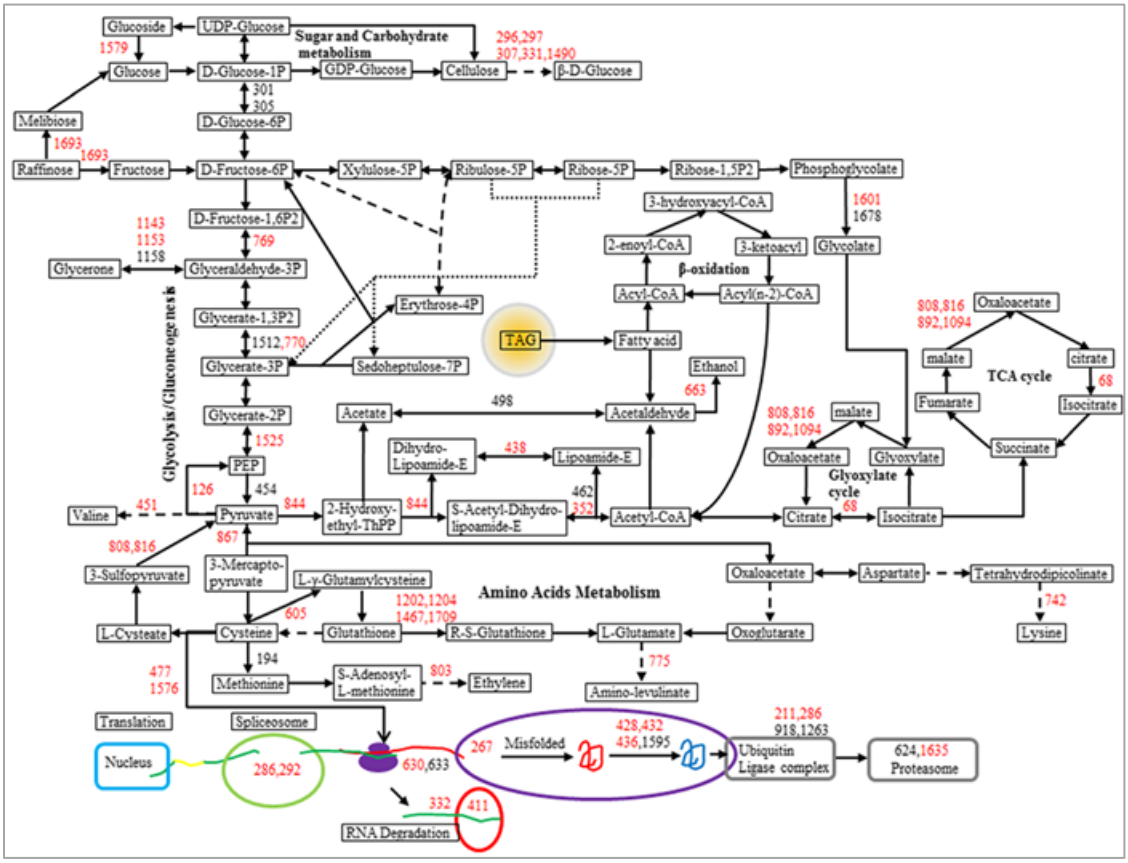


Figure 2.2. Metabolic pathways during seed germination of differentially expressed proteins in *Brassica napus* seeds. This pathway was developed from KEGG database (Adapted from Gu et al., 2019).

2. 8 Changes in the proteome during seed germination

The germination process initiates with the hydration of the seed through water uptake and ends with the protrusion of the radicle from the seed coat. This cellular hydration results in the breakdown of the stored reserve such as starch, sucrose, lipids, and proteins that provide energy to cells that helps them to rebuild the morphology from seed to plant. An added advantage of protein during seed germination compared to the other stored reserves, along with providing energy and key precursors for metabolism and structural formation, there are certain types of proteins or peptides or amino acids that execute the metabolic activities such as catalytic activity, production

of phytohormones, transportation as well as storage of molecules, supplying energy, repairing organelles, and maintenance of cellular activities. This plays a pivotal role in cellular activities and overall metabolism in plants. As an example, **Figure 2.2** demonstrates pivotal metabolic events that occur during seed germination of *Brassica napus* (Gu et al., 2019).

Studies suggest that the process of germination is programmed during seed maturation as after-ripening leads to changes in the seed proteome that prepares the seed for germination (Feng et al., 2018). The period after seed ripening contributes to changes in the seed proteome by modulating proteolytic activities, cellular signaling, translation of genes, and energy metabolism. Some studies also suggest that along with physiological state after the maturation process, the storage environment (including moisture, temperature, and oxygen) of the seed has an impact on germination (Carrera et al., 2008). The physiological processes of germination include the synthesis of reactive oxygen species (ROS) that modifies proteins, genes, and transcripts that are related to carbohydrate, energy, and amino acid metabolism, synthesis of secondary metabolites, genetic information processing, and anti-oxidative defenses. Moreover, changes occur in the metabolism, antioxidant, and plant hormone signaling that leads to the breakdown of seed dormancy, and therefore, germination occurs. At the initial stage of germination, rapid water uptake by dry seed triggers enlargement of seed in size and subsequently causes rupture of membrane structure, which leads to leakage of low molecular weight solutes and metabolites as well as water absorption (Koizumi et al., 2008; Ali and Elozeiri, 2017). The dearth of water within dried mature seeds makes cell organelles such as mitochondria, plastid as well as biomolecules like mRNA, proteins, enzymes, transcription factors and other biomolecules inactive (Rosental et al., 2014). Upon imbibition, these biomolecules are activated, and presumably by the energy supplied from anaerobic respiration as this is the pivotal pathway to serve energy for initiating cellular activities within the cell at the early stage of germination.

The storage proteins of legumes are localized in the vacuole of parenchyma cells present in seed cotyledons (Dey and Harborne, 1997). These compartments are known as protein storage vacuoles or protein bodies (Herman and Larkins, 1999). Even though the duration of seed germination may vary from species to species, the germination process proceeds with the uptake of water. As germination proceeds, the proteins and enzymes syntheses start taking place from the existing mRNAs within the cell during seed development. The energy required for this process is supplemented by anaerobic respirationor fermentation in the presence of pyruvate decarboxylase

(PDC), lactate dehydrogenase (LDH), and alcohol dehydrogenase (ADH) (He and Yang, 2013; Flikweert et al., 1996). Gradually, the partially active or inactive cellular organelles, for instance, mitochondria, plastids, and endoplasmic reticulum, etc. become fully functional (Czarna et al., 2016). During late phase I and early phase II, with the influence of red-light, enzymes belonging to the phytochrome superfamily become activated. The embryo of seed becomes activated and initiates the signal transduction and subsequent cellular activities in presence of phytochrome (absorb red lights), particularly the production of GA (gibberellin) (Toyomasu et al., 1998). Subsequently, GA is produced from geranylgeranyl diphosphate with the catalytic activities of enzymes like gibberellin 3-beta-dioxygenase, cytochrome P450 monooxygenases (consists of ent-kaurene oxidase and ent-kaurenoic acid oxidase), ent-copalyl diphosphate synthase, and ent-kaurene synthase (Tuan et al., 2019). Alongside, blue light is absorbed by cryptochrome that regulates transcription and cryptochrome-dependent proteolysis (Liu et al., 2011). Meanwhile, the DNA repair takes place, and the existing mRNA initiates protein synthesis with the help of prevailing low amount energy in form of adenosine triphosphate (ATP) from maturation stage followed by anaerobic respiration (Rosental et al., 2014). During phase II, an extensive molecular activity takes place in seed cells that are the critical factors for seed germination. While protein synthesis from existing mRNAs is ongoing, synthesis of protein from new mRNAs starts taking place in this phase (Galland et al., 2014). Studies have suggested mRNA degradation initiates either in phase I or in phase II. The signal transduced through GA for producing α-amylase and other hydrolytic enzymes, which are secreted into the cotyledons to drive the degradation of large molecular substances, such as starch, non-starch polysaccharides, and proteins, into smaller molecular compounds like reducing sugar, amino acids, peptides (Toyomasu et al., 1998; Ferreira 2009). The enzymes involved in the degradation are RNA endonucleases, RNA exonucleases, helicases, and polymerases (Houseley and Tollervey, 2009). Carbon and nitrogen assimilation are the two key pathways for plants to synthesize primary metabolites that are executed through the Calvin-Benson and GS-GOGAT cycle (Lehmann and Ratajczak, 2008; Galili et al., 2014). RuBisCo activase, ATP-synthase, NADP-malate dehydrogenase, nitrate reductase, nitrite reductase, glutamate dehydrogenase, and glutamate synthase enzymes are involved in these pathways (Lehmann and Ratajczak, 2008). In this phase, cellular organelles such as pro-mitochondria, and pro-plastids becomes active mitochondria, and plastids due to formation of inner membranes and synthesis of biomolecules to become fully functional (Pogson et al., 2015).

processes such as glycolysis, tricarboxylic acid cycle (TCA cycle), pentose phosphate pathway (PPP), and triose as well as hexose phosphate pool are enhanced for increased level of energy production, synthesis of amino acids, nucleic acids, triacylglycerols, photosynthetic apparatus, enzymes, and other biomolecules required for progression of germination (He and Yang, 2013). Even if protein synthesis from mRNA is also initiated in cells following hydration in presence of RNA polymerase, peptidyl transferase, and ribosomal proteins; after synthesis of transcription factors, and DNA polymerase transcription of new mRNA and eventually new protein synthesis takes place, which is one the deciding factor for germination (Rajjou et al., 2004). There is an increased abundance for enzymes involving in glycolysis, TCA cycle, PPP are found.

In addition, an abrupt shift in metabolism generates reactive oxygen species (ROS), which cause oxidative damage to the cell. In response, antioxidant enzymes such as ascorbate peroxidase (APX), superoxide dismutase (SOD), catalase 2 (CAT2), glutathione reductase (GR) are synthesized in cell (Kiran et al., 2020). Some studies suggest that ROS are detrimental to seed germination; however, some studies suggest that ROS accumulation can be beneficial for seed germination as this might result in better regulation of cellular growth, redox homeostasis, and protection against stress by triggering production of antioxidants (El-Maarouf-Bouteau and Bailly, 2008). Expression of various proteins such as heat shock proteins (HSP), adenosine kinase, chaperonin-haperonin tailless complex polypeptides, and translation elongation factor EF-Tu, prosystemin, systemin, chaperone aids in protein folding and shape formation for structural proteins were also observed in this phase. In addition, in the presence of blue light, cryptochrome becomes activated, which induces proteolytic activities (Liu et al., 2011).

In late phase II, proteolytic enzymes such as α-amylase, glucan glucohydrolases, β-D-glucan exohydrolases, α-galactosidase, invertase, cysteine proteases, aspartic protease, metalloproteases, serine carboxypeptidase, lipoxygenase are expressed. These enzymes degrade starch and non-starch carbohydrates, proteins, fatty acids into smaller molecules such as reducing sugars, amino acids, peptides, triacylglycerols (Albarracin et al., 2015). In addition, lipid mobilization is regulated by AT-hook motif-containing nuclear localized (AHL) protein that regulates lipid mobilization and fatty acid β-oxidation during seed germination, where TAG lipases SDP1 sugar dependent 1 (SDP1), and DALL5 and acyl-thioesterase KAT5 are involved in this process (Penfield et al., 2006).

A study on seed germination revealed that DNA and mRNA damage repairing also occurs during this phase and the enzymes includes in these processes are, photolyases, CPD photolyase, DNA alkyltransferases, 8-oxoguanine DNA glycosylase, apurinic/apyrimidinic endonuclease, ligase, DNA 3′-phosphatase, ARP endonuclease (Kiran et al., 2020; Córdoba-Cañero et al., 2014). A significant shift in metabolic processes results in the biosynthesis of new cell components which leads to radicle protrusion, mobilization of biomolecules for growth and development, and biosynthesis of other secondary molecules (Penfield et al., 2006). However, whether radicle protrusion requires mitotic activity differs from species to species. For instance, mitotic cell division is indispensable for tomato seed germination even though in some species, cell division is not required for the radicle protrusion (Gendreau and Corbineau, 2012; de Castro et al., 2000). The key proteins that are responsible for triggering cell cycle and division are cyclin-dependent kinase inhibitor 1, cell-cycle-associated protein kinase, mitotic cyclin, cytochrome c1, and mitochondrial tRNA methylthiotransferase (Barrôco et al., 2005). Seed germination process ends with radicle protrusion, and enter into the post-germination phase (phase III) where seedling establishment process takes place due to catabolism of stored food reserves. Remobilization of these reserves and radicle emergence are most evident in post-germination phases (Penfield et al., 2006; Tegeder, 2014). The degradation of ANFs (antinutritional factors) and other polyphenols, and dissociation of polysaccharides and minerals results in an improved bioavailability of nutrients. The formation of transporter proteins on membrane aids in mobilization between cells. Studies on transporter protein have identified the presence of sucrose binding protein, outer membrane lipoprotein, nitrate transporter, glucose binding protein, selenium binding protein, sec61 transport protein, ADP/ATP carrier protein, translocase of chloroplast, secretory carrier membrane protein, alpha-soluble NSF attachment protein, outer plastidial membrane protein, mitochondrial outer membrane protein, chloroplastic ferritins, mitochondrial 2-oxoglutarate/malate carrier protein, nascent polypeptide-associated complex, peroxisomal voltage-dependent anion-selective channel protein, outer membrane lipoprotein, and mitochondrial import inner membrane translocase subunit (Tim/Tom) in different organelles of cells (Pirovani et al., 2002; López-Millán et al., 2016; Rospert et al., 2002; Zhang et al., 2015; Hu et al., 2019; Fernie et al., 2020).

Although there is ambiguity regarding the cell wall biosynthesis and degradation during germination, a study by Buckeridge (2010) revealed that in seed storage cell wall, even if, there is

limited or no cellulose synthesis, other polysaccharides such as xyloglucan, arabinogalactan, mannan and galactomannan are present. The commonly found enzymes for cell wall biosynthesis in plants are glucan synthase, mannan synthase, galactosyltransferase, fuosyltransferase, xylosyltransferase, arabinosyltransferase, galacturonosyltransferase. Moreover, lentil is also well-known for producing raffinose, and enzymes involved in raffinose family oligosaccharides biosynthesis are galactinol synthase, raffinose synthase, stachyose synthase, verbascose synthase, and galactan-galactan galactosyl transferase. Cytoskeleton tube (actin and tubulin) formation are also observed during phase III that provide strength to the newly formed cell structure (Díaz-Camino et al., 2005). Moreover, even though storage proteins breakdown by proteolytic enzymes are initiated during late phase II, during post-germination phases the breakdown of storage proteins are enhanced (Müntz, et al., 2001). Studies have suggested that degradation of phytic acid, polyphenols and protease inhibitors also release polysaccharides, proteins and minerals during seed germination. In addition, folate, tocopherols, vitamin-C, polyphenols and GABA (gamma-aminobutyric acid) availability in the seeds are also found to be enhanced significantly (Patil and Khan, 2011; Mbithi-Mwikya et al., 2000; Chitra et al., 1996). It is suggested that these biomolecules transduce signals for seed germination. When cells get exposure to air, enzymes involved in ascorbate-dependent antioxidative pathway such as ascorbate peroxidase (APX), monodehydroascorbate reductase (MDAR), and dehydroascorbate reductase (DHAR) are found upregulated (Yin et al., 2014).

Chapter 3

Characterization of phases of germination and post-germination in seeds of seven *Lens* species

3.1 Introduction

Germination is a process during which mature seeds imbibe water and develop to the stage of embryo emergence. The germination process is initiated by water imbibition, the starting point for germination and growth of seedlings as metabolic events commence in hydrated cells (Nonogaki et al., 2010). The uptake of water results in a molecular shift during the subsequent period of germination and seedling establishment. Understanding the onset and duration of the distinct phases of the process is imperative for understanding germination.

The duration of seed germination varies from species to species (Xu et al., 2014). Physical seed characters affecting germination are seed size, and composition (Vázquez-Yanes and Orozco-Segovia, 1993; Blalogoe et al., 2020). The process of seed germination can be divided into three stages known as phase I, phase II, and phase III. Significant characteristic metabolic activity take place during each phase that leads to emergence of the embryo. In phase I, imbibition greatly impacts the hydration kinetics of the seed, resulting in changes in the seed coat, increases seed size, and results solute leakage, which enables smooth hydration of the seed during later stages of germination (Lechowska et al., 2019). During phase II, there is little further water absorption as seeds become saturated, while enzyme activation, synthesis, and other metabolic activities are enhanced. In phase III, both metabolic activities and water absorption are accelerated. However, the subsequent seedling growth is dependent on these cellular events.

Several studies were conducted previously on different crops for defining phases of seed germination (Cideciyan and Malloch, 1982; Bewley, 1997; Girolamo and Barbanti, 2012; Lutts et al., 2016; Ali and Elozeiri, 2017**)**. Some studies have also included descriptions of the metabolic events that place during these phases. However, in lentil, seed germination phases are not well characterized. Therefore, this study aimed to determine and describe the three distinct phases of seed germination in seven lentil species.

Table 3.1 Gene pool, species, genotype, center of origin, and photographs of seeds 7 lentil genotypes used in the experiments

Gene pool	Species	Genotypes	Center of origin*	Photograph
Primary	*L. culinaris*	CDC Redberry	Canada (Vandenberg et al., 2006). (Genome sequenced: Ramsay et al., 2021)	
	L. culinaris	Eston	Canada (https://doi.org/ PI 471917)	
	L. orientalis	BGE 016880	Israel (Fratini et al., 2004)	
	L. tomentosus	IG 72805	Sanliurfa, Turkey (https://doi.org/ 10.18730/7RPJ3)	
Secondary	*L. lamottei*	IG 110813	Lucena, Córdoba, Andalucía, Spain (https://doi.org/ 10.18730/8R6T3)	
	L. odemensis	IG 72623	Mardin, Turkey (https://doi.org/ 10.18730/7RH3D)	
Tertiary	*L. ervoides*	IG 72815	Kahramanmaras, Turkey (https://doi.org/10.18730/7RPWD) (Genome sequenced: Ramsay et al., 2021)	
Quaternary	*L. nigricans*	IG 72541	France (https://doi.org/10.18730/7REJ6)	

*Plant genetic resources accession data provided by: International Center for Agricultural Research in the Dry Areas (ICARDA), http://www.icarda.org/.

3.2 Materials and methods

3.2.1 Plant material and experimental design

The experiment was conducted in a laboratory at the Department of Plant Sciences at University of Saskatchewan, Saskatoon, Canada. For conducting imbibition experiments, six genotypes were selected from the primary, secondary and tertiary gene pools of lentil. These genotypes were CDC Redberry, BGE 016880, IG 72805, IG 110813, IG 72623, IG 72815, and belong to *Lens culinaris*, *L. orientalis*, *L. tomentosus*, *L. lamottei*, *L. odemensis*, and *L. ervoides*, respectively. For conducting a radicle and plumule emergence experiment, *L. nigricans* IG 72541 was added. The experiment was conducted in dark condition within a growth chamber where temperature and relative humidity were recorded using a data logger (Oneset HOBO U23 pro v2 temperature/relative humidity). Details of genotypes used for experiments are listed in **Table 3.1**.

3.2.2 Experimental procedure

Seeds of wild and cultivated species were scarified with a sharp knife at the convex portion of the seed surface to prevent damage to the embryo as well as the endosperm. Seeds were placed in a conical flask and soaked with 1% SDS (sodium dodecyl sulfate; Spruce Street, USA) and 70% ethanol for 1 and 2 min, respectively, then rinsed with water three times after chemical application. Seeds were then soaked in distilled water to determine the imbibition period. After the imbibition period, the water was poured off out and seeds were placed into a petridish where soaked germination paper was placed, and seeds were poured onto the paper. Another soaked germination paper was placed on the top to cover the seed and covered with a plastic lid to reduce moisture loss. The petri dish was then kept in dark condition in a growth chamber set 20 °C and with 90% relative humidity.

3.2.2.1 Collection of seed fresh weight

This experiment was conducted in 6 replications with 6 seeds in each replicate. After soaking, seed fresh weight was recorded each hour for 10 consecutive hours after soaking. For fresh seed weight measurement, seeds from petridishes were placed onto a paper towel and rubbed slightly to remove excess water from the seeds. The seeds were then put on a weighing paper and placed on an analytical scale with a static detector. The weight of seeds (in mg) was recorded immediately to prevent excessive water loss from the seed surface. Seeds were then placed back into the petridish

containing water. From the collected seed fresh weights, the average seed weight was calculated and was subjected to further analysis.

3.2.2.2 Collection and processing of microscopic images

For microscopic photographs, seeds were placed on a wet paper towel within a petridish, and photographs were taken using a Carl Zeiss light microscope. For the determination of phase I, micrographs were taken hourly intervals for 10 h. The diameter of each replicate of a total of 27 seeds (9 in each of 3 replicates) was collected. Then, to determine the duration of phase II, phase III, and phase IV, seeds were observed at 12-h intervals until radicle and plumule emergence.

3.2.3 Data collection and statistical analysis

Seed fresh weight and seed diameter data were recorded. Seed diameter data were obtained from ImageJ software by setting the scale of each image (imagej.nih.gov/ij/). Three measurements were taken from each seed, maintaining roughly a 45-degree angle between two adjacent lengths, and the average obtained from the three measurements was taken as the individual seed diameter for the analysis. The average diameter value from each seed sample was recorded for further analysis. The mean value, standard error as well as ANOVA were analyzed using R. Data were represented through MS Excel.

3.2.4 Determination of phase I, phase II, phase III, and phase IV

For the phase I determination, fresh weight of seeds were taken in each hour after removing excess water, curve was prepared with the fresh weight, and the first timepoint at which the seed fresh weight reached at maximum level due to rapid increase in seed fresh weight was determined as the endpoint of phase I. For the determination of phase II, the duration of morphological changes was recorded. Phase II for genotypes was determined when the radicle was about to emerge from the seed coat. Phase III is when radicle expansion occurs but no plumule emergence occurs. And phase IV is when the radicle and plumule are emerged from the seeds. The rapid change in fresh weight of seeds of all lentil genotypes indicated relatively rapid water uptake occurred.

3.3 Results

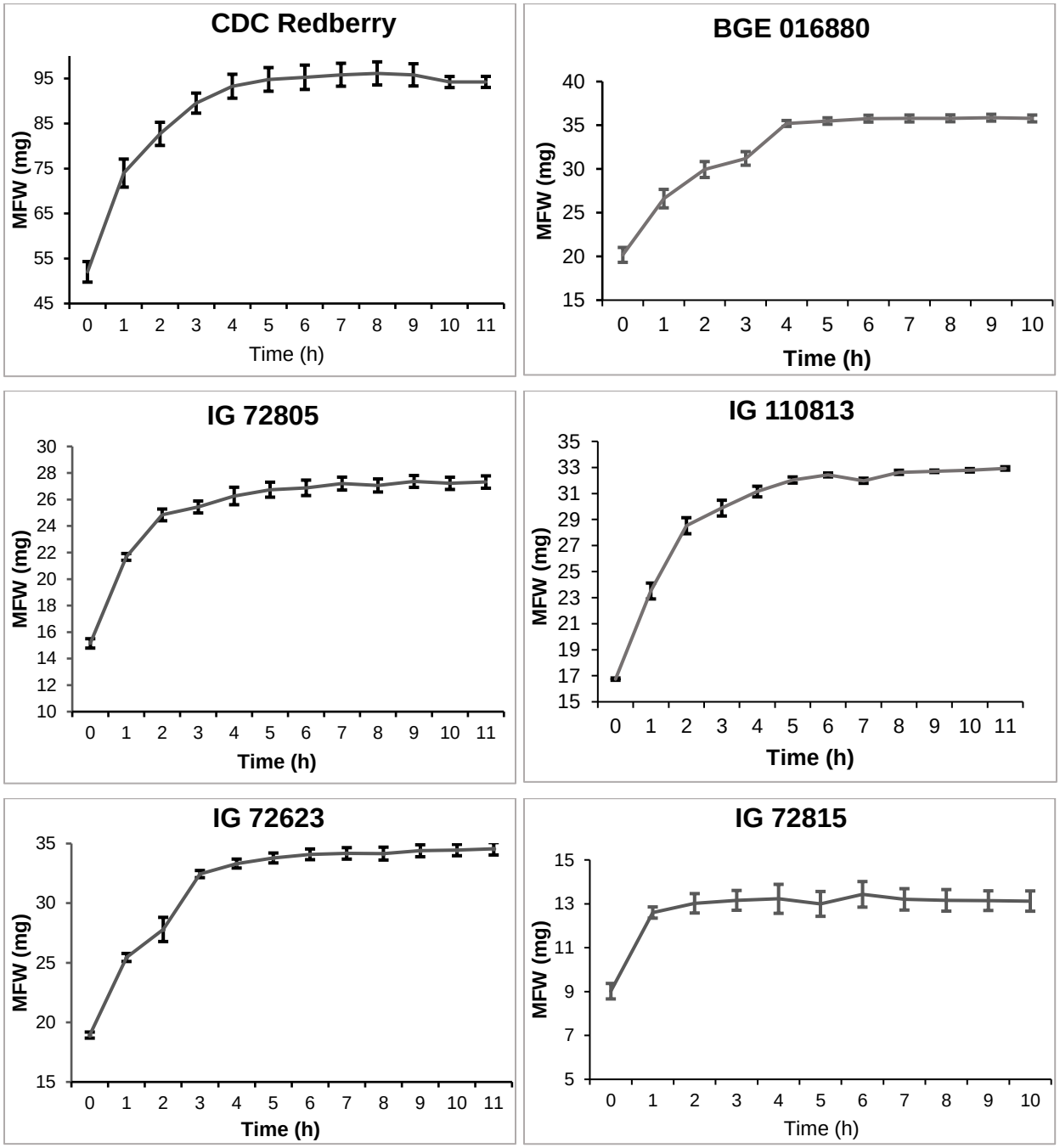


Figure 3.1 Seed imbibition curves of CDC Redberry, BGE 016880, IG 72805, IG 110813, IG 72623, and IG 72815. The X-axis represents duration after seed soaking up to 11 h, while the Y-axis represents average seed fresh weight as MFW.

Figure 3.1 shows the imbibition curves of 6 genotypes of *Lens* spp. The curve shows the change in the mean fresh weight of seed in each replicate. **Figure 3.1** revealed all the curves follow a

polynomial trend. Based on the imbibition curves, the endpoint of rapid water uptake in terms of weight gain was designated phase I. The duration of phase I for CDC Redberry, BGE 016880, IG 72805, IG 110813, IG 72623, and IG 72815 was 4, 4, 2, 3, 3, and 1 h after soaking (imbibition), accordingly (**Table 3.2)**. The mean weight gain during phase I for CDC Redberry (*L. culinaris*), BGE 016880 (*L. orientalis*), IG 72805 (*L. tomentosus*), IG 110813 (*L. lamottei*), IG 72623 (*L. odemensis*), and IG 72815 (*L. ervoides*) was 93.2%, 92.9%, 79.8%, 73.8%, 79.7%, and 70.0%, respectively (**Appendix A**).

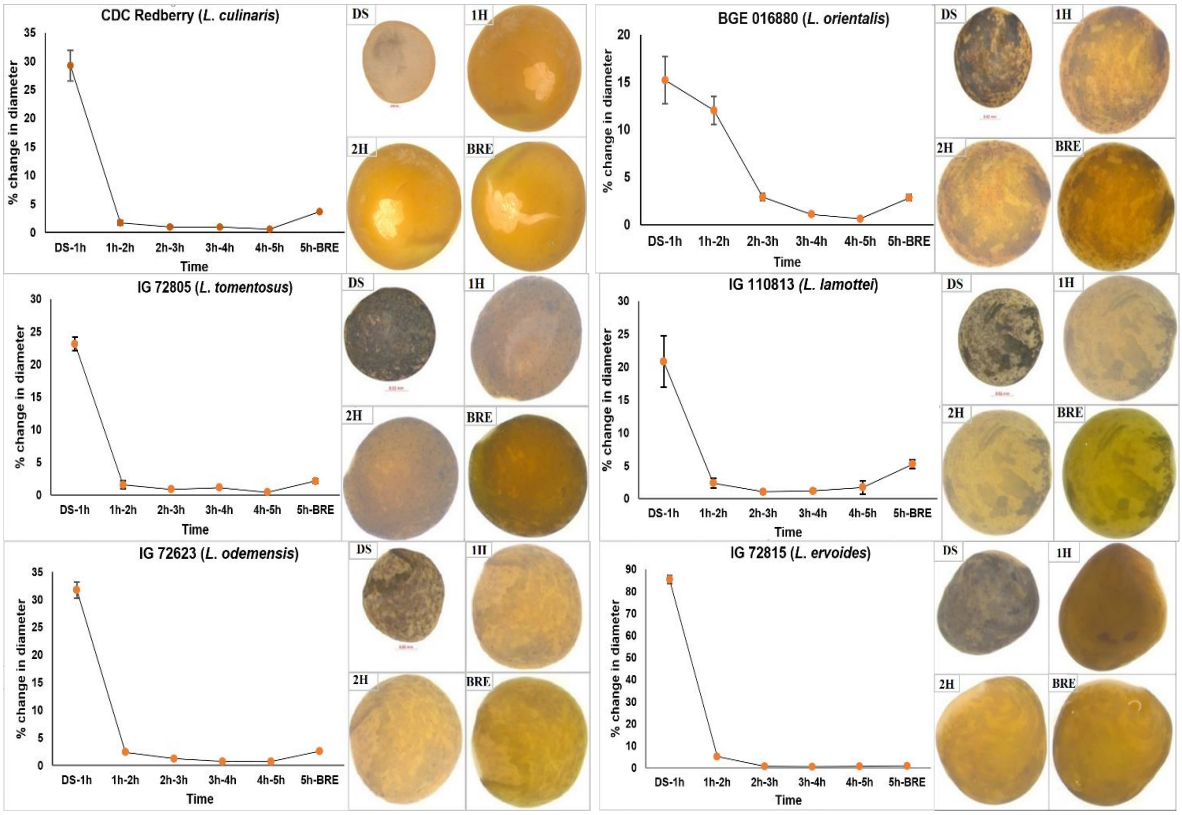


Figure 3.2 Percentage of change in seed diameter of CDC Redberry (*L. culinaris*), BGE 016880 (*L. orientalis*), IG 72805 (*L. tomentosus*), IG 110813 (*L. lamottei*), IG 72623 (*L. odemensis*), and IG 72815 (*L. ervoides*) during seed germination period. The micrographs of seed diameter show dry seed, imbibition after one hour, after two hours and before radicle emergence, represented as DS, 1H, 2H, and BRE, respectively.

Figure 3.2 shows the change in seed diameter during germination for CDC Redberry, BGE 016880, IG 72805, IG 110813, IG 72623, and IG 72815. The variation in seed diameter among the genotypes was significant (F-value = 84.87, $P < 0.0001$, **Appendix B**). The average initial seed diameter of CDC Redberry was 4.81 mm which was the largest, while the smallest seed diameter

was 2.58 mm for IG 72815. After CDC Redberry, the largest initial seed diameter was that of of IG 110813 followed by of BGE 016880, IG 72805, and IG 72623 (**Appendix B**). The seed diameter increased significantly in BGE 016880 after 2 h of imbibition, while in all other genotypes, the seed diameter increase reached maximum size up to 1 h of imbibition (**Figure 3.2**). The increase in seed diameter was the greatest in IG 72815 (*L. ervoides*) (around 85%) followed by in IG 72623 (32%) and CDC Redberry (29%). After imbibition to the end of phase II there was an increase between 2-5% in seed diameter in all genotypes except IG 72815, which had the smallest seed diameter, with less than 1% of seed diameter increase from phase I to phase II (**Figure 3.2**)

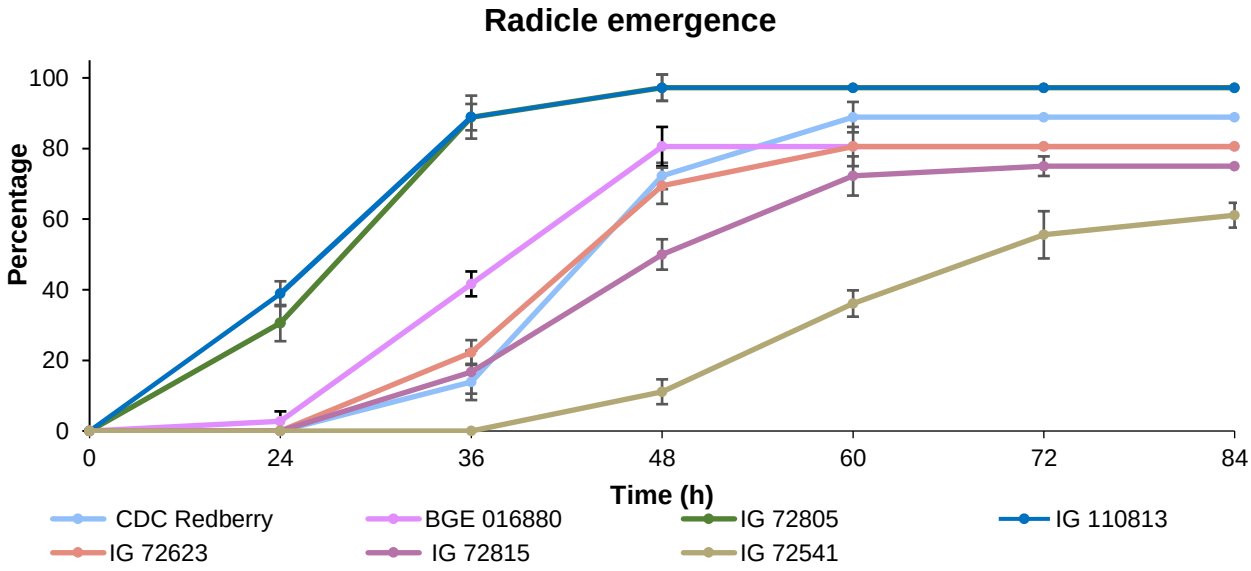


Figure 3.3.1 Seed germination curve showing cumulative percentage of radicle emergence for CDC Redberry (*L. culinaris*), BGE 016880 (*L. orientalis*), IG 72805 (*L. tomentosus*), IG 110813 (*L. lamottei*), IG 72623 (*L. odemensis*), IG 72815 (*L. ervoides*), and IG 72541 (*L. nigricans*) at different timepoints of seed germination. X-axis represents the timepoints at 12 h intervals after seed soaking, while Y-axis represents the percentage of radicle emergence.

Figure 3.3.1 shows the cumulative percentage of radicle emergence at different timepoints of seed germination. The results indicated that IG 110813 commenced radicle emergence earlier, within 24 h of soaking seeds. For IG 72815 and IG 72541, radicle emergence of viable seeds was completed in 72 h and 84 h, respectively. As Phase II can be defined as the time period after the end of phase I to before radicle emergence, phase II for CDC Redberry, BGE 016880, IG 72805, IG 110813, IG 72623, IG 72815, and IG 72541 was determined between 24 h-48 h, 24 h-48 h, 24 h-48 h, 15 h-36 h, 24 h-48 h, 24 h-60 h, and 36 h-84 h, respectively. Phase III occurs when the

radicle is fully elongated but the plumule is yet to emerge. All the lentil genotypes required a 12 h to 24 h period to reach plumule emergence from the timepoint of radicle emergence. Hence, the phase III for CDC Redberry (*L. culinaris*), BGE 016880 (*L. orientalis*), IG 72805 (*L. tomentosus*), IG 110813 (*L. lamottei*), IG 72623 (*L. odemensis*), and IG 72815 (*L. ervoides*) was determined between 36h-60h, 36h-60h, 36h-60h, 27h-48h, 36h-60h, and 36h-72h, accordingly (**Figure 3.3.2**). However, IG 72541 (*L. nigricans*) was the only genotype that took up to 48 h between radicle and plumule emergence. **Figure 3.4** shows the determined phases (0 – IV) of seed germination in CDC Redberry (*L. culinaris*), BGE 016880 (*L. orientalis*), IG 72805 (*L. tomentosus*), IG 110813 (*L. lamottei*), IG 72623 (*L. odemensis*), and IG 72815 (*L. ervoides*). The microscopic image represented for phase I at the timepoint that was determined by change in mean fresh weight of seeds from images taken at each hour after seed soaking of six species. Images of phase II, phase III, and phase IV were selected for representation by visual observation of individual seeds based on phase definitions cited in the review of literature chapter.

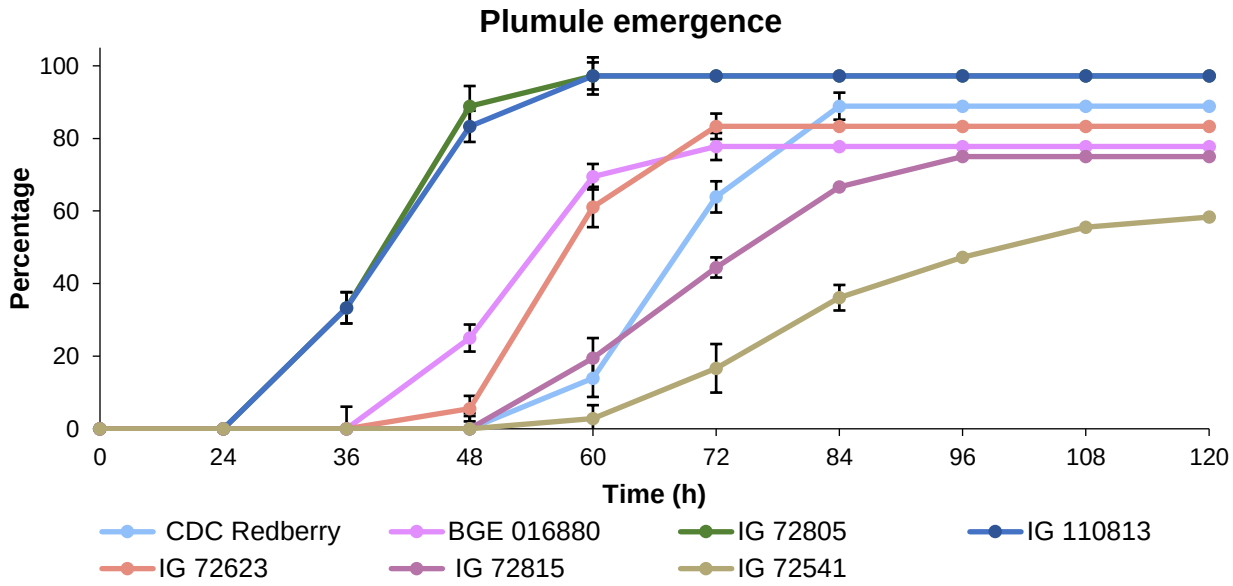


Figure 3.3.2 Seed germination curves showing cumulative percentage of plumule emergence for CDC Redberry (*L. culinaris*), BGE 016880 (*L. orientalis*), IG 72805 (*L. tomentosus*), IG 110813 (*L. lamottei*), IG 72623 (*L. odemensis*), IG 72815 (*L. ervoides*), and IG 72541 (*L. nigricans*) at different timepoints of seed germination.

Figure 3.4 Micrograph of five distinct phases (dry seed, phase I, phase II, phase III, and phase IV) of seed germination from a single seed of six lentil genotypes. A, B, C, D, E, and F represents seed of CDC Redberry, BGE 016880, IG 72805, IG 110813, IG 72623, and IG 72815, respectively.

3.4 Discussion

Water uptake in dry seeds commences with the rapid uptake in phase I through imbibition following a very limited uptake in phase II (Manz et al., 2005**;** Orozco-Segovia et al., 2007**;** Kestring et al., 2009). Therefore, phase I was determined by measurement of seed fresh weight. In this experiment, a wide range of increase in seed fresh weight was observed across the six genotypes studied.

Table 3.2 Determination of the duration of phase I of seed imbibition for seeds of CDC Redberry, BGE 016880, IG 72805, IG 110813, IG 72623, and IG 72815

Gene pool	Species	Genotype	Initial mean weight (mg)	Mean weight gain (%)	Phase I (h)
Primary	*L. culinaris*	CDC Redberry	41.12±0.18	93.23±1.64	4 h
	L. orientalis	BGE 016880	15.38±1.27	92.95±1.74	4 h
	L. tomentosus	IG 72805	12.41±4.67	79.85±0.37	2 h
Secondary	*L. lamottei*	IG 110813	15.61±1.12	73.77±2.48	3 h
	L. odemensis	IG 72623	16.01±2.14	79.75±0.93	3 h
Tertiary	*L. ervoides*	IG 72815	6.93±1.99	70.04±1.35	1 h

The highest percentage of seed weight gained to reach phase I was observed in the largest seeds (CDC Redberry), which had the highest initial seed weight (initial seed weight average was 41.12 mg) and gained 93.23% weight to reach phase I. IG 72815 had the lowest initial seed weight (6.93 mg) had a 70.04% weight increase during phase I (**Table 3.2**). Among the seeds with a medium range of initial weight, the increase in seed fresh weight varied from 73.8% to 92.9%. From this study, it was evident that the larger seeds took a longer period to maximum water absorbance compared to smaller seeds. A similar result was found in a study on seed germination of *Copaifera langsdorffii* where larger seeds absorbed water gradually compared to smaller seeds (Souza and Fagundes, 2014). Studies on imbibition with *Brassica napus* and *Lens culinaris* Medik seeds suggested that seed diameter increases over the period of germination and water uptake are higher in larger seeds than in smaller seeds (Lechowska et al., 2019; Al-Karaki, 1998). However, seed size did not have an impact on the time to reach phase I according to these studies. A study of seed water uptake also revealed that water absorption varies due to anatomical structure and adherence between the seed coat and embryo (Nakayama and Komatsu, 2008). A study of water absorption during phase II found very little or no absorption in scarified or even in un-scarified seeds, and

thus water absorption cannot be a determinant for phase II (Kestring et al., 2009), thus no fresh weight was measured for phase II determination.

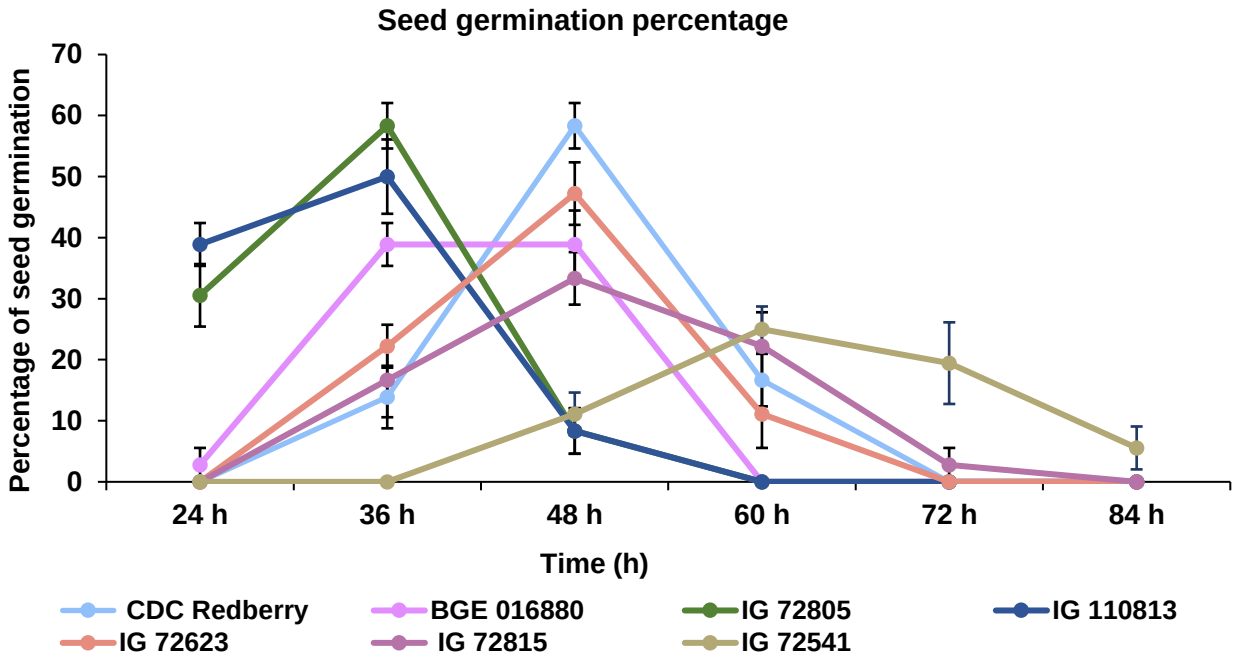


Figure 3.5 Percentage of seed germination of seven lentil seed species at different timepoints. The x-axis shows the time period after seed soaking while the y-axis represents the average percentage of seed germination.

This study on change in seed diameter during phases of seed germination in lentil species has also revealed that the change in seed diameter from phase I to phase II was ranging between 0.95% and 3.67% in all the genotypes studied. The increase in seed diameter was the highest at 1 h of imbibition, between 23.14% and 85.35% with an exception for BGE 016880 where seed diameter changed around 27.32% in 2 h of imbibition. In addition, CDC Redberry which had the largest seed size increased by around 37%, and IG 72815 which was the smallest seed size studied, increased by around 93% in seed diameter. By contrast, a study on seed size in *Hyptis suaveolens* showed that in the larger sized seeds increased 2 folds compared to small seeds (Mandal et al., 2010). Moreover, the increase in seed diameter was not associated with seed sizes (**Table 3.2**). As water absorption cannot be a determinant for phase II, images were taken at 12 h intervals that revealed a probable period of seed germination. Genotypes of all seven lentil species were studied for radicle and plumule emergence to determine the period of phase II and phase III. A previous

study on the velocity of seed size increase during germination showed that velocity of germination is higher in small seeds (Saeed and Shaukat, 2000), whereas the obtained result from lentil species in this experiment showed a different result. Among the genotypes studied, radicle emergence commenced within 36 and 48 h in lentil species with smaller seeds IG 72815 (*L. ervoides*) and IG 72541 (*L. nigricans*) hours, respectively; although among the genotypes having larger seeds, for instance, IG 110813 (*L. lamottei*), BGE 016880 (*L. orientalis*) and IG 72805 (*L. tomentosus*) took 24 h to reach radicle emergence, and the rest of the genotypes reached radicle emergence within 36 h of germination. This indicates that seed size of lentil species has no impact on the velocity of germination (**Figure 3.3.1, Figure 3.5**). However, this could possibly be influenced by the variability of species, by seed anatomy, and by composition of the seed coat. Phase II is defined as the period from full imbibition to the time when the radicle is about to emerge from the seed coat. The highest metabolic activities were observed in late-phase II of rice seed germination (He and Yang, 2013). In this study, samples for phase II was collected between 24h and 48h for CDC Redberry, BGE 016880, IG 72805, IG 72623; between 24h and 60h for IG 72815 and IG 72541, and between 16h and 24h IG 110813, respectively. Another study on the impact of seed size variation on germination percentage in *Rumex crispus* and *Rumex obtusifolius* has reported that large seeds are independent of germination percentage but that germination percentage was reduced in small seeds **(**Cideciyan and Malloch, 1982**)**. In this experiment, the germination percentage was found higher in larger seeds compared to small seeds. For instance, germination percentage in IG 72541 (*L. nigricans*) and IG 72815 (*L. ervoides*) was 61.1% and 75%, respectively, and in the other genotypes the percentage of seed germination was between 80.6% and 97.2%. This result indicates that small seeds have reduced germination percentage. Radicle emergence commenced within 36 h for all genotypes except for IG 72541 (*L. nigricans*). For plumule to emerge it required 12 to 24 h from radicle emergence for all the genotypes studied. Notwithstanding, in IG 72541 (*L. nigricans*) and IG 72815 (*L. ervoides*) for plumule and radicle emergence took a wider period than other genotypes studied. According to this study, the pattern of seed germination varied from species to species, except IG 110813 (*L. lamottei*), and IG 72805 (*L. tomentosus*) where the pattern of seed germination was similar (**Figure 3.5**).

3.5 Conclusions

From the above study, it was concluded that the duration of germination phases and proportion of germination percentage varied among genotypes across the seven lentil species. Imbibition period varied from species to species and in IG 72815 (*L. ervoides*) which had the the smallest seed, water uptake was faster. The rate of water uptake decreased after around 70% increase in mean initial fresh seed weight. The seed diameter change study indicated there was no significant change in seed diameter from phase I to phase II. The pattern of seed germination varied among species in terms of duration along with the percentage of radicle and plumule emergence. Across all of the *Lens* species studied, it took 12 to 24 h to reach Phase III from Phase II. In genotypes with small seeds, for instance, in IG 72815 the velocity of imbibition was faster although the time required to reach phase II and phase III were longer than the other genotypes with larger seeds.

Chapter 4

Influence of genotype, germination time and their interaction on protease activity and free amino acid profile

4.1 Introduction

Seeds of cultivated species of the Leguminosae family are a great source of protein, and lentil (*Lens culinaris* L.) is rapidly gaining recognition worldwide as a staple plant-based protein-containing food. Seed germination enhances the release of peptides and free amino acids that improve their digestibility and bioavailability in humans. The study of methods for improving protein quality is important for supplying quality protein as well as for genetic improvement. Seed germination enhances protease activity through breakdown of large proteins into peptides and amino acids. A study by Bautista et al. (2020) of protease activity and free amino acids (FAA) released during seed germination in lentil reported that after sprouting, there is increased protease activity that releases small peptides and FAA, resulting in improved digestibility. It was suggested that the increase of free amino acids is 4 to 6 times greater after sprouting compared to the unsprouted seeds (Bautista-Expósito et al., 2020). Another study that evaluated hydrolytic enzyme activities of lentils that during the 7 days of germination, the highest proteinase activity was found after 3 days of germination (Petrova et al., 2010).

Total protein amino acids available in the seed can be categorized into essential and non-essential amino acids. Amino acids are released from storage proteins during germination. These amino acids act as building blocks for synthesis of new proteins, as well as precursors for the synthesis of other biomolecules that are involved in seed germination and seedling establishment. Legume proteins are relatively high (around 50%) in the endogenous amino acids, arginine, aspartate, glutamate, leucine, and lysine, and are comparatively lower in methionine, threonine, tryptophan, and cysteine (Sosulski Holt, 2020). Seeds of the cultivated lentil are also rich in the endogenous amino acids such as arginine, aspartic and glutamic acids, and leucine, which represent almost 50% of their total amino acids. Lentils have relatively low amounts of EAAs like methionine, threonine, phenylalanine, tryptophan, histidine, valine, isoleucine, and leucine. Lentils are low in sulfur-containing amino acids such as cysteine and methionine. A comparative study of seeds of

cultivated lentil compared to wild lentil species revealed that the amino acid profile in *L. orientalis*, *L. ervoides*, and *L. nigricans* was identical to that of the cultivated lentil species (Bhatty, 1986). A more recent study that compared amino acids among different lentil species reported that *L. orientalis* contained higher levels of amino acids than other *Lens* species (Rozan et al., 2001). Moreover, a study of free amino acids (FAAs) in different cultivars of germinated seeds of cultivated lentils reported variation in levels of aspartic acid (Asp), serine (Ser), proline (Pro), alanine (Ala), cysteine (Cys), histidine (His), and lysine (Lys), methionine (Met) and tyrosine (Tyr) (Bautista-Expósito et al., 2020). Sulieman et al. (2008) reported that in lentil cultivars the EAAs and NEAAs increased after 3 days of sprouting, even if the level of sulfur-containing AAs including methionine and cysteine did not improve much through sprouting. Wild lentil species have primarily been used in lentil breeding to introduce resistance to biotic and abiotic stresses. This study was designed to compare the free amino acid profile and rotease activity in seeds of wild and cultivated lentil species during four different phases of seed germination to determine their potential for variation due to seed germination as well as in different species. The outcome from this study will provide an idea of changes in digestibility and availability of readily utilizable protein. Therefore, seeds from specific phase(s) of germination could be suggested for preparing nutritious foods as well as genotype(s) showing better changes in free amino acids during specific phase(s) could suggested for genetic improvement through breeding.

4.2 Materials and methods

This experiment was conducted at the Crop Science Field laboratory at the University of Saskatchewan. Two preliminary germination trials were conducted before performing protein analysis experiments using germinated seed samples of cultivated and wild lentil species.

4.2.1 Plant material and experimental design

This experiment was conducted at the Crop Science Field laboratory at the University of Saskatchewan. A total of 8 genotypes from 7 lentil species were selected for this experiment. The genotypes were CDC Redberry (*L. culinaris*), Eston (*L. culinaris*), BGE 016880 (*L. orientalis*), IG 72805 (*L. tomentosus*), IG 110813 (*L. lamottei*), IG 72623 (*L. odemensis*), IG 72815 (*L. ervoides*), and IG 72541 (*L. nigricans*). CDC Redberry was specifically used because it was designated as the genotype for sequencing the lentil genome (Ramsay et al., 2021). Temperature and relative humidity of the surrounding environment of place where experiment was conducted were measured using data loggers. 3-5 g of scarified seeds were used each time of sample collection, and this experiment was repeated three times.

4.2.2 Experimental procedure

4.2.2.1 Sample collection for protein analysis

For sample collection, seeds were scarified with a sharp knife at the top edge of the convex surface. Temperature and relative humidity were measured using data loggers; 3-5 g of scarified seeds were used for each time of sample collection. Seeds were washed in a conical flask, first soaked with 1% SDS followed by 70% ethanol for 2 min and then rinsed with water three times following each chemical application. Sampling was performed at different time endpoints of the germination process. Phase I of germination corresponded to the end of the imbibition phase. The washed seeds of CDC Redberry, Eston, BGE 016880, IG 72805, IG 110813, IG 72623, IG 72815, and IG 72541 were then imbibed within tubes for 4h, 4h, 2h, 3h, 3h, 3h, 1h, and 2h, respectively (following the imbibition curve in chapter 3). After completing phase I, the seeds were placed on a wet germination paper; a second wet germination paper was placed on top to cover the seeds and maintain moisture levels. Then the entire tray was covered with a perforated polyethylene bag to retain humidity and to ensure proper aeration during seed germination. The trays were then kept in the dark condition and data loggers were activated to measure the temperature and relative

humidity. Distilled water was sprayed at 12 h intervals to ensure sufficient moisture for seed germination. Mean seed weight of dry seeds were collected from individual genotype to ensure the collection of sufficient amount of samples from distinct phases of seed germination. For phase II, samples were collected based on the time interval of radicle emergence from soaking seeds found from Chapter 3. The seed samples for phase II were picked at two different times of that interval period.

The germinating seeds to reflect different timepoints in the process. For collecting phase III samples, seeds were observed intermittently, and collected on the basis of the radicle length reaching half of the seed diameter, or after 12-16 h from radicle emergence, whichever occurred first. Phase IV samples, corresponding to the time at which both plumule and radicle emerged from the seed coat, were also included for comparison with the post-germination phase. This experiment was repeated three times to collect samples for three biological replicates.

4.2.2.2 Sample drying and grinding

Seeds were collected into 20 mL tubes and kept for 1 h in -80 ºC freezer. The tubes were then transferred quickly to a freeze dryer after taking out the lids and placing aluminum foil sheets on the top of the tubes with slight perforation for removing air. Samples were freeze dried at -55 ºC and 0.012 mbar pressure for 60 h. Samples were then ground to fine powder using 6 mm and 4 mm metal beads in tubes for 8 min in an electrical grinder at maximum speed. For control samples, dry seeds were washed and processed in a similar way. Each of the obtained samples were divided into two technical replicates after grinding to avoid sample loss.

4.2.3 Determination of protease activity

4.2.3.1 Sample extraction and preparation

For extraction, 200 mg of sample was weighed in 2 mL tube and 1 ml of double-distilled water was poured into it. This suspension was then vigorously vortexed for a few minutes until mixed properly. Then the mixture was centrifuged at 25 ℃ and 1,500 rpm for 60 min using a thermomixer (Eppendorf, Madrid, Spain). After that, tubes were centrifuged at 25 ℃, 14,000 *g* for 10 min, and the supernatant was collected for protease activity analysis.

4.2.3.2 Evaluation of protease activity

Measurement of protease activity performed using the Pierce™ Fluorimetric Protease Assay Kit (ThermoScientific, Madrid, Spain). Three replicates of sample extracts, trypsin solutions (used as standard at a concentration range from 0 to 500 ng/mL) or distilled water (blank) were dispensed (100 μL) in 96-well plates and mixed with 100 μL of native casein (10 ng/mL) that was labeled using a large molar excess of fluorescein isothiocyanate. Fluorescence properties of this heavily labeled, intact protein substrate change dramatically upon digestion by proteases, resulting in a measurable indication of proteolysis. Plates were incubated for 10 min at 30 ℃ in darkness. Fluorescence read performed in every 2 min using a microplate reader (Biotek, Winooski, VT, USA) at excitation and emission wavelengths of 485 and 538 nm, respectively.

4.2.3.3 Quantification

The increase of relative fluorescence units (RFU) within 10 min of reaction was calculated and interpolated in the standard calibration curve were ΔRFU was plotted as function of trypsin concentration (mg). The results were expressed as ng of trypsin activity equivalents /mg of sample.

4.2.4 Evaluation of free amino acids

4.2.4.1 Sample extraction and preparation

For extracting free amino acids (FAAs), 200 mg of lentil seed powder from each sample were taken and homogenized in 2 mL of 0.01 M HCl solution containing 10 μmol/mL norvaline as internal standard. This suspension was then centrifuged at 2500 g for 15 min and the supernatant was collected for subsequent analysis by reversed-phase high-performance liquid chromatography (RP-HPLC) and diode array detection (DAD).

4.2.4.2 Sample separation through high performance liquid chromatography (HPLC-DAD)

Agilent 1200 chromatographic system (Agilent Technologies, Inc., Wilmington, E, USA) equipped with a G1329A automatic liquid sampler was used for amino acid analysis. Separation of amino acids was performed into an Agilent Zorbax Eclipse Plus C18 column (4.6 × 250 mm, with particle size of 5 μm) at 40 ºC. Two solvents (A and B) were used as the mobile phases where A consisted of 10 mM Na2HPO4: 10 mM $Na_2B_4O_7$, 5 mM $NaN0_3$ at pH 8.2, while B was composed of acetonitrile: methanol: water (45:45:10, v: v: v). The injection volume was 20 μL

and with a flow rate of 1.5 mL/min. The gradient flow for chromatographic separation was started from 2% B for 0.5 min followed by 57% B for 30 min then 100% B for 10 min. Between each run, the system was started back at normal condition and kept for 2 min for column re-equilibration.

4.2.4.3 Amino acid identification and quantitation

Amino acid detection was performed using automated derivatization in the autosampler. Derivatization reagents: borate buffers (0.4 M in water, pH 10.2), ortho-phthalaldehyde (OPA), 10 mg/mL in 0.4 M borate buffer and 3-mercapto proprionic acid) and 9-fluorenylmethoxycarbonyls (FMOC), 2.5 mg/mL in acetonitrile) were ready-made solutions supplied by Agilent. They were transferred from their container into an autosampler vial. DAD was set up for collecting two channels (Signal A 338 nm, to detect OPA derivitized amino acids and Signal B 262 nm, to detect FMOC-yls derivitized amino acids). Peak identification was performed by retention time comparison with amino acid standards. Standard solutions of 20 amino acids available from Agilent (1 nmol/μL) were used to prepare calibration curves. Calibration curves with a standard concentration range from 10 to 1000 nmol/mL of individual free amino acids were determined, with each concentration measured in triplicate. The linearity was evaluated by the calibration curves for each standard and least-squares regression line relating to the absorbance peak area.

4.2.4.5 Statistical analysis

All data were analyzed using Statistica software. Variance in concentrations of free amino acids (TFAAs, FNEAAs and FEAAs) among different phases was analyzed through the *post hoc* Duncan test at a 5% level of significance ($p \leq 0.05$). Data were represented through GraphPad Prism version 8.0 and MS excel.

4.3 Results

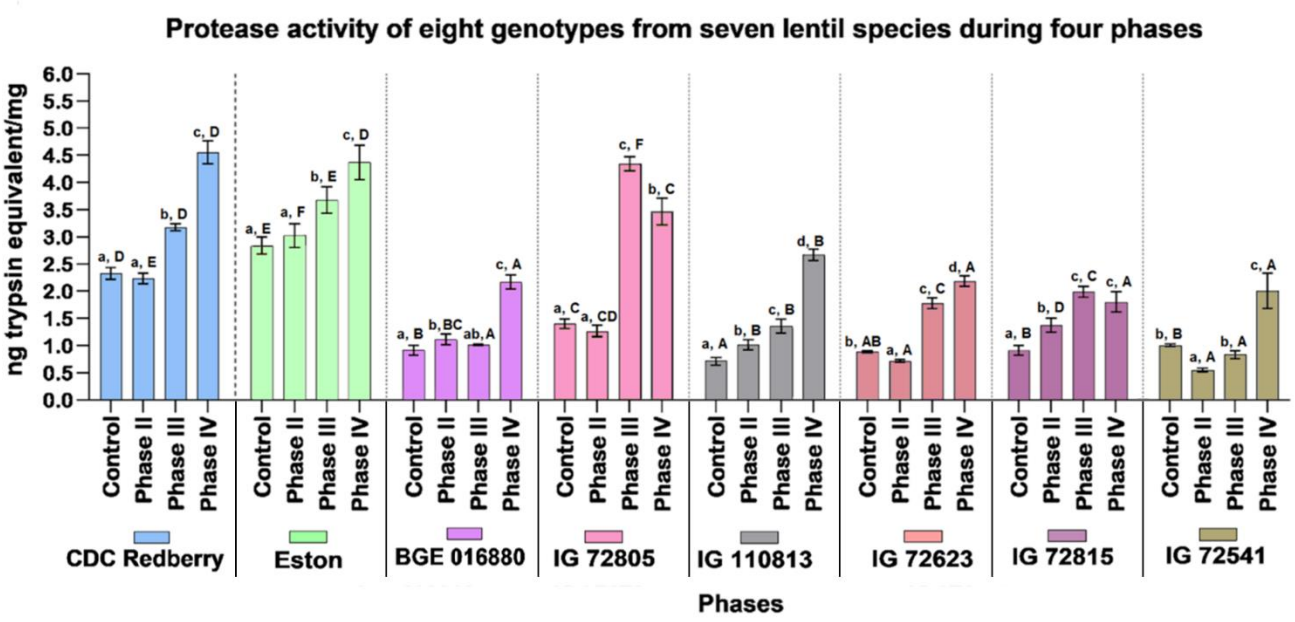


Figure 4.1 Protease activity of seven lentil species in four phases of lentil seed germination. Different lowercase letters indicate statistical differences among different phases ($p \leq 0.05$ post hoc Duncan test). Different uppercase letters denote statistical differences among samples at the same phase ($p \leq 0.05$, posthoc Duncan test). Values are for one biological replicate (n = 1).

Figure 4.1 illustrates the amount of protease activity of CDC Redberry, Eston, BGE 016880, IG 72805, IG 110813, IG 72623, IG 72815, and IG 72541 during control, phase II, phase III, and phase IV (**Appendix C**). Protease activity was the highest in Eston in dry seeds and phase II, while in phase III and phase IV, the highest protease activity was found in seeds of IG 72805 and CDC Redberry, respectively. A gradual increase in protease activity as the seed germination progress was found in the cultivated genotypes Eston, and CDC Redberry. A higher increase from phase III to phase IV was observed in BGE 016880, IG 110813, and IG 72541. The highest protease activity was found for CDC Redberry, Eston, BGE 016880, IG 110813, IG 72623, and IG 72541 seeds in phase IV was 4.55, 4.37, 2.17, 2.68, 2.18, and 2.01 mg trypsin equivalent/mg flour, respectively. However, in IG 72805 and IG 72815, the protease activity in phase III was 4.35 and 1.99 mg trypsin activity equivalents/mg flour, which was the highest among all the phases studied of these genotypes. For IG 72805, the protease activity increased 2.43-fold from phase II to phase III, which was the highest change between phases among all genotypes studied. Overall, the protease activity was the lowest in the control in all genotypes excluding CDC Redberry, IG 72805,

and IG 72623 in which protease activity was the lowest during phase II. In addition, in IG 72541, the protease activity was lower in both phase II (0.55 mg trypsin equivalent/mg flour) and phase III (0.83 mg trypsin equivalent/mg flour) than the control (1.01 mg trypsin activity equivalents/mg flour).

The total free amino acids (TFAAs) analysis of different genotypes of lentil species is shown in **Figure 4.2 (**also see **Appendix D)**. There was a significant increase observed in TFAAs in phase IV compared to all other phases in all the genotypes studied. The highest amount of TFAA was found in IG 110813, in control and phase IV - 14.52 mg/g flour and 37.81 mg/g flour, accordingly. Overall, in some wild species for instance, in BGE 016880, IG 110813, IG 72623, and IG 72815 the amount of TFAA was found to be significantly higher compared to the cultivated lentil CDC Redberry.

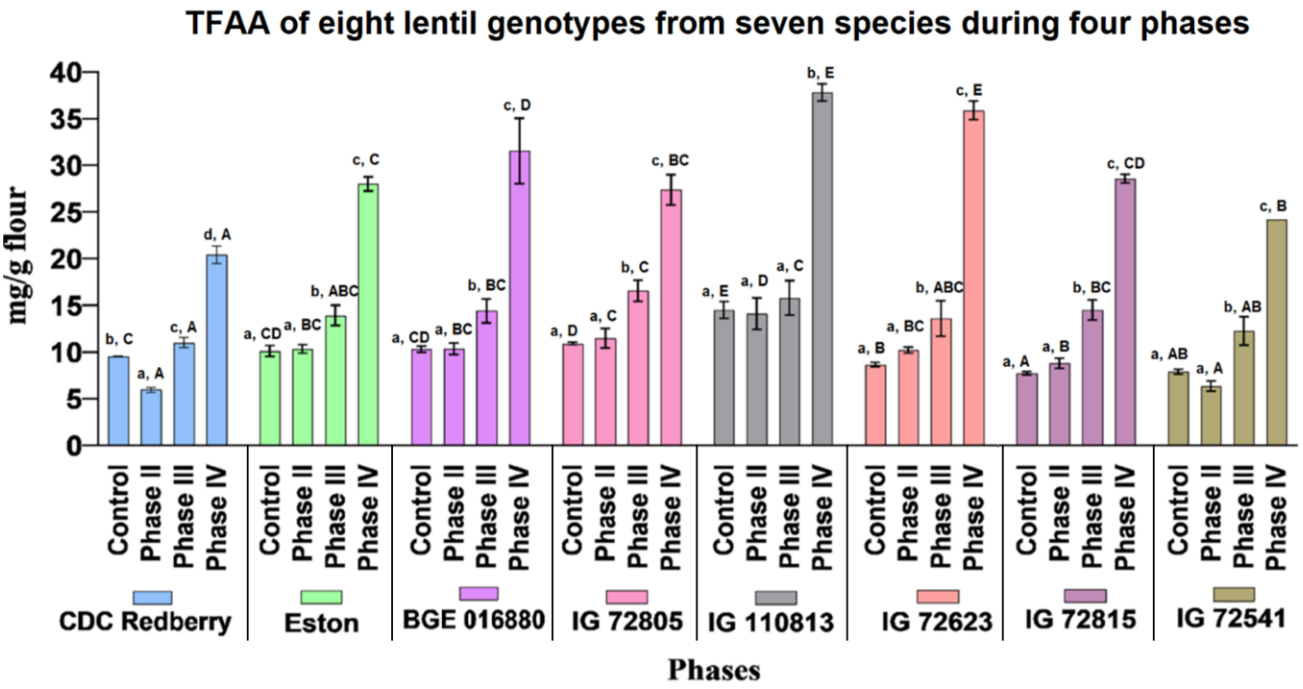


Figure 4.2 Total free amino acids (TFAA) of eight lentil genotypes at four stages (control, phase II, phase III, and phase IV) of germination. Different lowercase letters indicate statistical differences among different phases ($p \leq 0.05$ post hoc Duncan test). Different uppercase letter denotes statistical differences among samples at the same phase ($p \leq 0.05$, post hoc Duncan test). Values of one biological replicate (n = 1).

Essential free amino acids (FEAAs) and non-essential free amino acids (FNEAAs) analysis revealed that the amount of FEAA and FNEAA increased with germination, particularly from

phase III to phase IV (**Figure 4.3**). For instance, in BGE 016880, IG 110813, and IG 72623 the increase in FEAA was 13.44, 7.52, and 7.23 folds from control to phase IV. While the increase in FEAA between control and phase III in these genotypes was 2.96, 1.42, and 1.71 folds, respectively, which was comparatively lower than the increase between control and phase IV.

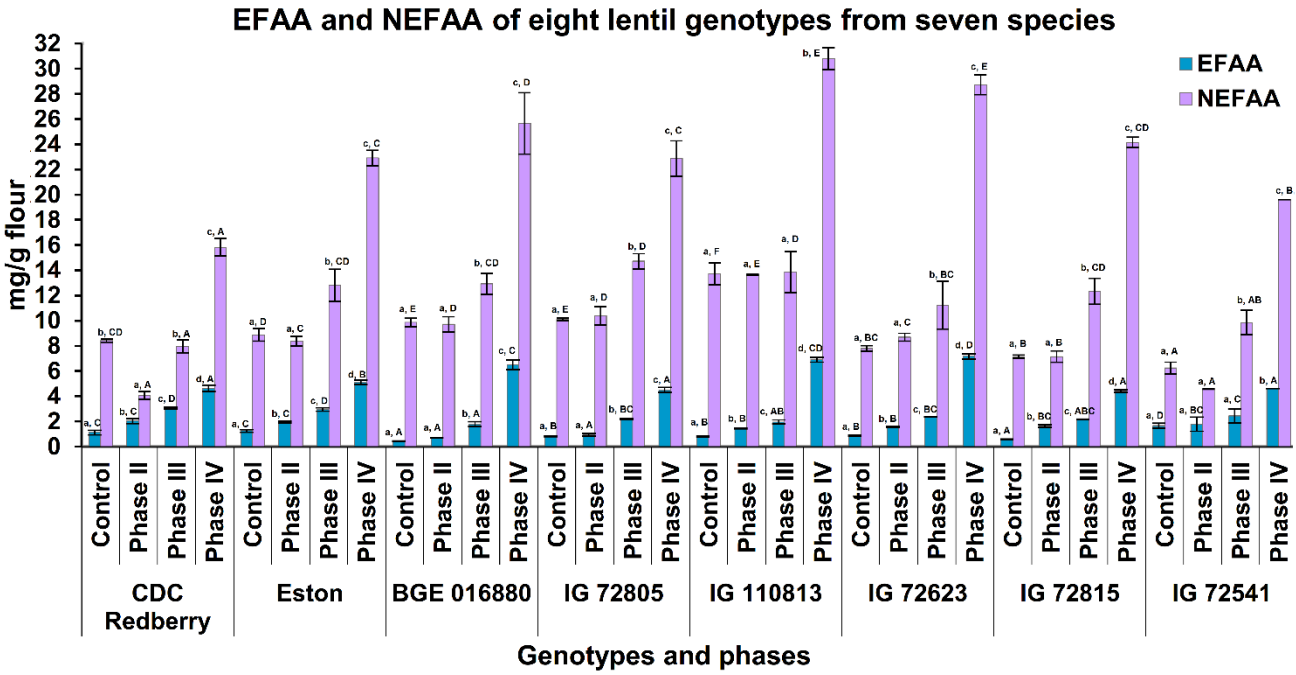


Figure 4.3 Essential (FEAA), non-essential (FNEAA) free amino acids in eight lentil genotypes at four stages (control, phase II, phase III, and phase IV) of germination. Different lowercase letters indicate statistical differences among different phases ($p \leq 0.05$ post hoc Duncan test). Different uppercase letter denotes statistical differences among samples at the same phase ($p \leq 0.05$, post hoc Duncan test). Values of one biological replicate (n = 1).

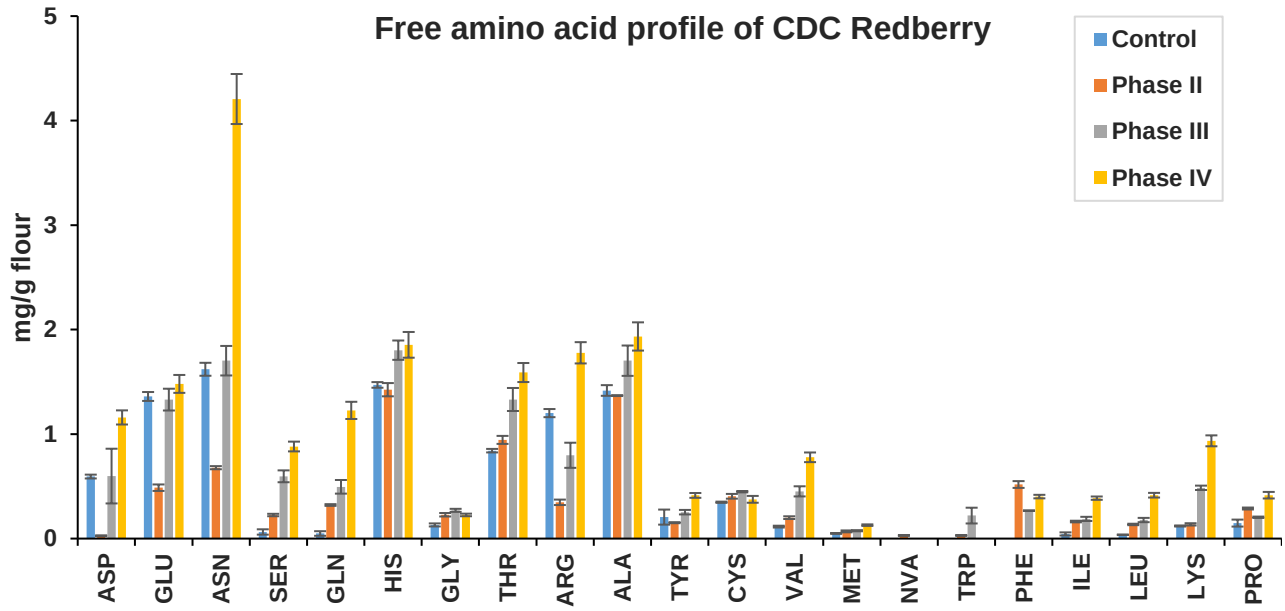


Figure 4.4.1 Individual free amino acids content of CDC Redberry (*L. culinaris*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

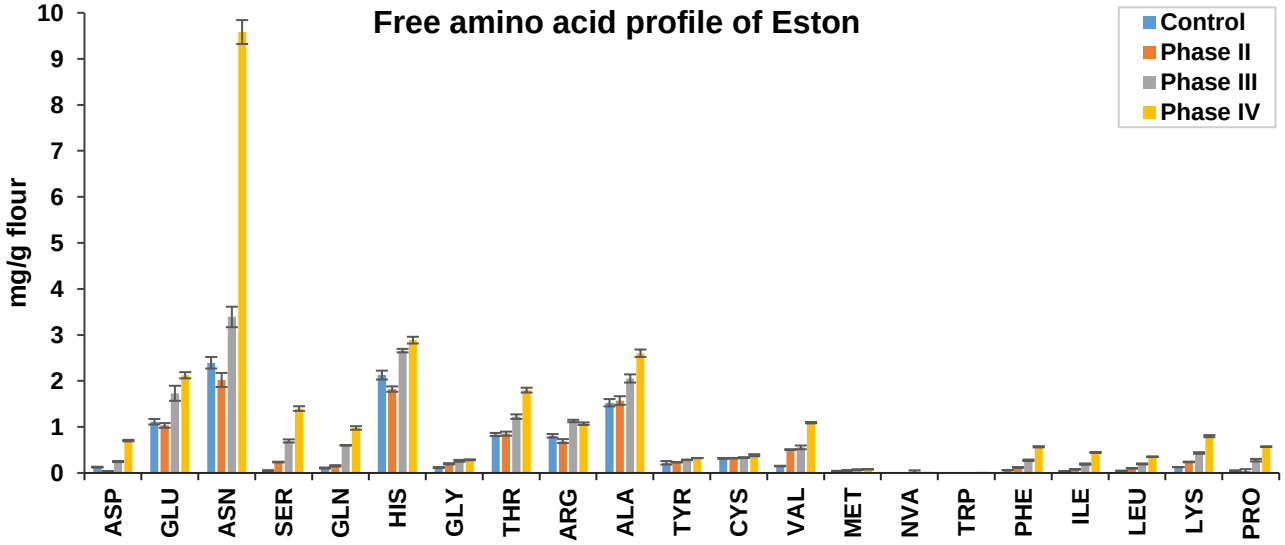


Figure 4.4.2 Individual free amino acids content of Eston (*L. culinaris*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine,

valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

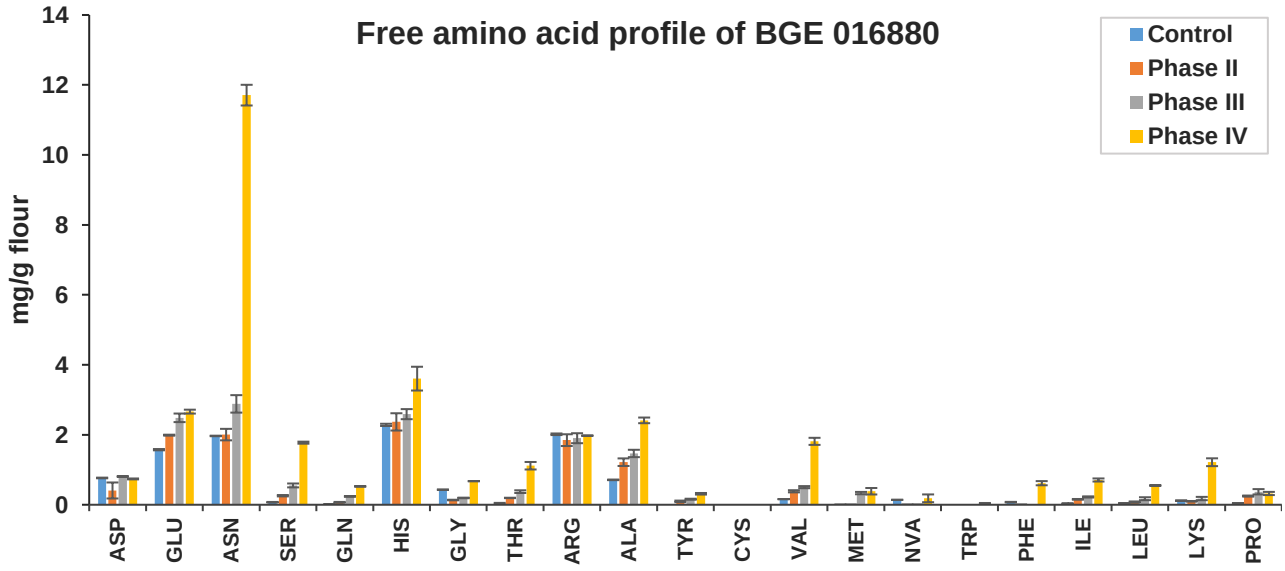


Figure 4.4.3 Individual free amino acids content of BGE 016880 (*L. orientalis*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

Similarly for NEFFA, for IG 72623, IG 72815, and IG 72541 there was an increase of 2.69, 2.38, and 2.14 folds from control to phase IV, while between control and phase III it was 0.44, 0.73, and 0.57 folds increase, accordingly. However, between phase III and phase IV, the highest increase of FEAA was observed in BGE 016880 (2.52 folds), followed by IG 110813 (2.65 folds), and IG 72623 (2.03 folds), respectively. Overall, the amount of increase in FEAA was higher than FNEAA from phase III to phase IV (**Figure 4.3**). From the individual amino acid analysis, asparagine (ASN) was found to be the most abundant amino acid followed by histidine (HIS) in control seeds of all lentil species (**Figure 4.4.1** to **Figure 4.4.8**). Among the control seeds of eight genotypes studied, in IG 110813, IG 72805, and BGE 016880 the amount of asparagine (ASN) was found to be 2.96, 2.22, and 2.28 mg/g flour, the highest level of all AAs. The histidine (HIS) content was between 2.96 and 2.22 mg/g in IG 110813, IG 72805, and BGE 016880, which is comparatively higher compared to the other genotypes studied. In some genotypes arginine (ARG), and glutamic acid (GLU) were also present in relatively high amounts in the wild

genotypes, while alanine (ALA) was present in higher amount in CDC Redberry and Eston. For instance, arginine (ARG) was present in 2.01 and 2.18 mg/g in BGE 016880 and IG 110813. Glutamic acids (GLU) were present in IG 110813, IG 72815, and IG 72623 between 1.53 and 1.42 mg/g flour. In CDC Redberry and Eston, alanine (ALA) was found 1.35 and 1.95 mg/g flour, respectively. The sulphur containing amino acids (cysteine and methionine) varied greatly among genotypes and phases of seven lentil species. Cysteine (CYS) was present in dry intact seeds of CDC Redberry, Eston, IG 72805, IG 72623, and IG 72815. Among theses, cysteine in IG 72805 was not detected in later phases as germination progressed. Conversely, in IG 110813 and IG 72541, where cysteine was not found at the initial stage, it was expressed later in the in germination phases, and in IG 72541 the level was increased about 4.4 folds from phase III to phase IV.

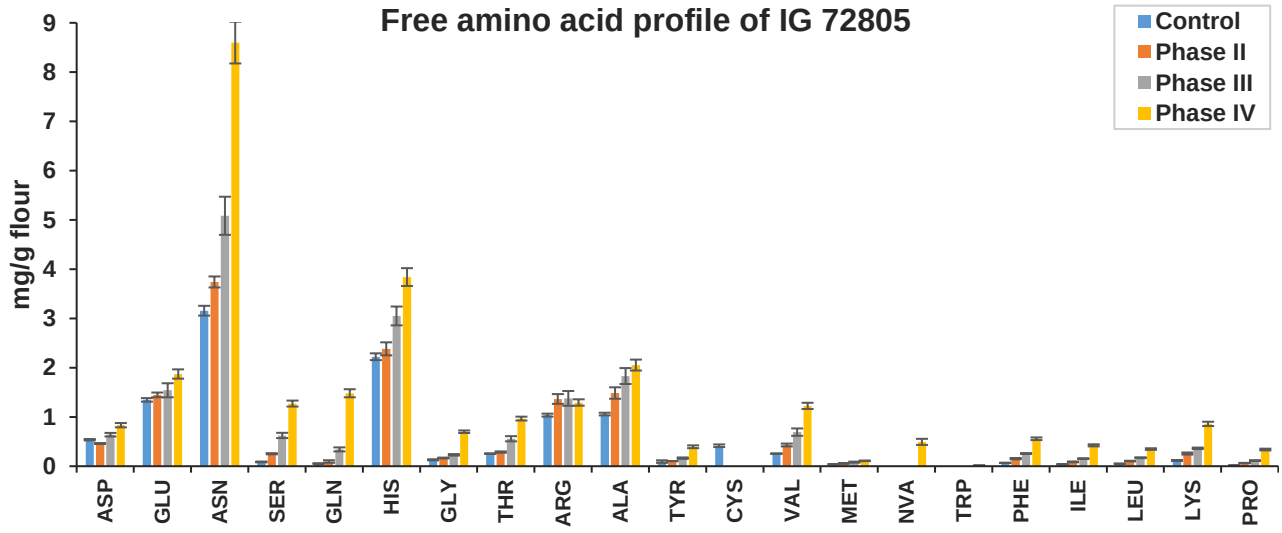


Figure 4.4.4 Individual free amino acids content of IG 72805 (*L. tomentosus*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

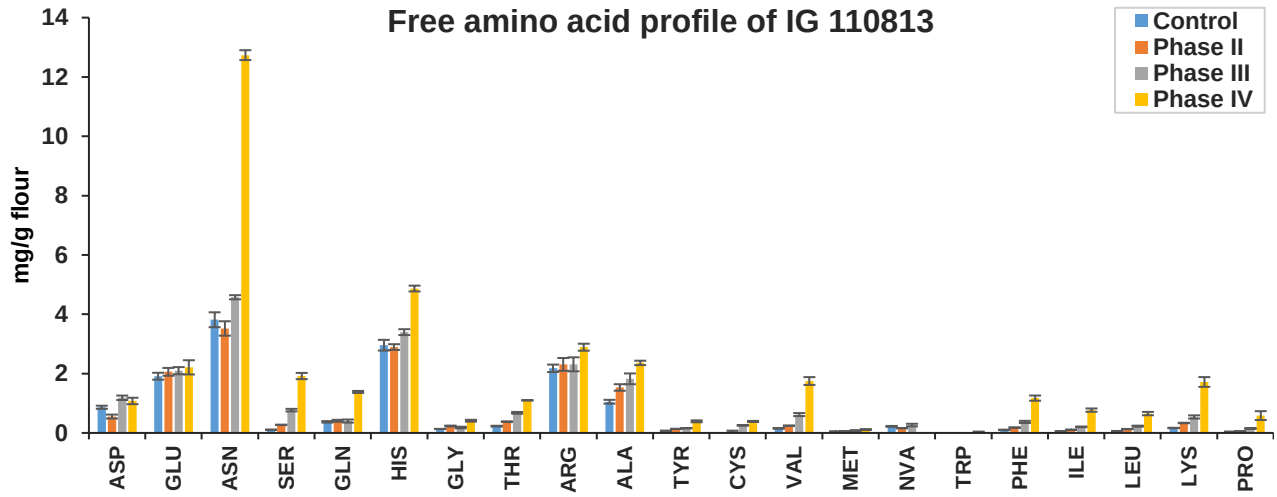


Figure 4.4.5 Individual free amino acids content of IG 110813 (*L. lamottei*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

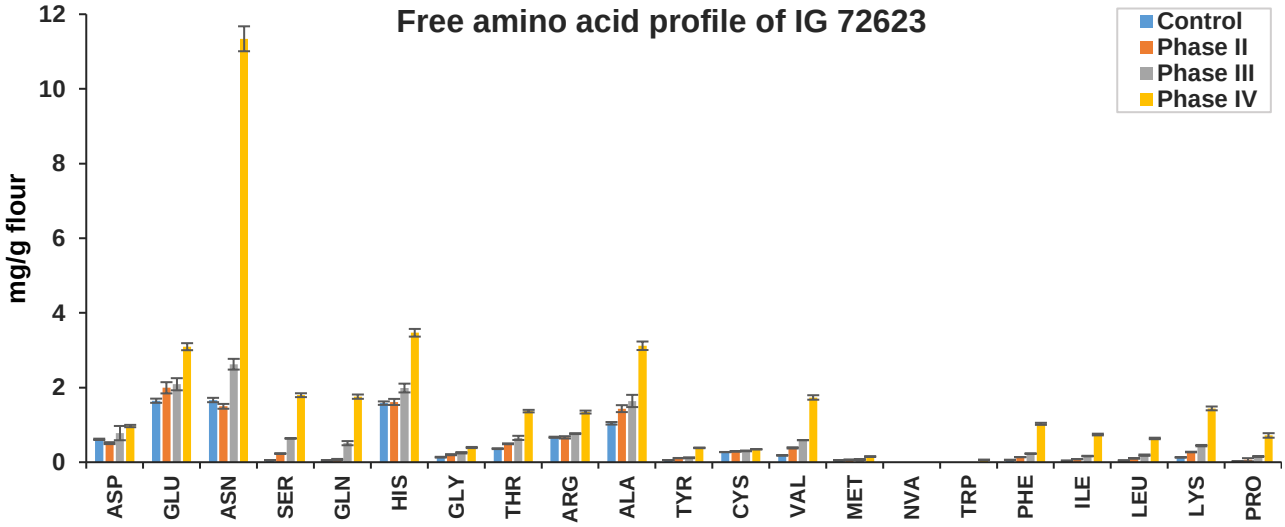


Figure 4.4.6 Individual free amino acids content of IG 72623 (*L. odemensis*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

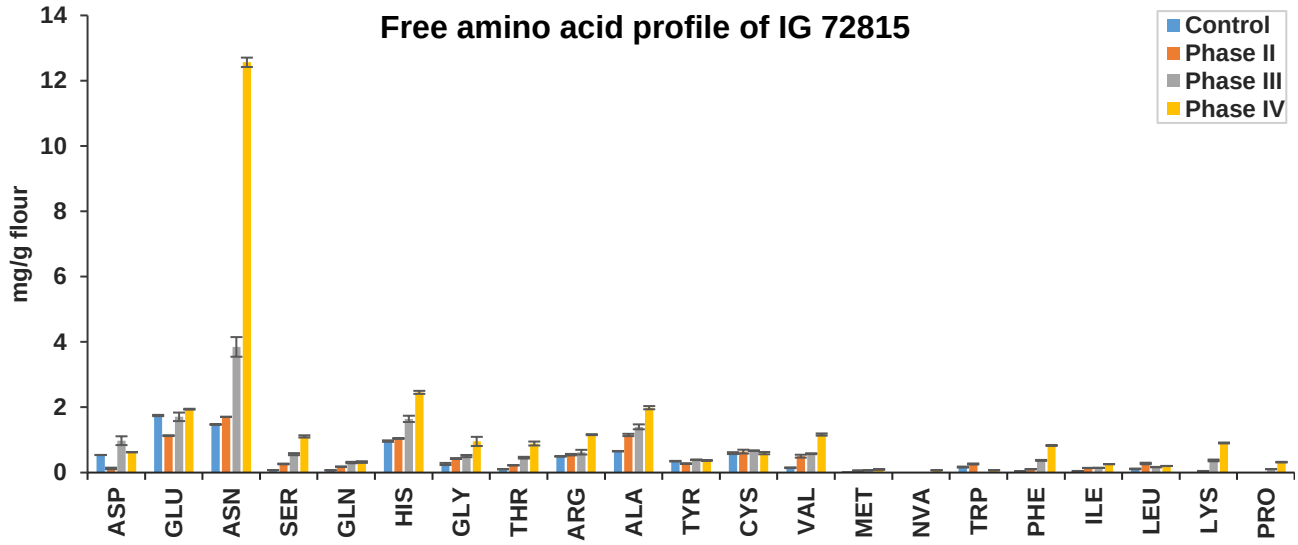


Figure 4.4.7. Individual free amino acids content of IG 72815 (*L. ervoides*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

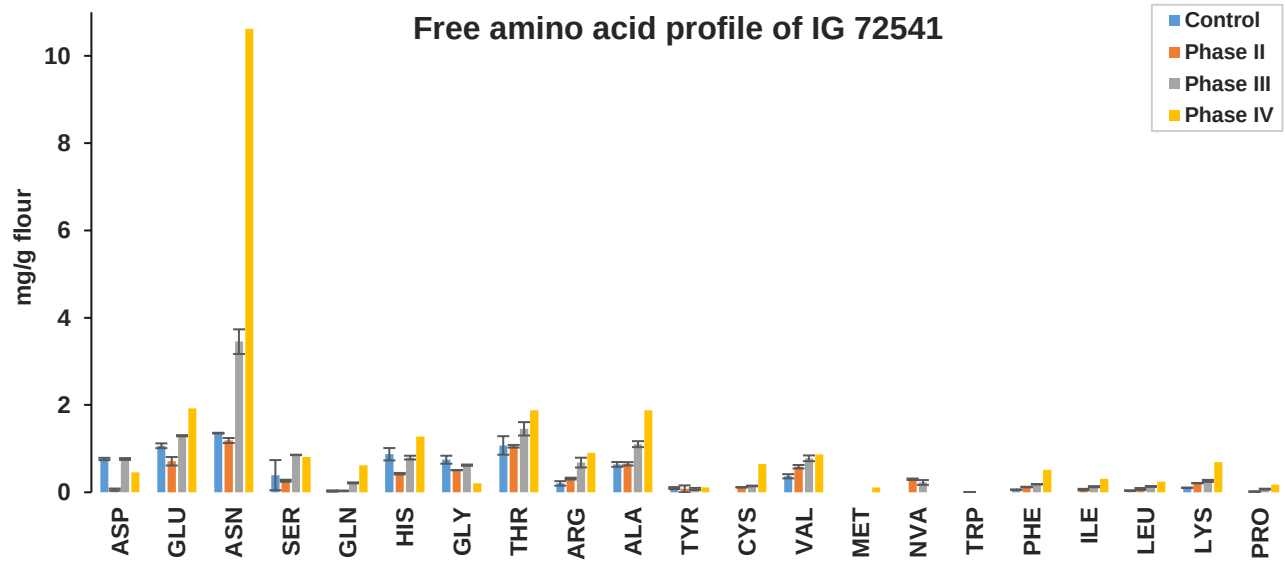


Figure 4.4.8 Individual free amino acids content of IG 72541 (*L. nigricans*) in four phases of seed germination. ASP, GLU, ASN, SER, GLN, HIS, GLY, THR, ARG, ALA, TYR, CYS, VAL, MET, NVA, TRP, PHE, ILE, LEU, LYS, and PRO represents aspartic acid, glutamic acid, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, norvaline, tryptophan, phenylalanine, isoleucine, leucine, lysine, proline, respectively.

However, in CDC Redberry, cysteine levels decreased by around 16.32% from phase III to phase IV. Methionine was present in control seeds of all genotypes excluding IG 72815, BGE 016880, and IG 72541. Unlike cysteine, the level of methionine augmented as the germination progressed in all the genotypes studied. Serine is the amino acid that increased the highest compared to the other FAAs from control to phase II, which was about 1.54-3.60 folds except for IG 72541 (*Lens nigricans*), where it reduced 33% in phase II compared to the control seeds, according to this study.

The other free amino acids that increased at least two-fold between control and phase II were valine (VAL), threonine (THR), leucine (LEU), isoleucine (ILE), proline (PRO), and glutamine (GLN).The most abrupt increase in the level of free amino acids occurred between phase II and phase III. Depending on genotypes the amount of aspartic acid (ASP) rocketed followed by glutamine (GLN) in phase III. In addition, lysine (LYS) and phenylalanine (PHE) increased around 6.9 folds and 2.63 folds in IG 72815; cysteine (CYS) increased 5.17 folds in IG 110813; proline (PRO) and serine (SER) increased 3.49 folds and 2.23 folds in IG 72541; and tryptophan (TRP) increased 6.40 folds in CDC Redberry, respectively. A gradual increase was observed in all the available amino acids while reaching phase IV. Although the increase in amino acids was comparatively lower than that was between phase II and phase III, a higher increase was observed in proline (PRO), phenylalanine (PHE), isoleucine (ILE), lysine (LYS), asparagine (ASN), glutamine (GLN), glycine (GLY), serine (SER), leucine (LEU), valine (VAL), and cysteine (CYS).

4.4 Discussion

Seed germination improves protein quality, which is determined by changes in the amino acid pattern, digestibility, and interference of other biomolecules such as indigestible carbohydrates or inhibitors, and relative proportions as well as availability of different amino acids (El-Mahdy and El-Sebaiy, 1985). This study was performed based on collected samples from the four distinct phases (control, phase II, phase III, and phase IV) to determine changes in protein digestibiligy in terms of protease activity and release of free essential and non-essential amino acids in distinct phases. It was found that the proteolytic activities were enhanced as the seeds progress towards germination. The protease activity analysis of seeds has revealed that the protease activity was the highest during phase III and phase IV, starting from radicle elongation to plumule emergence. The protease activity in dry seeds (control) was higher in cultivated lentils compared to the other wild lentil species. Conversely, the increase in protease activity during phase III and phase IV was higher in wild species compared to the cultivated species (**Table 4.1**).

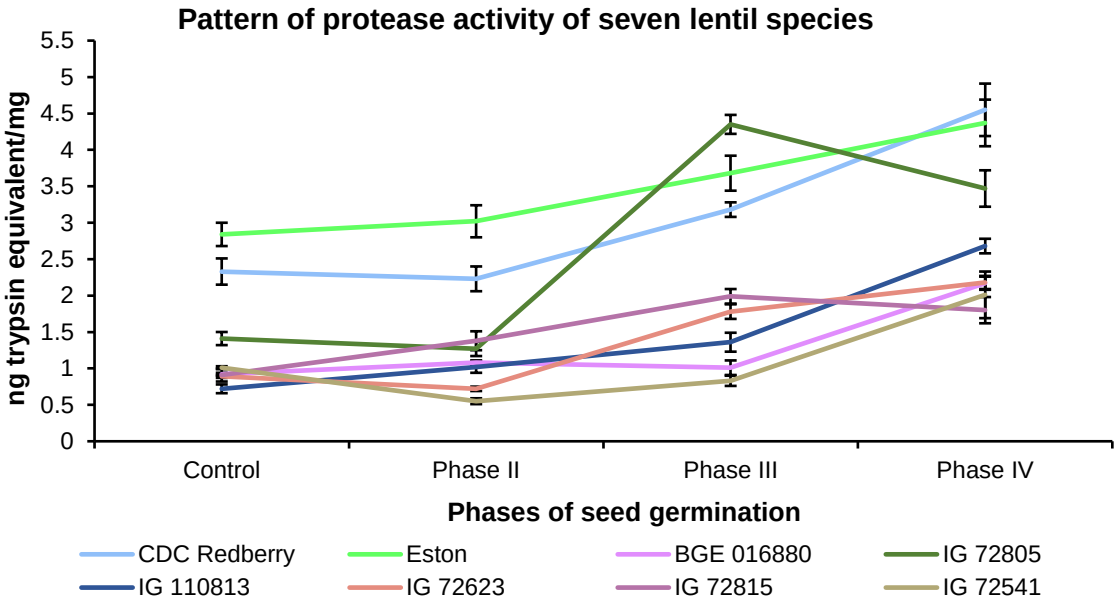


Figure 4.5 Pattern of protease activity in four phases of seed germination in different lentil species. The x-axis represents four phases of seed germination while the y-axis represents ng trypsin equivalent/mg.

A study of lentil proteolytic activity during seed germination showed a rapid increase of amino acids and nitrogen and decreased levels of proteinase inhibitors. An earlier study also showed the

reduction of half of the α-galactosides, as well as a reduction of about 80% of storage proteins and polysaccharides in lentil seeds having sprout lengths of 1-2 cm due to seed germination (Petrova et al., 2010). A study of proteolytic activity during the early stage of germination in lentil seeds revealed that proteolysis cleaves the acidic subunit of legumin and a larger subunit of vicilin improves solubility, foaming capacity, and absorbability (Bamdad et al., 2009a). Proteolytic activity enhances the release of soluble non-protein nitrogen and smaller size peptides resulting in increased solubility in seed (Bamdad et al., 2009a). A previous study revealed that the storage proteins are hydrolyzed to soluble peptides and free amino acids (FAAs) which serve as precursors for the synthesis of new proteins and also contribute to producing nitrogen-containing compounds in seedlings (Marcone and Kakuda, 1999). Studies on the digestibility of green gram, cowpea, lentil, and chickpea reported there was an increased digestibility of starch and protein as well as increased bioavailability of iron and calcium after germination for 24 h (Ghavidel and Prakash, 2007; Nkhata et al., 2018). Overall, there was an observed increase in FAAs and protease activity in phase III and phase IV observed (**Figure 4.5**).

Table 4.1 Percentage change in protease activity from Control to Phase II, Phase II to Phase III, and Phase III to Phase IV of seed germination for eight genotypes of lentil from seven species

Genotype	**Species**	**Control - Phase II**	**Phase II - Phase III**	**Phase III - Phase IV**
CDC Redberry	*L. culinaris*	-4.29%	42.60%	43.082%
Eston	*L. culinaris*	6.34%	21.85%	18.75%
BGE 016880	*L. orientalis*	17.39%	-6.48%	114.85%
IG 72805	*L. tomentosus*	-9.93%	242.52%	-20.23%
IG 110813	*L. lamottei*	41.67%	33.33%	97.059%
IG 72623	*L. odemensis*	-19.10%	147.22%	22.47%
IG 72815	*L. ervoides*	51.65%	44.20%	-9.55%
IG 72541	*L. nigricans*	-45.55%	50.91%	142.17%

With regard to the FAA release during seed germination, the FNEAAs were greater in proportion to FEAAs. In addition, both the FNEAAs and FEAAs were increased in the post-germination phases, and therefore, the level of TFAAs was increased in phase III and phase IV. A study of the amino acid content of seed germination of pea and lentil also reported that there was an increase in FAAs as found in this study (Kuo et al., 2004). Among the FEAAs isoleucine (ILE), lysine (LYS), leucine (LEU), valine (VAL), and threonine (THR) were increased significantly during

germination. Among the genotypes, BGE 016880 showed the highest increase in FEAAs followed by IG 72623 and IG 110813. Phenylalanine (PHE) was one of the FEAAs that increased in all genotypes during germination, excluding CDC Redberry, in which the level of PHE fluctuated among the phases studied, and in BGE 016880, phenylalanine was reduced during phase III, and phase IV of seed germination. A study of changes in FAAs during seed germination of soybean revealed that the content of 23 FAAs, two acidic peptides, and 17 amino acids was increased, after the hydrolysis seeds tyrosine and phenylalanine appeared (Kasai et al., 1966). A study with peas, lentils and beans (*Phaseolus vulgaris*) revealed that there was an increase in GABA (ɣ-aminobutyric acid) (Kuo et al., 2004). Kuo et al. 2004 has also reported no GABA content in any stages of seed germination studied. Another study with five species of lentil revealed obtaining higher GABA content in germinated seedlings (Rozan et al., 2000). A study of amino acids of lentil reported that 38.10% of total amino acids of seed protein is occupied by EAAs (Fouad and Rehab, 2015). This study has also reported an increase in the content of leucine, lysine, phenylalanine, and valine due to seed germination. Another finding from Fouad and Rehab 2015 is that the tryptophan content is lower in lentil seeds compared to the other amino acids. However, the tryptophan (TRP) content was almost missing in all the genotypes studied with exceptions for CDC Redberry and IG 72815, whereas it was present in phase III and phase II as well as control, accordingly. In this study, the level methionine and cysteine increased in phase IV was found. A study of Sudanese lentil cultivars reported an increase in sulfur-containing amino acids (methionine and cysteine) due to sprouting (Sulieman et al., 2008).

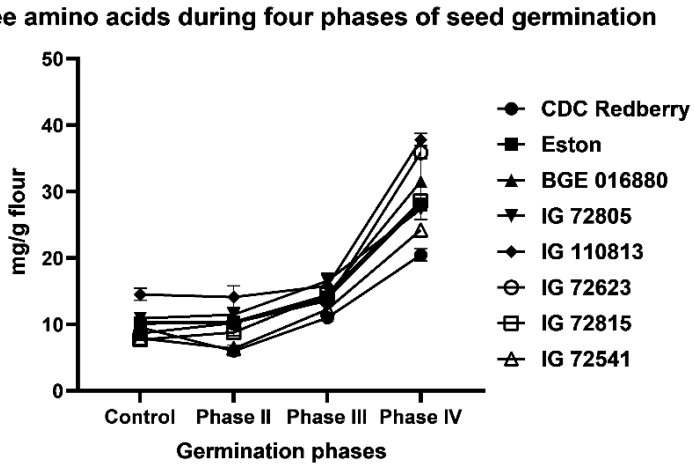


Figure 4.6 Pattern of change in total amino acids during four stages of seed germination of eight genotypes from seven lentil species.

FNEAAs increased at different level in different genotypes among all the phases studied. The level of FNEAAs were also increased in post-germination phases (phase III and phase IV). A study of beans, lentils and peas revealed that the histidine, glutamate, glycine, arginine, tyrosine and tryptophan contents decreased in beans during germination, while the free protein amino acids were increased in lentils and peas post germination (Kuo et al., 2004). These results support with the findings from this study across lentil species. The overall pattern of change in total amino acids was similar (**Figure 4.6**).

The control seeds of each genotype studied was also compared for their amino acid composition. It was found that the total FNEAA and TFAAs was higher among the genotypes from primary and secondary gene pool. For instance, in CDC Redberry, Eston, BGE 016880, IG 72805, and IG 110813 the amount of FNEAAs was 8.43, 8.87, 9.86, 10.11, 13.71 mg/g flour, respectively. With regard to the TFAAs, it was 9.55, 10.12, 10.31, 10.93, 14.52, and 8.66 mg/g flour in CDC Redberry, Eston, BGE 016880, IG 72805, IG 110813, and IG 72623, accordingly. However, the FEAA was comparatively higher among the individuals from primary and quaternary gene pools CDC Redberry (1.12 mg/g flour), Eston (1.25 mg/g flour), and IG 72541(1.66 mg/g flour), respectively.

With regard to the individual free amino acid profile from control seeds, IG 110813 (*L. lamottei*) was found as the best genotype. This species, however, has yet no scope for improvement through traditional breeding as they are not genetically compatible for interspecific breeding. Regarding the individual EAAs, THR was found higher in CDC Redberry (0.84 mg/g flour), Eston (0.84 mg/g flour), and IG 72541 (1.07 mg/g flour). HIS was found higher in IG 110813 (2.96 mg/g flour), Eston (2.13 mg/g flour), and BGE 016880 (2.28 mg/g flour), respectively. The VAL content was the highest in IG 72541 which is 0.37 mg/g flour followed by in IG 72805 that is 0.26 mg/g flour. Another EAA LEU was only found in IG 72541 that is 0.11 mg/g flour. The MET content is higher in CDC Redberry and IG 72623 compared to the other genotypes studied, which is 0.05 mg/g flour. The TRP content was significantly higher in IG 72815 with 0.17mg/g flour. The ILE content was the highest in CDC redberry and LYS was overall present in all the genotypes studied. A study by Rozan et al. 2000 with 5 lentil species revealed the absence or scarcity of Ile, Leu, Lys (trace), Met, Cys, and Phe (trace) in cultivated lentil. This study has revealed the presence of some EAAs in the cultivated lentils, which indicates that breeding can improve the EAAs in cultivated lentil.

However, even if IG 110813 (*L. lamottei*) and IG 72541 (*L. nigricans*) is unable to hybridize with the genotypes from other species, a molecular breeding approach could be an effective tool to improve the essential amino acid content.

A study of amino acids during germination of two varieties of soybean corroborated that there was no significant difference between them in terms of the patterns of free and total amino acids (Kasai et al., 1966). A study of fenugreek seeds also revealed a reduction in soluble protein and increase in net total free amino acids due to increase in production (El-Mahdy and El-Sebaiy, 1985). A study of cultivated lentil seed germination also elucidated that the total amino acids, free essential and non-essential amino acids, protease activity, phenolic compounds, reducing and non-reducing sugars increased upon germination, while oil content as well as antinutritional factors including tannins and phytic acid were reduced (Fouad and Rehab, 2015). This study has also stated that the change is in proteolytic activity was irrespective of pH (El-Mahdy and El-Sebaiy, 1985). The ratio of EAAs to NEAAs was increased in subsequent phases with the progression of seed germination, according to this study.

The release of bioactive peptides is one of the important determinants of protein qualitative changes. Therefore, this study aimed to identify the bioactive peptides that are naturally released due to seed germination process post gastrointestinal digestion. Due to instrumental limitation, the other studies could not be accomplished within the timeframe. Therefore, even if this study was attempted, it could not reach to data acquisition. Hence, a bioactive peptide identification study is suggested to explicitely explain the changes in protein quality due to seed germination.

4.5 Conclusion

Higher protease activity is indicative of higher degree of hydrolysis, and therefore, better digestibility upon consumption. Total amino acids, free essential amino acids, and free non-essential amino acids contents were increased during the germination process. The rate of increase in essential free amino acids was higher than non-essential amino acids, which indicates that higher amounts of bioaccessible essencial amino acids are present in post-germination phases of lentil, especially in phase IV. The increase of individual free amino acids varied greatly depending on genotype and showed some variation among species. Depending on the release of free individual amino acids in phase IV, gentoypes IG 72623 (*L. odemensis*), IG 110813 (*L. lamottei*), and BGE 016880 (*L. orientalis*) were the superior genotypes. The amount of sulfur-containing amino acids was increased in some lentil genotypes during seed germination. Possible molecular and traditional breeding approaches could be taken for the overall genetic improvement with regard to improvement of the profile of essential amino acids. Although all the digestibility factors were not studied, the protease activity and free amino acid release corresponded to the improvement of nutritional quality in terms of bioavailability and digestibility of lentil seed proteins. Therefore, the findings of the study are that germination improved the nutritional quality.

Chapter 5

Comparison of effects of germination on the proteome among the eight genotypes of seven lentil species

5.1 Introduction

Proteins are one of the major accumulated reserves for seeds of legumes, and therefore are one of the staple sources of protein from plant origin (Tahir et al., 2011). With protein content of around 26 % of dry weight, the proteins in legumes are often compared to meat as a dietary supplement. Lentil seeds are currently one of the main sources of dietary protein for over one billion people and provide opportunities for greater use in new plant-based protein foods (Jarpa-Parra, 2018; Khazaei et al., 2019). The seed reserves are also essential for supplying energy and primary components during seed germination and seedling establishment (Bewley et al., 2012).

From a quality perspective, the protein fraction of lentils is dominated by globulins followed by albumins. Lentils are also rich in essential amino acids, although, like most legumes, with low content of the sulfur-containing amino acids (methionine and cysteine) and tryptophan. Lentil proteins are sources of bioactive peptides, the short amino acid sequences encrypted in the native proteins released during digestion that provide health benefits (Marambe and Wanasundara, 2014; Khazaei et al., 2019). One major disadvantage of lentil protein compared to meat protein is the presence of antinutritional factors, including trypsin and Bowman–Birk protease inhibitors. These can negatively affect protein digestibility, and amino acid and peptide bioavailability (Benincasa et al., 2019). Many seed processing techniques, for example, cooking, autoclaving, germination, microwave, irradiation, spray and freeze-drying, fermentation, and extrusion have been shown to improve protein quality by stopping the functionality of proteins and antinutritional factors through hydrolysis (Sá et al., 2019). Germination causes the storage proteins to hydrolyze into peptides and amino acids by proteolytic enzymes, thereby decreasing protease inhibitors and increasing the content and bioavailability of amino acids and bioactive peptides (Benincasa et al., 2019).

The composition of seed reserves differs from species to species (Fenner et al., 2005) and the seed germination and establishment of seedlings depend on the availability of stored reserves in seeds,

including carbohydrates, lipids, storage protein, and other mineral nutrients (Kitajima and Myers, 2008; Soriano et al., 2011). Lentil species are grouped into primary (*L. culinaris, L. orientalis, L. tomentosus),* secondary (*L. lamottei, L. odemensis),* tertiary (*L. ervoides)* and quaternary (*L. nigricans)* gene pools (Wong et al., 2015). A study of protein content involving 72 accessions of different lentil genotypes showed a variability between 10.5 and 23.7% (Kumar et al., 2016). Kumar et al. 2016 reported the range of protein content of dry seeds of *L. culinaris, L. orientalis, L. tomentosus, L. odemensis, L. ervoides,* and *L. nigricans,* respectively, as 22.6-23.1%, 15.7-20.7%, 23.1-23.9%, 20.1%, 12.5-20.5%, and 18.1-22.6% of dry seed.. Wild lentil species have also been studied for their potential as sources of resistance to biotic and abiotic stresses, but relatively little information is available on their differential protein expression through different phases of germination (Dogra et al., 2013; Bai et al., 2021).

Study of the proteome (proteomics) during seed germination of some plant species showed that there is a significant increase in the number of proteins during germination. However, even though the proteome profile study over time after imbibition is available, proteome profile studies after staging individual seed are limited or unavailable. Quantitative analysis of proteomic studies are required to reveal the overall expression of proteins in terms of quantity to obtain a transparent overview of over-expressed and under-expressed proteins during seed germination. In maize, 134 and 191 proteins were identified in dry seed and in 24 hour-imbibed seeds, respectively (Guo et al., 2013). In *Brassica napus,* 50, 23, 14, 11, and 55 differentially expressed proteins were found at 6, 12, 18, 24, and 36 hours after imbibition (Gu et al., 2016). Since all seeds of individual lentil genotypes do not germinate simultaneously after imbibition, in this study germination phases of individual seeds were determined for studying the proteome profile. Genotypes from each species in four distinct germination phases were analysed to allow observation of comparative differences in the proteome, or differentially expressed proteins during germination through a proteome profile study.

5.2 Materials and methods

5.2.1 Plant material and experimental design

Eight selected genotypes representing the 7 lentil species were used for this experiment. The selected genotypes included the two cultivated genotypes CDC Redberry (*L. culinaris*, 2018 entry S-99), Eston (*L. culinaris*, 2018 gw. 103), and one genotype of each of the wild species: BGE 016880 (*L. orientalis*, 2018 wild parent entry 31), IG 72805 (*L. tomentosus*, 2018 entry 20), IG 110813 (*L. lamottei*, 2018, entry S-77), IG 72623 (*L. odemensis*, entry 10), IG 72815 (*L. ervoides*, wild parent increase entry 1), and IG 72541 (*L. nigricans*, wild line 97). For each genotype, samples were collected at three germination phases II, III, and IV) and control samples were the intact seeds washed with Mili-Q water. Samples were collected using three biological and two technical replications. From each genotype, therefore, a total 24 [(phases) 4 x (biological replicate) 3 x (technical replicate) 2] samples were collected for analysis.

5.2.2 Experimental procedure

5.2.2.1 Sample collection

The samples were collected following the methods described in chapter 4.

5.2.2.2 Sample processing

5.2.2.2.1 Drying and grinding samples

The samples were dried and ground following the previous method (details in chapter 4).

5.2.2.2.2 Extraction of proteins from seed samples

For protein extraction from seed samples, 30 mg of the powdered sample in a 2 mL microtube was mixed with 500 µL of extraction solution at 1200 rpm in an Eppendorf ThermoMixer C) for 5 min at 95°C. The extraction solution was prepared by mixing 4% sodium dodecyl sulfate (SDS) with 100 mM tris/HCl at pH 7.6. After cooling to room temperature, this mixture was sonicated for 5 min followed by centrifuging at 16,000 g for 30 min (20 °C); the supernatant was collected into a new 2 mL microtube. Then the pellet was re-suspended with 300 µL of extraction solution by vigorous vortexing and the centrifugation step was repeated to collect the supernatant to make the final volume of 800 µL solution containing lentil seed proteins.

5.2.2.2.3 Determination of protein quantity

A BCA assay (bicinchoninic acid assay) was performed for measuring protein concentration. The BSA (bovine serum albumin, ThermoFisher, Rockford, USA) was used as the standard and the standard curve was prepared by evaluating 0,100, 250, 500, 750, 1000, 1250, and 1500 µg/mL concentration of BSA in three replications of each concentration. From these samples, 15 and 30 times diluted samples of each diluted sample were poured in triplicate into 96 well plates. 200 µL of pierce BCA protein assay reagent A (ThermoFisher, Rockford, USA) and B (ThermoFisher, Rockford, USA) in 50:1 ration was poured into all the standards and samples. The plate was immediately covered and put in an incubator at 37°C for 30 min. The protein concentration was measured using a BioTek Synergy HT plate reader (Gen5 version 9.2). The mean concentration of the three time applications of two different dilutions (15X and 30X) was determine by measuring 6 samples per replicate.

5.2.2.2.4 Hydrolysis of protein samples

Based on the concentrations obtained from BCA assay, samples were diluted with Milli-Q water to 1 µg/µL concentration, and 50 µg of protein was poured into a 1.5 mL microtube with 10 mM of DTT (Dithiothreitol, ThermoFisher, Canada) and incubated at 75 °C for 30 min in an eppendorf thermomixer c. This solution was then transferred onto the top of 3 KDa filter inside the centrifuge tube for digestion. 200 µL of 8 M urea in 0.1M Tris/HCl at pH 8.5 was poured on the top of the filter and the lid was closed. This mixture was then mixed gently and centrifuged at 14000 rcf for 15 min. The centrifugation was repeated after adding 200 µl of 8 M urea in 0.1 M Tris/HCl (pH 8.5). Following that, 100 µl of IAM (iodoacetamide, Sigma-Aldrich, Saint Louis, USA) solution was poured on top of the filter and incubated for 30 min in the dark without mixing, and then centrifuged at 14,000 g for 10 min. Following that, 100 µL of 8 M urea (ThermoFisher, Ottawa, Canada) in 0.1M Tris/HCl (pH 8.5) was poured and centrifuged at 14,000 g for 15 min; this step was repeated two times. Then, 100 µl of 500 mM ABC was poured into the same filter and centrifuged at 14,000 g for 10 min; this step was also repeated two times. Then 40 µl 50mM ABC with 0.4% SDC (sodium deoxycholate, Sigma-Aldrich,) was poured on top of the filter membrane. Trypsin (sequencing grade modified trypsin,Promega, V511A) was added into the mixture at the ratio of protein to trypsin at 100:1 and that mixture was then mixed at 800 rpm for 1 min at 37°C, followed by incubated for 16 h in a wet chamber at 37 °C. Following this step, the filter was

transferred into a new collection tube and centrifuged at 14,000 g for 10 min. Then 60 µL of 0.5 M NH_4HCO_3 (ABC) was added and centrifuged again at 14000 g x 10 min. Then, on top of the same filter, 6 µL of pure formic acid was added and centrifuged at 14,000 g for 10min.

5.2.2.2.5 Purification of peptide samples

The filtrate was then collected for peptide desalting. Oasis HLB 1cc (10 mg) extraction cartridges (Ireland) were used for desalting. The cartridges were first rinsed twice with 1 mL acetonitrile followed by two rinses using 1 mL 0.2% FA in water. The sample was then loaded into the cartridge. Samples were then rinsed three times with 1 ml of 0.2% FA in water, and then again rinsed with 1 ml of water. Finally, 250 µl of 70% ACN (acetonitrile, Thermofisher, LC-MS grade) with 0.1% FA was poured to elute peptides to a 1.5 mL microtube. This peptide mixture was then dried in a SpeedVac for 6 h.

5.2.2.3 LC-MS analysis

For LC-MS analysis, all the reagents used were of LC-MS grade from Thermofisher optima. The samples were reconstituted with 50 µL of 2% can with 0.1% FA in water in 1.5 ml tubes, then vortexed for 3-4 min and sonicated for 5 min. This mixture was centrifuged for few seconds and 20 µL of the mixture was transferred into the LC vials prior to loading through autosampler. Solvents A and B, respectively, at 0.01% FA in water and 80% ACN in 0.1% FA were used; 0.1% FA in water was used for the loading pump. For the seal wash, 10% methanol in 80% ACN was used.

Chromatographic separation was performed by C18-reverse phase column. Samples in vials were then loaded onto the nanoLC for peptide separation through elution at different gradients of solvents. The solvent flow rate was at 0.3 µL/min. The initial loading phase was for 5 min with 2.5% of solvent B and 97.5% of solvent A combination. After that for next 80 min solvent B was increased linearly to 25% followed by a comparatively higher increase in concentration of B (40%) in next 10 min. Following that, the ration of solvent B was increased abruptly to 95% at 0.35 µL/min flow rate, and this gradient was steady for 5 min. At the last phase, solvent B was rapidly reduced to 2.5% in 5 min. After that, 30 min of gap was maintained between two samples to run for washing and equilibration. A solvent blank (0.1% FA in water) run was performed between

samples from two different phases, while there was no solvent blank used between the replicates of sample phase.A NanoLC instrument (Thermo Scientific™, Germany) was coupled with Q Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Nanospray Flex™ ion source was used where ESI was the method of ionization. A fullMS followed by ddMS^2 approach was used for peptide identification. Precursor ions were scaned through data dependent acquisition (DDA) method, in which scan was acquired at a 70K resolution between 350 and 2000 m/z. In precursor ion data acquisition, AGC target was set to 1e6 with the maximum IT 30ms per individual scan window for the samples studied. The precursor ion then was fragmented through HCD (high collision dissociation) with normalized collision energy (NCE) of 27 MeV. In the MS/MS was performed 17.5K resolution with 5e4 AGC target and maximum IT of 64 ms per microscan. TopN was selected as 10 based on resolution and scan speed. In addition, loop and MSX count were 10 and 1, respectively.

5.2.2.4 Data processing

Raw data acquired through Xcalibur was processed through Mascot Distiller (Matrix Science, v2.7.1.0). The peak list found through Macot Distiller was analyzed through Mascot Daemon (Matrix Science, v2.7) for identifying proteins from the peptide fragments. The peaks were screened for monoisotopic mass with 5 ppm deviation in precursor ion mass. In addition, mass tolerance for MS/MS was ±0.05 Da with 2+, 3+, and 4+ charge. Digestion with trypsin with maximum 2 missed cleavage was allowed to find protein identities. SwissPort and Pride_contaminants were the databases used for identifying the proteins within each sample. For better accuracy of obtaining protein identities, carbamidomethylation was selected as fixed modification, while oxidation and phosphorylation was selected as variable modifications. Outputs file from Mascot Daemon was generated with 5% level of significance ($p<0.05$). Further analysis was performed through MapMan ontology (https://mapman.gabipd.org/mapman).

For quantification, the raw file from Xcalibur was analyzed in Mascot Daemon with scans procecessed throught Mascot Distiller. precursor ion m/z tolerance was 20 ppm, while for MS/MS the tolerance was 0.2 Da. The average of three most abundant peptide intensities that reached

above the threshold of chromatogram was exported in excel file format for each biological replicate.

5.2.2.5 Data representation

As there was lack of reference sequences for lentil protein in the annotation software, *Viridiplantae* taxonomy was selected to obtain the maximum number of protein homologs from various plant species. The output file was then exported into an Excel file to manually screen individual protein homologs from different species. A Venn diagram was prepared from the protein identities found through primary screening for the representation of identified proteins from distinct phases of seed germination. Proteins categorized through MapMan in different functional categories were also screened for comparing protein homologs present in different phases. For the representation of the functional annotation graph, proteins involved in each subcategory were counted for individual samples, and the mean value of all category totals was taken for preparing the annotation graph with the proportion of proteins from each category. The identification output from Mascot Daemon was also annotated through KEGG (Kyoto Encyclopedia of Genes and Genomes) for explaining major pathways within the cell in the seed germination phases. As there is no lentil proteome enlisted in the KEGG server, other common legume species including *Glycine max, Glycine soja, Lotus japonicas, Arachis ipaensis, Lupinus angustifolius, Medicago truncatula, Arachis duranensis, Vigna unguiculata* non-legume species *Arabidopsis thaliana, Arabidopsis lyrata, Capsella rubella, Camelina sativa, Eutrema salsugineum, Brassica rapa, Brassica napus, Brassica oleracea, Oryza sativa japonica and Zea mays* were used for obtaining protein homologs involved in a specific pathways.

5.2.3 SDS-PAGE experiment

Protein was extracted from phase 0, phase III, and phase IV samples at pH 8 and pH 8.5 following the procedure in chapter five. The extracted proteins were quantified by Bio-Rad DC Protein Assay. 100 µl of sample or BSA standard (BSA standards for preparing standard curve) was taken into test tube with 500 µl of reagent A and 4 ml of reagent B and vortexed immediately. This mixture was then kept for 15 min and 500 µl was poured into the cuvette for absorbance reading at 750 nm in a microplate reader (Thermo Scientific Multiskan GO). The standard curve was prepared using BSA (sigma Aldrich, Saint Louis, USA) concentration range from of 12.5 µg/100

μL, to 200 μg/100 μL. Each sample was diluted 2 and 5 times with two replicate in each dilution for the detection of absorbance. SkanIt software v4.1 was used to quantify the protein based on the standard curve absorbance of BSA standards.

For runing the samples into gel, 100 μg of protein from samples was taken into individual 0.5 mL appendorf tubes. 100 μL of 2X laemmli sample buffer (Bio-Rad, Canada) containing 65.8 mM Tris-HCl, pH 6.8, 2.1% SDS, 26.3% (w/v) glycerol, 0.01% bromophenol blue was added. Following that 5% 2-mercaptoethanol (v/v) was added to the protein mixture and MiliQ water was added to make the total volume upto 100 μL. The mixture was then vortexed and thoroughly mixed for 5 min at 95°C. For large gel electrophoresis, 40 ml of 10% resolving gel was prepared with 13.3 mL of 30% Acrylamide/Bis-acrylamide solution (Bio-Rad). 10 mL 1M Tris and 16.28 mL of distilled water was added to the solution. The mixture was then agitated gently and just before casting, 20 μL and 200 μL of TEMED and 10% APS was added for polymerization. For casting large gel, two glass plates were installed perpendicular to the stand floor after adjusting the floor with balancing bubble. Rubbers were set below the glass plates and clamps were set at the two edges of the glass plates. Resolving gel was kept for 25 min to polymerize. Then the 15 ml of 6% stacking gel was prepared following the same procedure and kept for 30 min to polymerize. Combs were place on the top just after pouring the stacking gel. The pH of resolving gel was 8.8, while the pH of stacking gel was 6.8. The mini gel (Bio-Rad) was also casted following similar method using same ratio of chemicals. Both the mini and large gel after polymerization was place in the respective running chambers. The running buffer containing 0.1M Tris-HCL, 0.1M tricine, 0.1% SDS at pH 8.3 was poured into the chamber. 50 μL of samples was then poured into gel well. The large gel (Bio-Rad) run conducted at 15°C temperature with 12 mAh voltage for 16 h, while for the mini gel (Bio-Rad) 5 μL of samples were loaded into the well and was run at 120 V for 175 min. For both gel run 250 kDa to 2 kDa marker ladder (Precision Plus Protein™ Dual Xtra Prestained Protein Standards, Bio-Rad) of same amount as sample was poured into a well. After the complete run, the large and mini gels were taken out and placed on a container containing 12% TCA in a shaker for 15 min. After that, the gels were soaked into 0.1% Coomassie Brilliant Blue R-250, 40% methanol, and 10% acetic acid in a shaker. After gels were taken out after 2 h and then soaked overnight with 40% methanol, and 10% acetic acid solution in a shaker. After that picture was taken for gels in gel documentation chamber (ChemiDoc MP, Bio-Rad).

The protein of interest was cut and transferred into an appendorf tube. For in-gel digestion, gel slice was cut into small pieces (1-2 mm) with the help of sterile sharp knife and placed into 1.5 mL tubes. 25 mM NH4HCO3, 50% ACN solution was poured to cover all the slices and vortexed for 10 min. then the supernatant was discarded. This cleaning steps was repeated two times. The gel pieces were dried completely in a Speed Vac. Then 25 µL of 10 mM DTT in 25 mM NH4HCO3 was added to dried gels followed by vortexing and spinning was performed for proper mixing. This reaction was kept for 30 min 56°C in a mixer (Eppendorf thermomixer). After that the supernatant was removed and 25 µl of 55 mM IAM was added to the gel pieces followed by vortexing and spinning was performed for proper mixing, and kept in the dark for 30 min at room temperature. Then the supernatant was removed and ~100 µl NH4HCO3 was added and mixed properly. The supernatant was then removed and ~100 µl NH4HCO3 in 50% ACN was added and mixed thoroughly. After removing the supernatant and following this step twice, the gel pieces were dried in a Speed Vac. For digestion, 15 µL trypsin solution was added and mixed thoroughly. After a short spinning of the suspension, it was kept at 37°C in an incubator for 16 hours. For the extraction of peptides, 30 µL of 50% ACN/5% formic acid was added to the solution. This suspension was thoroughly mixed by frequent vortexing and sonication and the peptide solution transferred to a clean new tube. This extraction steps was repeated and peptides were accumulated. Then again this peptide solution was dried in a Speed Vac and final volume was made to 10 µL. Purified peptides were eluted through C18 ZipTip (Millipore) using 5% formic acid solution. This purified peptide solution was again dried and re-constituted into 0.2% FA (5% ACN) solution for LC-MS acquisition. LC MS analysis and data processing was performed same as explained in sections **5.2.2.3** and **5.2.2.4**.

5.2.4 Statistical analysis

The statistical analysis for BCA assay was performed using online version SAS (Statistical Analysis System). The quantitation values of protein intensities for *L. nigricans* obtained through Mascot were analyzed through MetaboAnalyst 5.0.

5.3 Results

Table 5.1 Protein concentration of CDC Redberry, Eston, BGE 016880, IG 72805, IG 110813, IG 72623, IG 72815, and IG 72541 from BCA assay in four phases of seed germination. The uppercase letters indicate the variation among genotypes and the lowercase letters indicate the variation among the phases

Genotype	Protein concentration in four phases of seed germination (µg/mL)			
	Phase 0 (control)	**Phase II**	**Phase III**	**Phase IV**
CDC Redberry	$9.39\pm0.78^{C,a}$	$9.09\pm0.74^{C,b}$	$8.53\pm0.58^{C,b}$	$8.36\pm0.96^{C,c}$
Eston	$9.13\pm0.41^{A,a}$	$8.78\pm0.98^{A,b}$	$8.86\pm0.48^{A,b}$	$8.33\pm0.36^{A,c}$
BGE 016880	$10.72\pm0.27^{C,a}$	$10.33\pm0.44^{C,b}$	$9.91\pm0.32^{C,b}$	$8.81\pm0.25^{C,c}$
IG 72805	$10.62\pm0.35^{D,a}$	$9.38\pm0.21^{D,b}$	$9.19\pm0.34^{D,b}$	$8.19\pm0.45^{D,c}$
IG 110813	$13.12\pm0.17^{B,a}$	$11.12\pm0.18^{B,b}$	$12.27\pm0.25^{B,b}$	$12.27\pm0.26^{B,c}$
IG 72623	$10.13\pm0.21^{E,a}$	$9.65\pm0.23^{E,b}$	$11.52\pm0.17^{E,b}$	$9.60\pm0.27^{E,c}$
IG 72815	$14.23\pm0.16^{E,a}$	$12.20\pm0.10^{E,b}$	$10.63\pm0.14^{E,b}$	$11.96\pm0.08^{E,c}$
IG 72541	$11.38\pm0.22^{D,a}$	$11.73\pm0.52^{D,b}$	$10.27\pm0.28^{D,b}$	$09.24\pm0.14^{D,c}$

Table 5.1 shows the protein concentrations obtained from BCA assay from phase 0 or control seeds, and for seeds in phase II, phase III and phase IV of germination. The results showed there were significant variation among the genotypes, phases and their interactions ($p \leq 0.0001$, **Appendix E**). The F value for genotypes, phases and their interactions was 70.21, 16.62, and 5.36, respectively (**Appendix E**). Overall, in phase IV samples, the concentration of protein was reduced in samples compared to phases II and III, and control seeds (**Table 5.1**). In addition, in the control seeds (phase 0) the protein concentration was the highest compared to all the three phases studied. The highest protein concentration found among all the genotypes and phases was in phase 0 of IG 72815 (*L. ervoides*), which was 14.23 µg/ml, while the lowest protein concentration observed in IG 72805 (*L. tomentosus*), which was 8.19 µg/ml in phase IV. The highest amount of reduction in protein concentration was observed in IG 72805 (*L. tomentosus*) around 23% in phase IV compared to the control seeds. Moreover, the amount of protein concentration among different phases of seed germination showed that the protein concentration was decreasing as germination progressed. However, in IG 72541 (*L. nigricans*) an increase in protein concentration was observed in phase II compared to phase 0, while in IG 110813 (*L. lamottei*) there was an increase observed in phase III compared to phase II (**Table 5.1**).

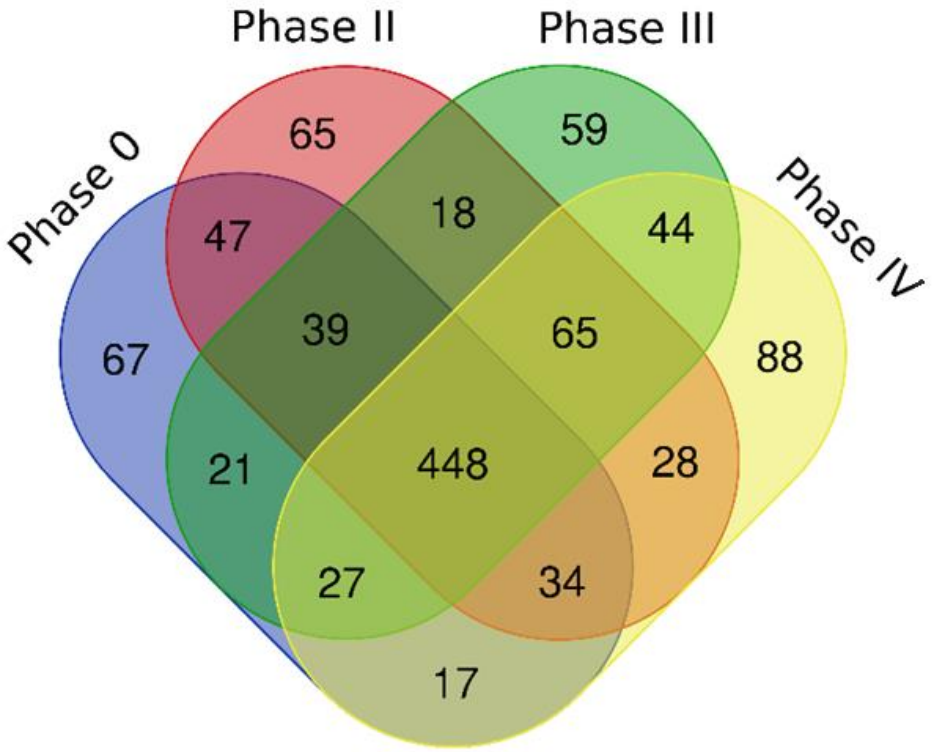


Figure 5.1 Venn diagram representation of the comparison and contrast of proteins identities involved in different biological processes in phase 0 (control), phase II, phase III and phase IV of *L. nigricans* during germination.

The Venn diagram (**Figure 5.1**) represents the common and uniquely identified protein sequences of *L. nigricans* among four different phases of seed germination. This Venn diagram was prepared by accumulating protein identities that were present in at least 50% of the samples in each phase. The screened protein identities were 700, 744, 721, and 751 in phase 0 (control seeds), phase II, phase III, and phase IV, respectively (**Figure 5.1**). This shows that in phase IV, the highest number of unique proteins identified (88 proteins), followed by phase 0 where 67 unique proteins were identified. A total of 448 protein identities were common among all the phases found in the Venn diagram. Seventeen proteins were found in common between phase 0 and phase IV, the lowest number of common proteins.

Upon further screening, proteins involved in DNA damage response (ADP-ribosyltransferase), plastidial fatty acid synthase activity (ketoacyl-ACP synthase III), phenolics biosynthesis (chalcone synthase CHS), protein modification (methyltransferase), protein translation (isoleucine-tRNA ligase, poly-P/G elongation factor, ribosome recycling factor), glycinebetaine biosynthesis (betaine-aldehyde dehydrogenase), vesicle trafficking (SNARE membrane fusion

complex), glycolysis (phosphofructokinase), and mitochondrion/peroxisome binary fission were found common between phase 0 and phase II. Similarly, between Phases III and IV, cytokinesis (callose synthase), and alanine-tRNA ligase; between phases 0 and III, pectin modification and degradation of bifunctional alpha-L-arabinofuranosidase and beta-D-xylosidase); between phases 0 and IV, salvage pathway (adenine phosphoribosyltransferase APT) and fermentation (alcohol dehydrogenase ADH) related proteins were found common. Similarly, in Phase II and Phase III Serpin protease inhibitor protein was present. In addition, in Phase III and Phase IV protein responds to light.red/far red light, cysteine-type peptidase activities, serine-type peptidase activities, sucrose degradation, glyoxylate cycle (malate synthase), hydrolysis (hydrolases), coenzyme metabolism (S-adenosyl methionine SAM), tetrahydrofolate interconversion, cell wall organization (callose synthase), and photorespiration (serine hydroxymethyltransferase) were present.

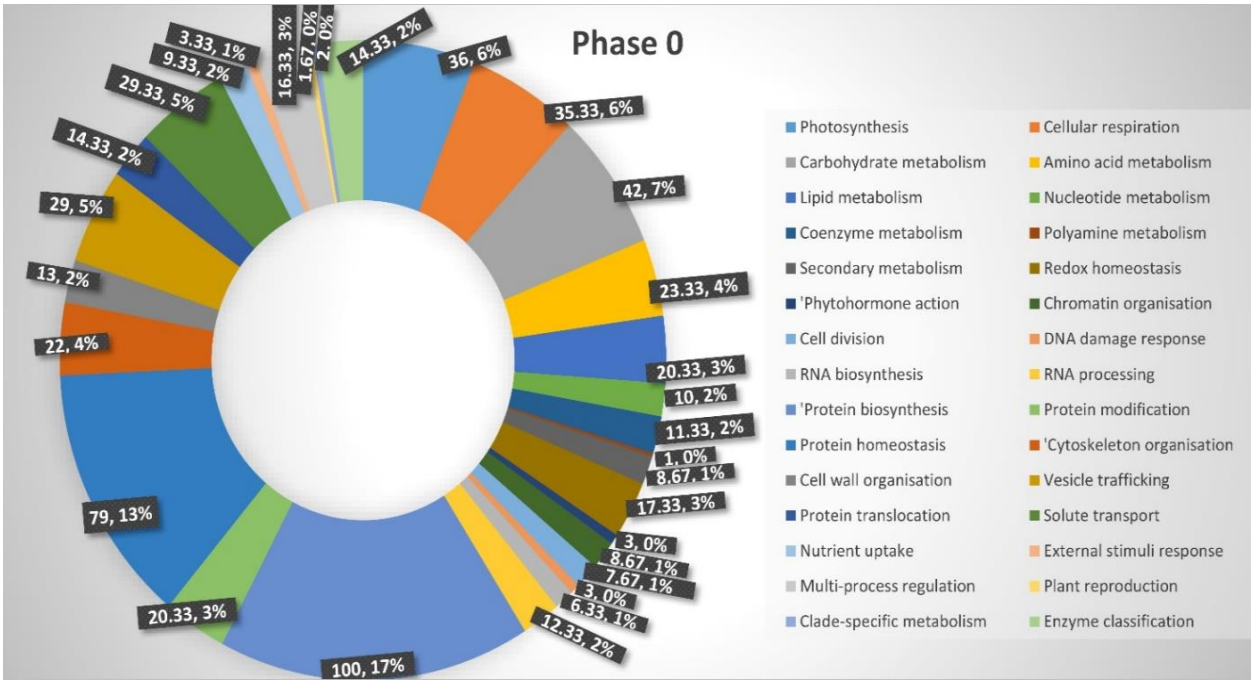


Figure 5.2 The diagram represents the number and proportion of proteins involved in each biological process in intact seeds of the identified proteins of *L. nigricans* during germination.

The total non-redundant proteins identified through functional annotation in phase 0, phase II, phase III, and phase IV were 594, 527, 487, and 535, respectively (**Appendix F**). A total of 30 functional categories of proteins were accumulated to prepare the diagram with the percentage of protein identities (with numbers) found under each functional category (**Figure 5.2-5.5, Appendix F, Appendix G**). As the protein identities were searched against different plant species, the result generated the same protein from multiple species. Therefore, proteins involved in each category and subcategory were screened, and the percentage shown in the diagram is the percentage based on the total number of proteins identified in each category of each phase after screening. The numbers of protein homologs identified were higher in number in the control seeds among all the phases studied for the categories involving photosynthesis, carbohydrate metabolism, coenzyme metabolism, protein biosynthesis, protein translocation, protein homeostasis, cytoskeleton organization, cell wall organization, vesicle trafficking, solute transport, nutrient uptake, and plant reproduction. In addition, in phase II, the number of protein homologs involved in amino acid metabolism, RNA biosynthesis, and protein modification was increased, and was higher compared to all other phases. Similarly, in phase III, phytohormone action and chromatin organization related proteins increased in number in contrast with the other phases. In phase IV, cellular respiration, lipid metabolism, secondary metabolism, redox homeostasis, and various enzymes were higher in number than in the other phases. Among the total number of protein homologs identified, the proportion of protein homologs identified in protein biosynthesis and protein homeostasis was 17% (100 proteins) and 13% (67 proteins), respectively, which were the highest among all the categories. In contrast, the lowest number of proteins identified are 1, 1, and 2 in polyamine metabolism, plant reproduction, and in clade-specific metabolism, respectively.

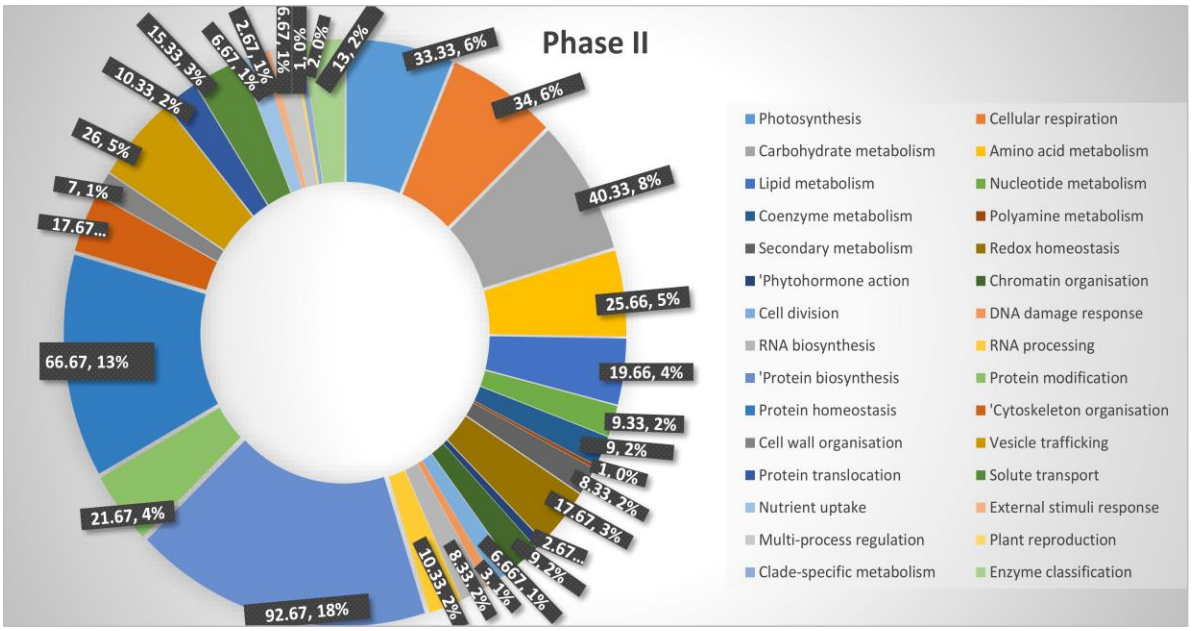


Figure 5.3 The diagram represents the number and proportion of proteins involved in each biological process in germination phase II of the identified proteins for *L. nigricans*.

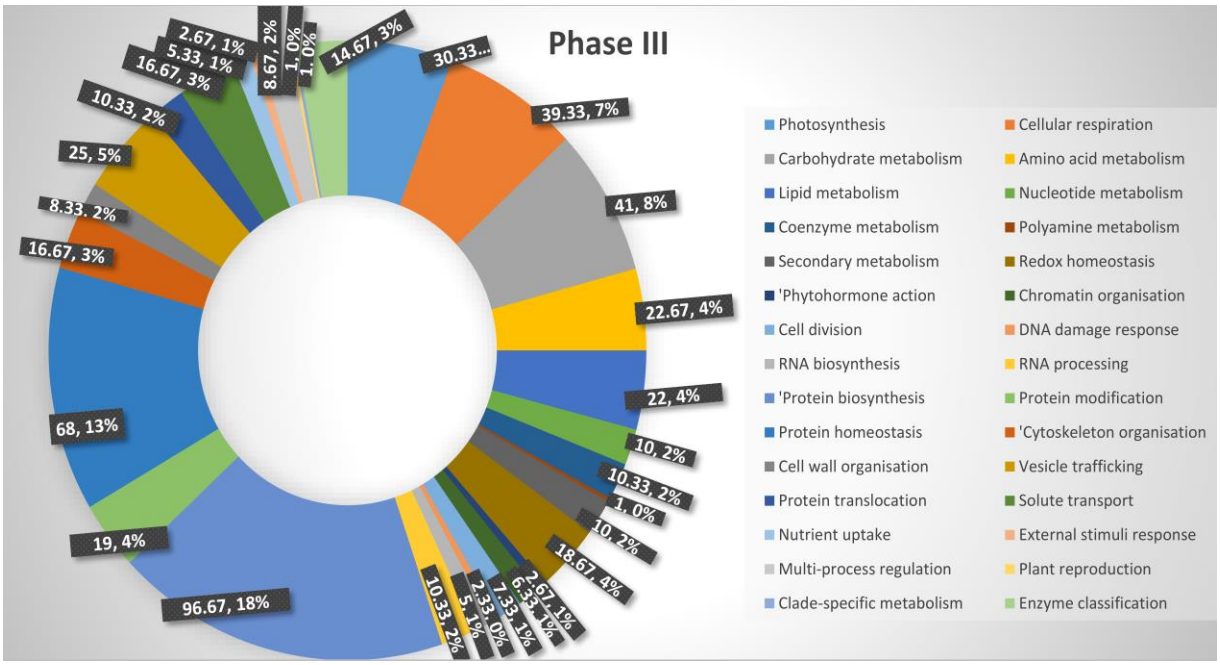


Figure 5.4 The diagram represents the number and proportion of proteins involved in each biological process in germination phase III of the identified proteins for *L. nigricans*.

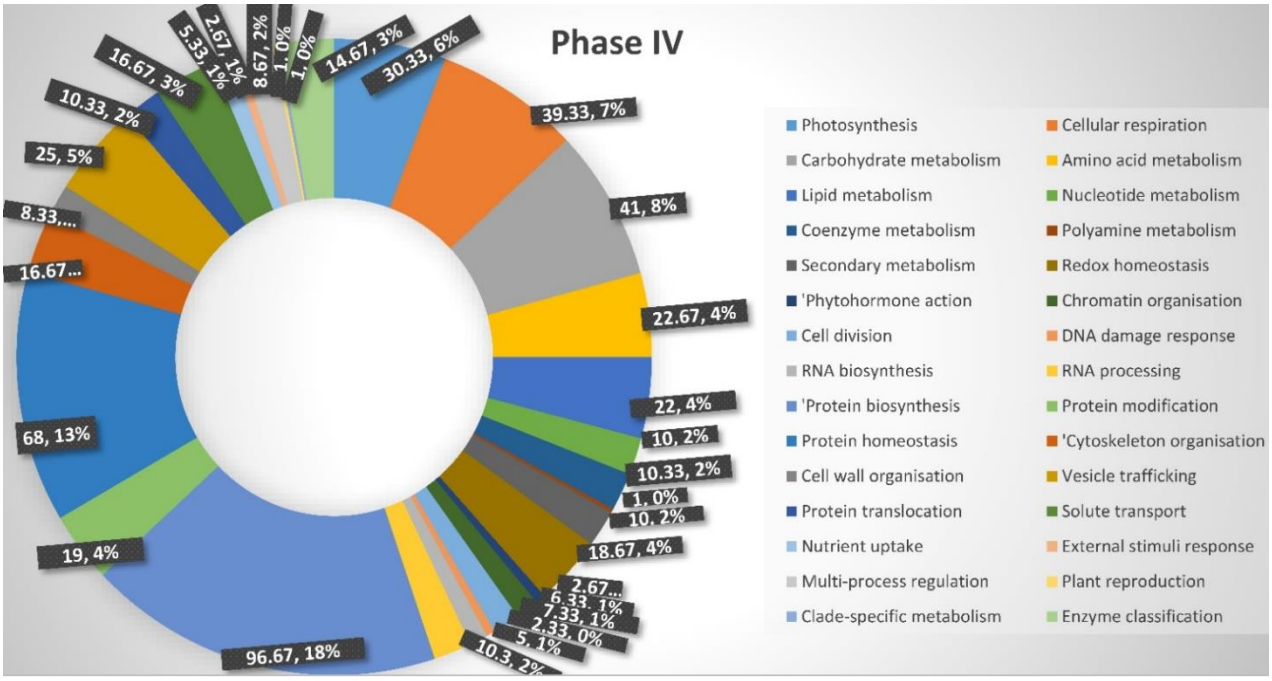


Figure 5.5 The diagram represents the number and proportion of proteins involved in each biological process in germination phase IV of the identified proteins of *L. nigricans*.

The obtained protein sequences were compared among the phases of seed germination. The functional categories included photosynthesis, cellular respiration, carbohydrate metabolism, amino acid metabolism, lipid metabolism, nucleotide metabolism, coenzyme metabolism, polyamine metabolism, secondary metabolism, redox homeostasis, phytohormone action, chromatin organization, cell division, DNA damage response, RNA biosynthesis, RNA processing, protein biosynthesis, protein modification, protein homeostasis, cytoskeleton organization, cell wall organization, vesicle trafficking, protein translocation, solute transport, nutrient uptake, external stimuli response, multi-process regulation, plant reproduction, clade-specific metabolism, and enzyme classes. **Table 5.2.1** to **Table 5.2.5** shows the number of unique proteins found in each phase of seed germination. The number of proteins involved in multi-process regulation was the highest in phase 0 and reduced to half during the other phases (**Appendix C**). The number of solute transport and protein homeostasis related proteins were also reduced in all the phases studied compared to the control. With regard to individual proteins, 54 proteins from 20 different pathways were identified as differentially expressed among different phases of seed germination, but not in dry seed (control or phase 0). For representation, protein identities from each sample was annotated from the processed output file with a 5% level of

significance. For the representation of uniqe protein, a protein that is expressed in distinct phase(s) among atleast two of the four samples was selected. In addition, the protein that is only found in one specific phase was reffered as unique protein for that particular phase. Then protein identities were expressed in at least 50% of the samples were represented in the tables (**Table 5.2.1**, **Table 5.2.2**, **Table 5.2.3**, **Table 5.2.4**, **Table 5.2.5**). All the photosynthesis related proteins were expressed in each phase including the control except NADH dehydrogenase subcomplex M (ndhf) which was found in phase III and phase IV. Hexokinase enzyme, (involved in sucrose degradation to glucose and fructose) as well as plastidial phosphoglucose isomerase (involved in starch biosynthesis) were found in phases II, III and IV. However, the protein involved in raffinose biosynthesis was present in phase II and phase IV, but absent in phase III. Among the proteins expressed for protein synthesis, component eif2-beta, eif-iso4f unwinding complex (component eif-iso4g), eif2-alpha component, isoleucine-tRNA ligase, and pre-60S ribosomal subunit nuclear export factor (ARX1) were found in all of the three phases studied. The ribosome recycling factor (RRF) was found in phase II and mitochondrial tRNA import factor (TRIC) was found in both phases II and III, respectively. With regards to fatty acid metabolism, malate synthase, ATP-dependent citrate lyase heterodimer alpha chain, sterol delta24 reductase were found in phases, II, III, and IV. Similarly, some of the proteins involved in protein homeostasis including conjugation E2 protein, M18-class aspartyl aminopeptidase (DAP), and base subcomplex regulatory component RPT1 were expressed in phases II, III, and IV. Protein modification proteins were expressed in phase PP2A-class phosphatase was expressed in phase II and IV. The proteins related to cell wall organization and protein translocation showed diversity of expression in different phases. For instance, callose synthase protein was expressed in phases III and phase IV, while matrix polysaccharide-trafficking kinesin (FRA) was expressed in phase II and IV. The other cell wall organization related proteins expressed in all three phases were alpha-L-arabinofuranosidase activities, bifunctional alpha-L-arabinofuranosidase and beta-D-xylosidase, Class-A endo-1,4-beta-glucanase (KOR). AAA-ATPase motor complex regulatory component YCF2 and Cell-plate formation.phragmoplastin (DRP1), which participates in protein translocation and cell division, respectively, was expressed in all three phases (phase II, phase III, and phase IV). Export karyopherin (XPO1) was expressed in phases III and IV. Coenzyme metabolism related proteins that were expressed in three phases were transferred phase assembly factor (COG0354m/IBA57.1), and pyridoxal 5-phosphate synthase complex (synthase component).

With regard to proteins participating in cytoskeleton organization, profilin actin nucleation protein and WEB1-PMI2 actin filament reorganization complex component PMI2 were found in all germination phases and in phase III, respectively. Protein cell-plate formation phragmoplastin DRP1, which is involved in cell division, was found in all three phases (phase II, phase III, and phase IV). RNA processing and interference activity related proteins were found in phase III and phase IV including RNA-induced silencing complex (RISC) assembly and export karyopherin (XPO1). An enzyme responsible for initiation isoprenoid biosynthesis pathway, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase, was also found in phases III nd phase IV. During phase II, phase III and phase IV, vesicle trafficking activity seems to improve in the presence of coat protein complex scaffolding component Sec31, VPS4-VTA1 ESCRT-disassembly complex (ATPase component VPS4/SKD1), sRNA cargo-loading helicase (RH11/37), AP-1 trans-Golgi network cargo adaptor complex large subunit gamma (AP1G), Small gtpase (RAB5), RAB-GTPase GDP-dissociation inhibitor (RAB-GDI). Pertaining to redox homeostasis, 4-hydroxyphenylpyruvate dioxygenase (HPPD), and gamma-glutamyl: cysteine ligase was found in phase III, while methionine S-enantiomer sulfoxide reductase (msra) was found in phase IV. A carrier-mediated solute transport protein NCS-2 family solute transporter (NAT) was expressed in all three phases. A carrier-mediated anion transporter (NRT1/PTR) was expressed in phase II, and phase IV but was absent in phase III. JA (jasmonic acid) biosynthesis protein oxophytodienoate reductase (OPR3) was found in all three phases.

The unique proteins found in phase II were cytochrome b6/f complex apocytochrome b component PetB, ubiquitin-ligase U-Box E3 ligase, Chaperone (AKR2), and ubiquitin-fold protein (SUMO), which were found to be involved in photosynthesis, protein homeostasis, protein translocation, and protein homeostasis, respectively. The RAM signaling and glutathione S-conjugate transporter protein, ATRIP transducer kinase, HOPS/CORVET tethering component VPS41/VAM2, HOPS-specific vesicle tethering protein, lambda glutathione S-transferase, RAM signalling scaffold component (MO00), and ARF-GTPase activation class I ARF-GAP was found in phase III only. The highest number of proteins is uniquely found in phase IV. These proteins include glucose-6-phosphate dehydrogenase, pyrophosphate-dependent phosphofructokinase, kinase co-activator (ILITHYIA/GCN1), PP2A-class phosphatase complex regulatory component B2, ATP-dependent citrate lyase alpha chain, ribonucleotide (RN) purine synthetase (PUR5), S-adenosyl methionine synthetase (MAT), CAMK protein kinase (CDPK), methionine S-methyltransferase, OPC-8:coa

oxidase (ACX1), multifunctional enzyme (MFP), assembly phase scaffold protein (NBP35), cytochrome c reductase component MPP subunit alpha, UDP-L-arabinose kinase, fatty acid elongase (FAE) component enoyl-CoA reductase (ECR), monofunctionial hydroxyacyl-CoA dehydrogenase, isocitrate lyase, ribonucleotide (RN) synthetase (PUR5), iron-sulfur cluster ISC system transfer scaffold protein (NFU4/5), MPP mitochondrial signal peptidase subunit alpha, 1,4-beta-glucan synthase (CSLC), TIC translocation channel (Tic110), nuclear pore complex nucleoporin (NUP58), RopGEF guanine nucleotide exchange factor (PRONE).

With regard to assigned, but not annotated proteins, the numbers found in phase 0 (control), phase 2, phase 3, and phase 4 were 57, 62, 65, and 65, respectively (**Appendix H**). Unassigned as well as annotated proteins were present in all three phases and phase 0 (control seeds) were malate dehydrogenase, defensin-2, albumin-2, chaperonin (20 kDa), defensin-1, late embryogenesis abundant protein (11 kDa), and wound-induced basic protein. ABA-responsive protein, actin-interacting protein, RETICULATA-RELATED protein, developmentally-regulated G-protein, and PLAT domain-containing protein were all the other phases except control. The abundance of endochitinase, and vestitone reductase enzymes was increased during phase 4. During phase 3 and 4, chromoplast-specific carotenoid-associated protein and alpha-mannosidase proteins were found.

With regards to the carbohydrate degradation enzymes beta-amylase was found in all the phases studied while the starch-debranching isoamylase (ISA3) was found in all the phases except phase IV, which is an indication that starch degradation initiates right from the beginning and slowdown from phase IV. One cell wall degradation enzyme beta-galactosidase, was present in phase 0 and phase II, while was absent in phase III and phase IV. Malate synthase is an enzyme that participates in glyoxylate cycle was absent in phase 0, while was present in the subsequent phases. methylthioalkylmalate synthase in phase III, methylated. 2-isopropylmalate synthase that involved in leucine and chloroplastic malate biosynthesis process was found in all the phases except phase IV.

Table 5.2.1 List of protein identities that were uniquely found in control or phase 0 thorugh MapMan ontology. Proteins identities found in at least half of the four selected samples in each phase are represented in the table

Category	Sub-category	Protein	%*	Function**
Protein biosynthesis	Aminoacyl-tRNA synthetase activities	Arginine-tRNA ligase	75	Helps in protein translation and amino acid editing or proofreading
Carbohydrate metabolism	Sucrose biosynthesis	Cytosolic fructose-1,6-bisphosphatase	100	Releasing inorganic phosphate from fructose-1,6-bisphosphate
RNA processing	Spliceosome-associated non-snrnp MOS4-associated complex (MAC)	MAC associated component MOS2	75	Disease resistance
Carrier-mediated transport	VIT family	Metal cation transporter (MEB)	75	Cation transporter (in vacuoles)
Cell wall organisation	Pectin modification and degradation (polygalacturonase activities)	Regulatory protein of polygalacturonase activity	50	Depolymerization and solubilization of cell wall polyuronides during ripening
Protein homeostasis	Protein quality control (Hsp90 system)	Hsp90/Hsp70-co-chaperone (HOP)	75	Chaperone-mediated protein complex assembly
Vesicle trafficking	Multi-pathway trafficking regulation (vesicle tethering)	HOPS/CORVET tethering component VPS50	75	Regulating membrane fusion at the tonoplast and the prevacuolar compartment
Photosynthesis	calvin cycle	NADPH-dependent malate dehydrogenase (NADP-MDH)	100	NADP-dependent form is essential for the photosynthesis C4 cycle
Vesicle trafficking	multi-pathway trafficking regulation (CCV)	AP-4 vacuole cargo adaptor subunit mu (AP4M)	100	Trans-Golgi network
Photosynthesis	photophosphorylation (photosystem I)	PS-I component psaj	50	Chloroplast thylakoid membrane in photosystem I helps in photosynthesis
Protein homeostasis	ubiquitin-proteasome system (conjugation)	RUB/NEDD8 ubiquitin-fold protein (RUB)	50	Modification-dependent protein catabolic process, ubiquitin protein ligase binding

***** Percentage of presence among four selected samples from phase 0 or control. ** Functions were collected from https://www.uniprot.org/

Table 5.2.2 List of protein identities that were uniquely found in phase II thorugh MapMan ontology. Proteins identities found in at least half of the four selected samples in each phase are represented in the table

Category	Sub-category	Protein	%*	Function**
Protein biosynthesis	Organelle machinery-Translation termination	Ribosome recycling factor (RRF)	50	Termination of chloroplastic protein biosynthesis
Protein homeostasis	Ubiquitin-proteasome system (SUMO conjugation-sumoylation)	Ubiquitin-fold protein (SUMO)	75	Selective ubiquitin function (protein sumoylation)
Protein translocation	Chloroplast (outer envelope insertion system)	Chaperone (AKR2)	50	Chaperone activity in chloroplast
photosynthesis	photophosphorylation	cytochrome b6/f complex apocytochrome b component PetB	75	electron transfer between PSII and PSI

***** Percentage of presence among four selected samples from phase II. ** Functions were collected from https://www.uniprot.org/

Table 5.2.3 List of protein identities that were uniquely found in phase III thorugh MapMan ontology. Proteins identities found in at least half of the four selected samples in each phase are represented in the table

Category	Sub-category	Protein	%*	Function**
Multi-process regulation	RAM (Regulation of Ace2 and Morphogenesis) signalling	RAM signalling scaffold component (MO25)	50	Sense environmental conditions, phytohormones, and other external factors
External stimuli response	Vacuolar sequestration	Glutathione S-conjugate transporter	75	Arsenite transport and response to arsenic containing substance
	Plastid movement	WEB1-PMI2 actin filament reorganisation complex component PMI2	50	Response to blue light
Vesicle trafficking.	multi-pathway trafficking	HOPS/CORVET tethering component VPS41/VAM2	75	formation and sorting of endosomal cargo proteins into MVBs
Redox homeostasis	Glutathione-based redox regulation	lambda glutathione S-transferase	75	Glutathione dependent reduction of S-glutathionylquercetin

***** Percentage of presence among four selected samples from phase III. ** Functions were collected from https://www.uniprot.org/

Table 5.2.4 List of proteins identities that were uniquely found in phase IV thorugh MapMan ontology. Proteins identities found in at least half of the four selected samples in each phase are represented in the table

Category	**Sub-category**	**Protein**	**%***	**Function****
Carbohydrate metabolism	Oxidative pentose phosphate pathway	Glucose-6-phosphate dehydrogenase	75	Pentose phosphates for fatty acid and nucleic acid synthesis
Cellular respiration	Glycolysis (cytosolic)	Pyrophosphate-dependent phosphofructokinase	75	Carbohydrate degradation
Protein biosynthesis	Pre-Initiation Complex (PIC) module eif2	Kinase co-activator (ILITHYIA/GCN1)	75	Protein biosynthesis
Lipid metabolism	Fatty acid degradation (core beta-oxidation)	Multifunctional enzyme (MFP)	75	Fatty acid catabolism
Amino acid metabolism	Aspartate group amino acid biosynthesis (S-methylmethionine cycle)	Methionine S-methyltransferase	75	S-methylmethionine (SMM) biosynthesis, methylation
Coenzyme metabolism	Iron-sulfur cluster assembly CIA machinery	Assembly phase scaffold protein (NBP35)	50	Component of plastids, mitochondria, cytosol, and nucleus (biogenesis)
Redox homeostasis	Methionine sulfoxide reductase activities	Methionine S-enantiomer sulfoxide reductase (msra)	50	Redox of methionine sulfoxide and oxidative stress reduction
Phytohormone action	Jasmonic acid.biosynthesis	OPC-8:coa oxidase (ACX1)	50	Desaturation of long-chain mono-unsaturated substrate, JA biosynthesis in root tips
Protein modification	Phosphorylation (serine/threonine phosphatase	PP2A-class phosphatase complex regulatory component B2	50	Salt stress response
Carbohydrate metabolism	Nucleotide sugar biosynthesis	UDP-L-arabinose kinase	75	Converts and transports arabinose

Category	Sub-category	Protein	%*	Function**
Nucleotide metabolism	Ribonucleotide (RN) anabolism	Ribonucleotide (RN) purine synthetase (PUR5)	50	Ribonucleotide inosinemonophosphate biosynthesis
Lipid metabolism	Fatty acid metabolism	Fatty acid elongase (FAE) component enoyl-CoA reductase (ECR)	50	Long-chain fatty acids elongation
		Monofunctionial hydroxyacyl-CoA dehydrogenase	75	Peroxisomal fatty acid beta-oxidation (seed germination). Benzoic acid (BA) biosynthesis.
		Isocitrate lyase	50	Storage lipid mobilization in plant seedling.
	Fatty acid metabolism (citrate shuttle)	ATP-dependent citrate lyase alpha chain	100	Biosynthesis: cytosolic acetyl-CoA, isoprenoids, and flavonoids. Fatty acids elongation, and seed growth
Coenzyme metabolism	Iron-sulfur cluster assembly machinery	Iron-sulfur cluster ISC system.transfer scaffold protein (NFU4/5)	50	Iron-sulfur cluster assembly
	S-adenosyl methionine (SAM) cycle	S-adenosyl methionine synthetase (MAT)	75	Formation of S-adenosylmethionine, hydrolysis of the triphosphate, biosynthesis of lignin
Protein modification	Targeting peptide maturation	MPP mitochondrial signal peptidase subunit alpha	75	Transfer electrons for transmembrane transport and the ATP synthase
	Phosphorylation	CAMK protein kinase (CDPK)	100	Signal transduction
Cell wall organisation	Hemicellulose (xyloglucan) biosynthesis	1,4-beta-glucan synthase (CSLC)	50	Synthesis of the xyloglucan backbone

Category	Sub-category	Protein	%*	Function**
Protein translocation	Chloroplast inner envelope	TIC translocation channel (Tic110)	75	Importing protein precursor into chloroplasts
Protein translocation	Nuclear pore complex (NPC)	Nucleoporin (NUP58)	75	Nucleocytoplasmic trafficking. Regulatory roles in the gibberellin and auxin signaling pathway
Multi-process regulation	ROP-GTPase regulatory system	RopGEF guanine nucleotide exchange factor (PRONE)	75	Activates of Rop (Rho of plants) GTPases

***** Percentage of presence among four selected samples from phase IV. ** Functions were collected from https://www.uniprot.org/

Table 5.2.5 List of proteins identities that were differentially expressed in three distinct phases (phase II, phase III, and phase IV) thorugh MapMan ontology. Proteins identities found in at least half of the four selected samples in each phase are represented in the table

Category	Sub-category	Protein	P0*	PII*	PIII*	PIV*	Function**
Photosynthesis	Photophosphorylation	NdhF: NADH dehydrogenase-like (NDH) component	NA	NA	50	50	Transfer electron in light reaction
Carbohydrate metabolism	Sucrose degradation	Hexokinase	NA	100	100	100	Glucose export to cytosol
	Starch biosynthesis	Plastidial phosphoglucose isomerase	NA	50	75	75	Starch biosynthesis, gluconeogenesis, glycolysis pathway
	Oligosaccharide metabolism	Galactinol-sucrose galactosyltransferase	NA	50	00	100	Raffinose biosynthesis
Lipid metabolism (Fatty acid degradation)	Glyoxylate cycle	Malate synthase	NA	50	100	100	Glyoxylate cycle (S-malate from isocitrate)
	Citrate shuttle	ATP-dependent citrate lyase alpha chain	NA	50	50	100	Biosynthesis: Lipid and flavonoid, acetyl-CoA

Category	Sub-category	Protein	P0*	PII*	PIII*	PIV*	Function**
Protein biosynthesis	Ribosome biogenesis-large ribosomal subunit (LSU) processome	Pre-60S ribosomal subunit nuclear export factor (ARX1)	NA	75	75	50	regulate cell growth, division, especially the auxin activity
	mRNA loading	Eif-iso4f unwinding complex (eif-iso4g)	NA	50	50	50	Biotic defense
	EIF2 Met-tRNA binding factor activity	Eif2-alpha component	NA	100	100	50	Translation initiation
	Organelle machinery-Mitochondrial ribosome biogenesis	Mitochondrial tRNA import system.import factor (TRIC)	NA	100	100	00	Protein transporter. Mitochondria organization
	Aminoacyl-tRNA synthetase activities	Isoleucine-tRNA ligase	NA	100	100	100	Activate tRNA
Solute transport	APC superfamily	NCS-2 family solute transporter (NAT)	NA	75	100	100	Transmembrane transporter
Protein homeostasis	Proteolysis (metallopeptidase activities)	M18-class aspartyl aminopeptidase (DAP)	NA	75	50	50	Aminopeptidase, metallopeptidase activity
	Ubiquitin-proteasome system (26S/19S regulatory particle)	Regulatory component RPT1	NA	100	100	100	Protein degradation
	Membrane-anchored Ubiquitin-fold (MUB)	Conjugation E2 protein	NA	100	50	50	Selective protein polyubiquitination
Cell wall organisation	Cellulose (cellulose synthase complex CSC)	Matrix polysaccharide-trafficking (FRA)	NA	50	NA	50	Regulate cellulose microfibrils for growth
	Pectin modification and degradation (rhamnogalacturonan)	Bifunctional alpha-L-arabinofuranosidase and beta-D-xylosidase	NA	50	100	50	Releases xylose and arabinose by xylan catabolism
	Cellulose (cellulose synthase complex CSC)	Class-A endo-1,4-beta-glucanase (KOR)	NA	100	75	50	Cell elongation and cytokinesis
	Callose	Callose synthase	NA	NA	75	50	Callose synthesis during cytokinesis
Cell division	Cytokinesis (cell plate formation)	phragmoplastin (DRP1)	NA	100	100	100	Cytokinesis

Category	Sub-category	Protein	P0*	PII*	PIII*	PIV*	Function**
Coenzyme metabolism	Iron-sulfur cluster assembly (mitochondrial ISC system)	Transfer phase assembly factor COG0354m/IBA57.1	NA	75	100	50	Fe/S cluster biogenesis (Folate-dependent)
	Pyridoxalphosphate biosynthesis	Pyridoxal 5-phosphate synthase complex (synthase component)	NA	100	100	100	Response to stress (osmotic, oxidative, salinity)
Redox homeostasis	Tocopherol biosynthesis	4-hydroxyphenylpyruvate dioxygenase (HPPD)	NA	50	75	100	Tyrosine catabolism
	Glutathione biosynthesis	Gamma-glutamyl:cysteine ligase	NA	50	50	50	Glutathione biosynthesis
Vesicle trafficking	Anterograde trafficking (Coat protein II (coatomer machinery)	Coat protein complex scaffolding component Sec31	NA	100	100	100	Protein transport (ER to GA)
	Endocytic trafficking (ESCRT-mediated sorting)	VPS4-VTA1 ESCRT-disassembly complex (atpase component VPS4/SKD1)	NA	75	100	50	Membrane protein transport
	Exocytic trafficking (cross-kingdom extracellular vesicle trafficking)	sRNA cargo-loading helicase (RH11/37)	NA	75	100	100	Energy (ATP hydrolysis), regulation of gene expression
	Multi-pathway trafficking regulation (clathrin coated vesicle CCV machinery)	AP-1 trans-Golgi network cargo adaptor complex large subunit gamma (AP1G)	NA	50	75	100	Protein sorting at Golgi network
	Multi-pathway trafficking regulation (RAB5-RAB7-dependent pathway)	Small GTPase (RAB5)	NA	50	50	75	Endocytosis
	Multi-pathway trafficking regulation (RAB-GTPase membrane association)	RAB- GTPase GDP-dissociation inhibitor (RAB-GDI)	NA	100	100	100	Protein transport (vesicle-mediated)

Category	Sub-category	Protein	P0*	PII*	PIII*	PIV*	Function**
Cytoskeleton organisation	Microfilament network (actin polymerisation)	Profilin actin nucleation protein	NA	100	100	100	Building cytoskeleton structure
Protein translocation	Chloroplast (inner envelope TIC translocation system)	AAA-ATPase motor complex.regulatory component YCF2	NA	75	100	100	ATP binding and hydrolysis in stroma
Photosynthesis	Photophosphorylation (photosystem II)	reaction center component D2/PsbD	NA	50	50	NA	
Lipid biosynthesis	Phytosterol metabolism (plant sterol pathway)	Sterol delta24 reductase	NA	75	NA	75	Cell elongation
Protein modification	Phosphorylation (serine/threonine phosphatase)	PP2A-class phosphatase complex component A	NA	50	NA	100	Resistance to phosphatase inhibitors
RNA processing	mRNA silencing pathway	RNA-induced silencing complex (RISC) assembly	NA	NA	50	75	Oxidative stress reduction
RNA processing	Nucleus (nucleocytoplasmic transport)	Export karyopherin (XPO1)	NA	NA	50	50	Oxidative stress reduction
Protein homeostasis	Proteolysis (cysteine-type peptidase activities)	C1-class protease (Papain)	NA	NA	50	50	Proteolysis
Secondary metabolism	Methylerythritol phosphate (MEP) pathway (terpenoids)	4-hydroxy-3-methylbut-2-enyl diphosphate synthase	NA	NA	75	100	Salicylic acid (SA), non-mevalonate pathway, isoprenoid biosynthesis
	Retrograde trafficking (Coat protein I coatomer machinery)	ARF-gtpase activation class I ARF-GAP protein	NA	NA	50	100	Growth of root hairs

***** Percentage of presence among four samples from phase0, phase II, phase III, phase IV. phase0, phase II, phase III, phase IV is represented as P0, PII, PIII, PIV ** Functions were collected from https://www.uniprot.org

Protein peak intensities within samples were log transformed and no other normalization was performed (**Supplementary Table 1**). One-way ANOVA with the obtained intensities of proteins revealed 14 proteins that were differentially expressed among the different stages of seed germination (**Figure 5.6; Table 5.3**). These proteins included thioredoxin-like protein Clot, casparian strip membrane protein, mitochondrial-processing peptidase subunit alpha, disease resistance response protein Pi49, disease resistance response protein DRRG49-C, 40S ribosomal protein S7-1, malate synthase, glyoxysomal, bifunctional L-3-cyanoalanine synthase/cysteine synthase 2, seed biotin-containing protein SBP65, vestitone reductase, protein SLE1, protein SLE2, carrot ABA-induced in somatic embryos 3, and ABA-responsive protein ABR18 (**Table 5.3**). Partial Least-Squares Discriminant Analysis (PLS-DA) maximizes the covariance principal components one and two, and therefore represents the correlation among all the samples in distinct phases with their replicates (**Figure 5.7**). **Figure 5.7** shows that variations exists among the replicates of phase 0 or control and phase II, despite minimizing covariance, while all the samples from phase IV highly correlated with each other.

The correlation heatmaps show the correlation among expressed protein features in terms of their quantities (**Figure 5.8**). Proteins that were highly expressed in phase 0 or control, were downregulated in subsequent phases, and had the lowest quantity in phase IV seeds were: late embryogenesis abundant protein 47, Protein SLE2, Carrot ABA-induced in somatic embryos 3, Protein SLE1, and Phytochrome 1 (**Figure 5.9**). Some proteins had their lowest expression in phase 0 or control and were upregulated in subsequent phases, with the greatest quantity appearing in phase IV. This group included disease resistance response protein DRRG49-C, cysteine proteinase 15A, ABA-responsive protein ABR17, probable mediator of RNA polymerase II transcription subunit 36b, glyoxysomal malate synthase, and NADH dehydrogenase [ubiquinone] iron-sulfur protein 2 (**Figure 5.9**). However, flagellar radial spoke protein was the lowest in control seeds and was the highest in phase III followed by a decline in phase IV.

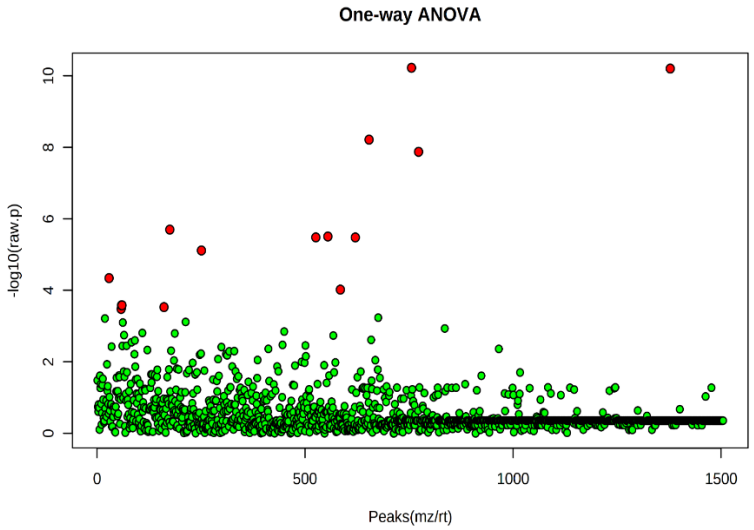


Figure 5.6 One-way ANOVA showing variation in expression among all four different phases of seed germination. The red circles indicate the proteins that were variably expressed.

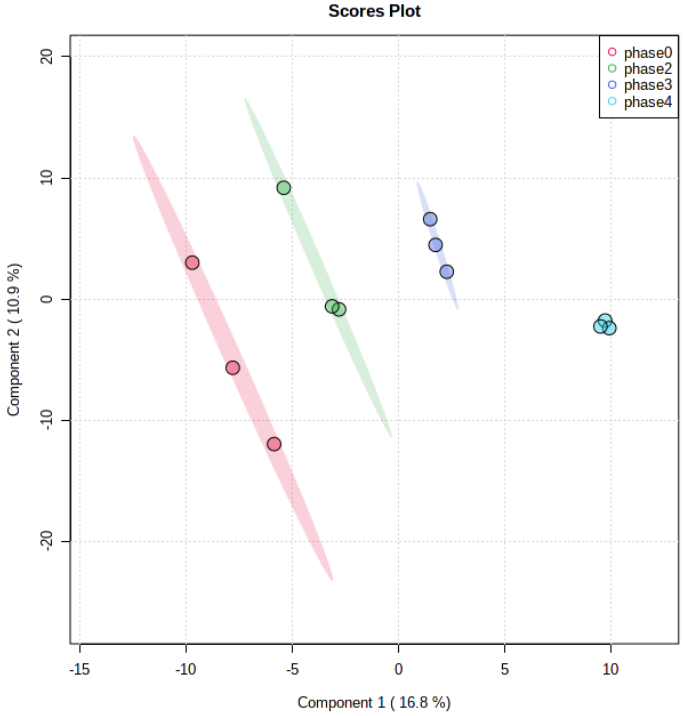


Figure 5.7 PLS-DA (Partial Least-Squares Discriminant Analysis) for four phases (control, phase II, phase III, and phase IV) for three replicates from each phase. The variance displayed is the explained variance for X for each sample containing quantitative value of proteins

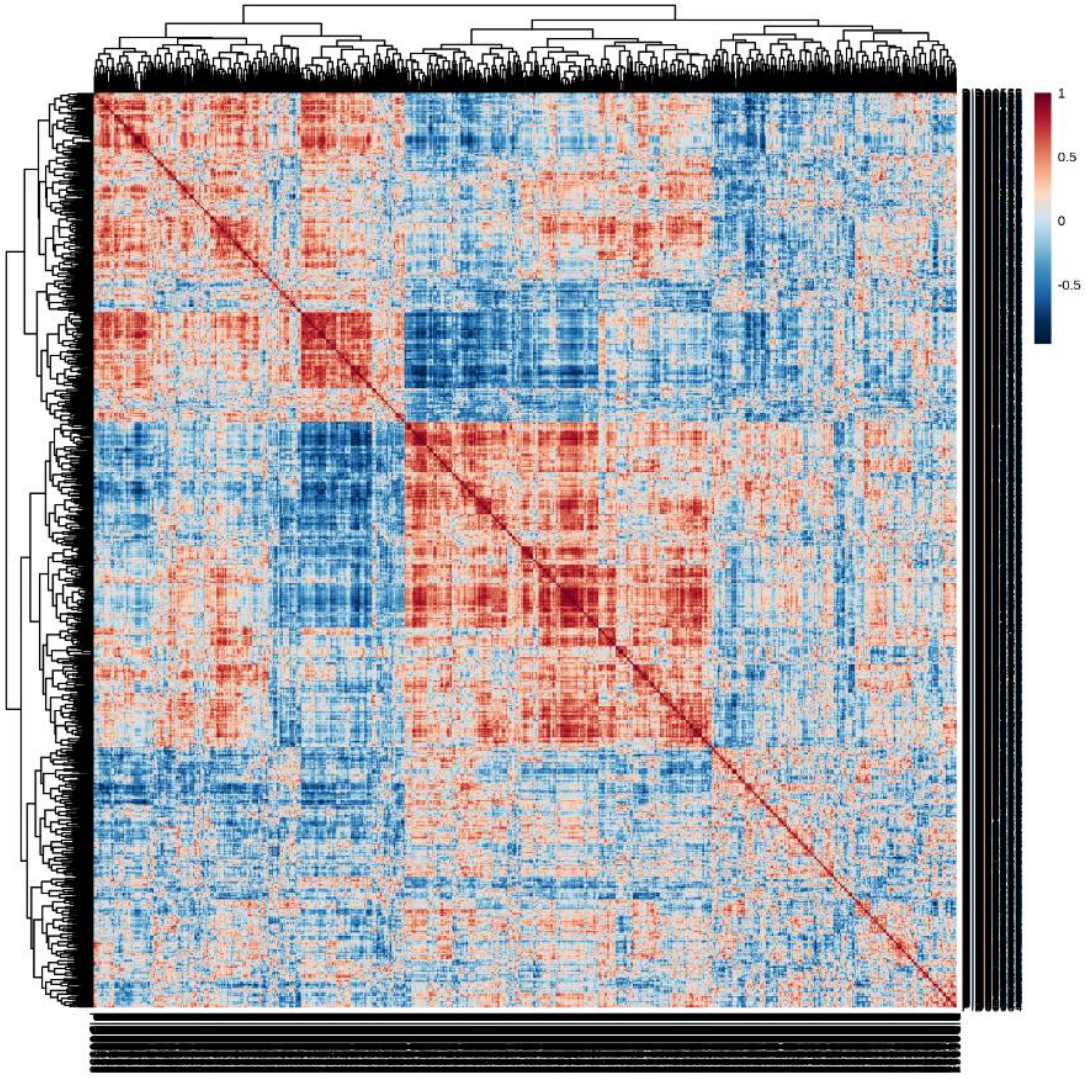


Figure 5.8 Correlation heatmaps of proteins expressed during the experiment. Red zones indicate higher levels of correlation - blue zones indicate lower levels of correlation among protein features.

Table 5.3 List of 14 proteins that were expressed in different quantities during four phases of seed germination as obtained through one-way ANOVA

Protein name	**Species**	**F-value**	**P-value**	**FDR**
Thioredoxin-like protein	*Oryza sativa* subsp. *japonica*	1199.7	5.95E-11	4.71E-08
Casparian strip membrane protein	*Solanum tuberosum*	1184.3	6.26E-11	4.71E-08
Mitochondrial-processing peptidase subunit alpha	*Solanum tuberosum*	375.13	6.09E-09	3.06E-06
Disease resistance response protein Pi49	*Pisum sativum*	307.93	1.33E-08	5.01E-06
Disease resistance response protein DRRG49-C	*Pisum sativum*	85.852	2.00E-06	0.000603
40S ribosomal protein S7-1	*Arabidopsis thaliana*	76.552	3.12E-06	0.000623
Malate synthase, glyoxysomal	*Glycine max*	75.41	3.30E-06	0.000623
Bifunctional L-3-cyanoalanine synthase/cysteine synthase 2	*Solanum tuberosum*	75.36	3.31E-06	0.000623
Seed biotin-containing protein SBP65	*Pisum sativum*	37.735	4.55E-05	0.006841
Vestitone reductase	*Medicago sativa*	30.843	9.55E-05	0.013067
Protein SLE1	*Glycine max*	23.306	0.000262	0.030341
Carrot ABA-induced in somatic embryos 3	*Daucus carota*	23.304	0.000262	0.030341
ABA-responsive protein ABR18	*Pisum sativum*	22.535	0.000295	0.031728
Protein SLE2	*Glycine max*	21.804	0.000332	0.033264

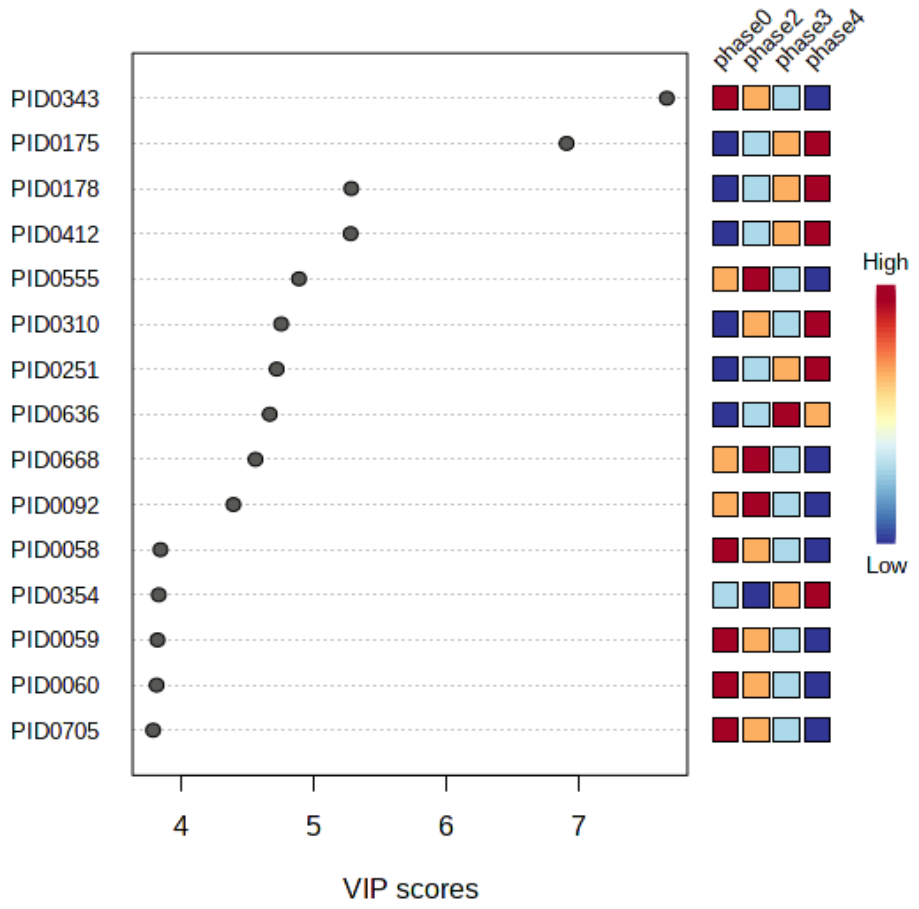


Figure 5.9 The variable importance in projection (VIP) of PLS-DA. The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group under study. The The x-axis lists PID0343, PID0175, PID0178, PID0412, PID0555, PID0251, PID0310, PID0636, PID0668, PID0092, PID0058, PID0354, PID0059, PID0060, and PID0705 which represent the following proteins: late embryogenesis abundant protein 47, disease resistance response protein DRRG49-C, cysteine proteinase 15A, ABA-responsive protein ABR17, 40S ribosomal protein S7-1, glyoxysomal malate synthase, probable mediator of RNA polymerase II transcription subunit 36b, flagellar radial spoke protein 1, chloroplastic fructokinase-like 2, chloroplastic glyceraldehyde-3-phosphate dehydrogenase B, protein SLE2, NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, carrot ABA-induced in somatic embryos 3, protein SLE1, and Phytochrome 1, respectively.

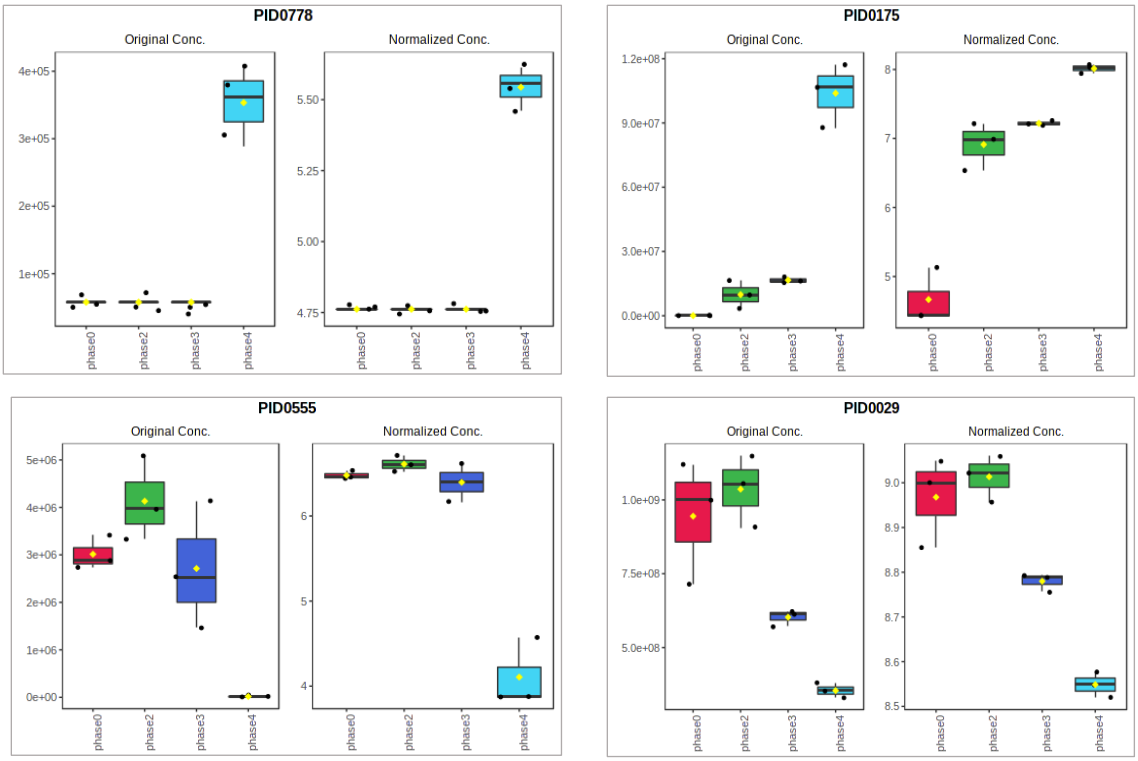


Figure 5.10 Differentially expressed proteins among different phases of seed germination in lentil. PID0778, PID0175, PID0555, and PID0029 represent disease resistance response protein Pi49, disease resistance response protein DRRG49-C, 40S ribosomal protein S7-1, and seed biotin-containing protein SBP65, respectively.

The thioredoxin-like protein is a thiol-disulfide oxidoreductase that helps in chloroplast development by regulating plastid-encoded polymerase (PEP) dependent chloroplast transcription, was reduced 2.42 fold after reaching phase IV from phase 0. Mitochondrial-processing peptidase subunit alpha proteolysis of mitochondrial sequence from imported precursor proteins, and the quantity of this protein was reduced 2.59-fold in phase IV compared to phase 0. According to the seed germination studies, the mitochondria becomes fully functional during phase II (Ali and Elozeiri, 2017) and presumably in subsequent stages a major portion of mitochondrial-processing peptidase may get degraded. Among the defense related proteins, the amount of resistance response protein Pi49 and disease resistance response protein DRRG49-C reduced 2.6135 and 10.681 times, respectively in phase IV compared to control. Additionally, ABA-responsive protein ABR18 that transduces defense response to biotic stimulus was decreased 2.38 times. Conversely, vestitone reductase that induces defense reactions in plants was increased 1.22 times in phase IV compared to control seeds. A study of the proteome of *Brassica napus* has revealed that biotic and abiotic stress-related proteins can be activated or inhibited during seed germination (Gu et al.,

2019). The expression of some germination-related proteins were found to change significantly among the phases of seed germination studied in this experiment. Seed biotin-containing protein SBP65, which is a growth-limiting enzyme, as well as a source of biotin, was decreased 4.57 times in phase IV. Glyoxysomal enzyme malate synthase that participates in the TCA cycle pathway was increased by 2.4 fold in phase IV compared to phase 0. Germination was also induced by carrot ABA-induced in somatic embryos 3 that is upregulated in response to abscisic acid. This protein is expressed in embryogenic cells to induce embryogenic cell production, even without the requirement of auxin (Ogata et al., 2005), which was found to increase by 5.38 times in phase IV. Proteins SLE1 and SLE2, which are found in the cotyledons, were increased 5.37 and 5.31 times, respectively. However, the role of these proteins in seed germination is still unexplored. **Figure 5.10** displays some of the proteins that were expressed variably among phase 0, phase II, phase III, and phase IV, with and without normalization of protein quantities. Overall, both the original and normalized peak intensities explained the variation in expression among the phases. However, in terms of disease resistance response protein DRRG49-C expression was significantly higher in phase II and phase III post normalization.

The **Figure 5.11** shows different banding pattern in SDS-PAGE of phase 0 or control, phase III, and in phase IV in IG 72541 (*L. nigricans*). Difference in banding pattern was clearly visible from between 25 kDa to 10 kDa (**Figure 5.11**). Three gel bands of higher intensities visible (S1, S2, and S3 in **Figure 5.11**) were selected for in-gel digestion for peptide and protein identification. The list of identified peptide sequences and number of unique peptides can be found in **Appendix J**.Vicilin, convicilin, seed linoleate 9S-lipoxygenase, legumin type B were identified. Vicilin and convicilin was identified from S1, S2, and S3; whereas linoleate 9S-lipoxygenase and legumin type B was identified from S2, and S3, respectively (**Appendix J**).

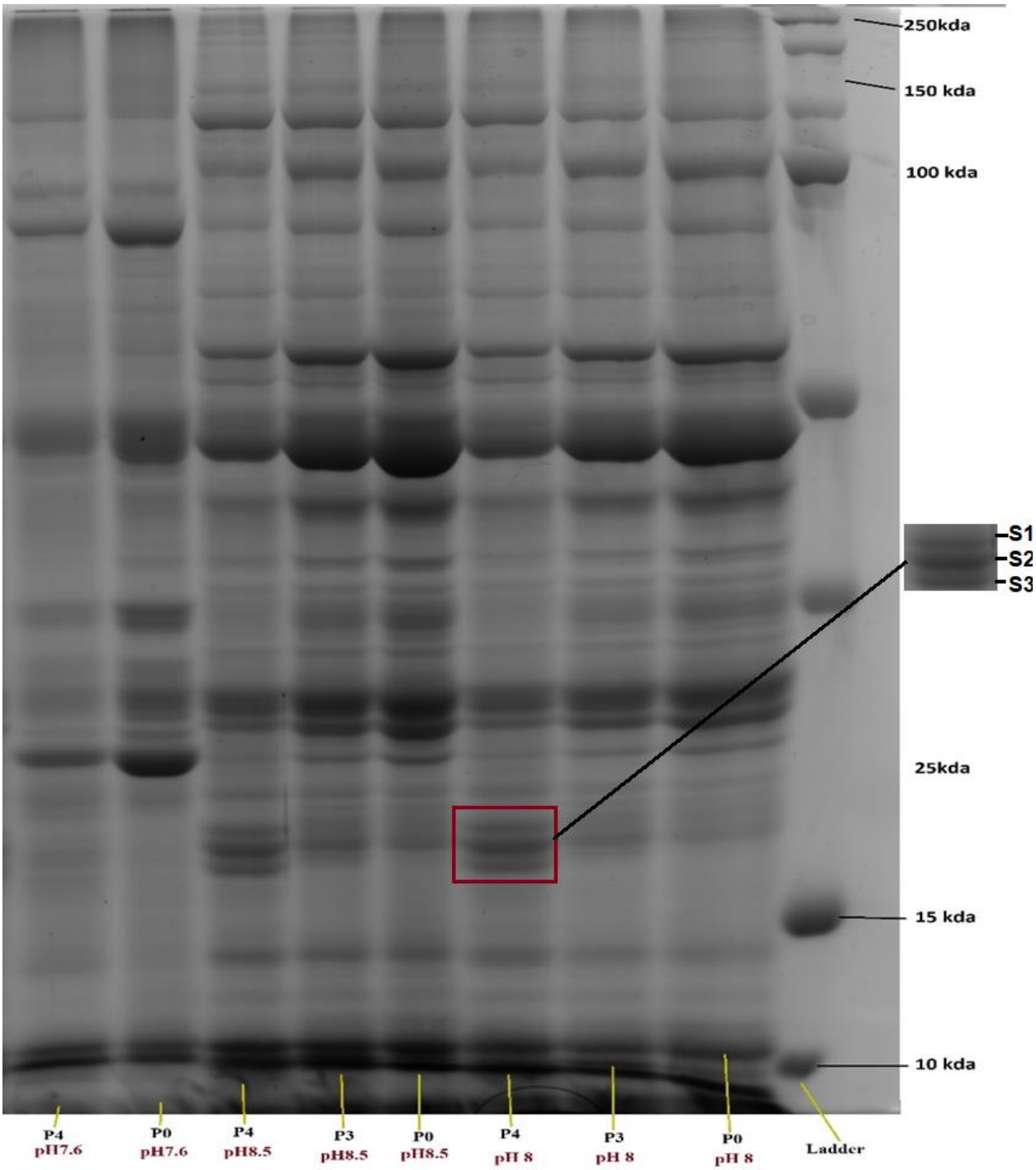


Figure 5.11 Photograph of seed protein banding pattern IG 72541 using SDS-PAGE at 10% gel concentration at pH 7.6, pH 8, and pH 8.5. P0, P3, and P4 represents phase 0, phase III, and phase IV, respectively. The marker ladder represent the weight of protein range from 2 kDa to 250 kDa in the right lane.

5.4 Discussion

The total protein concentration from BCA assay results revealed that the concentration of protein reduced as the germination progressed. A study of seed germination of *Pongamia pinnata* also revealed that during seed germination the protein concentration decreases and during seedling establishment the reduction of protein concentration is the highest (Kesari and Rangan, 2011).

The total number of proteins was counted using protein identities using Venn diagrams and also by counting manually. Through the Venn diagram, the screened total protein identities were 700, 744, 721, and 751 in phase 0 (control seeds), phase II, phase III, and phase IV, respectively, Through manual screening, the number of total protein identities found was 593, 527, 487, and 534 in phase 0 (control seeds), phase II, phase III, and phase IV, accordingly. The reason for the difference is that there was not enough proteome information regarding lentil available at the annotation software. Therefore, the proteome of lentil was searched against different legume, cereals, and oil crops for functional annotation that has generated the number of copies for proteins found in different species. With regard to the unique proteins found, through Venn diagrams, proteins present in all the phases from the same species were detected as similar, and others were then detected as unique proteins.

A few proteins involved in photosynthesis, protein biosynthesis, carbohydrate metabolism were present in phase 0 (control seeds), but not found in the phases of seed germination (**Table 5.2.1**). For instance, sucrose metabolism cytosolic fructose-1,6-bisphosphatase which catalyzes the irreversible reaction of releasing inorganic phosphorus from fructose-1,6-bisphosphate, was absent in the other phases. This enzyme regulates sucrose metabolism and the absence of this enzyme aids in glycolysis as fructose-1,6-bisphosphate, one of the key intermediates of glycolysis (Ogawa et al., 2015). In the photosynthesis process, NADPH-dependent malate dehydrogenase (NADP-MDH) and PS-I component PsaJ were found only in control seeds. With further query through the Kyoto Encyclopedia of Genes and Genomes, it was found that all the components and enzymes were not detected through this method, either in control or in the germination phases (**Figure 5.12**). The reason could the overabundance of seed storage proteins - some studies have found that due to the abundance of seed storage proteins, several mitochondrial proteins were not detected through mass spectrometry (Miernyk and Hajduch, 2011; Czarna et al., 2016). This could also be

a similar phenomenon for plastids. However, ferredoxin--NADP+ reductase [EC:1.18.1.2] became undetectable or possibly degraded after imbibition of lentil seeds in this study. This may indicate that all the photosynthesis components are not prepared for photosynthesis, even in phase IV, while the source of energy supply is through the glycolysis, pyruvate metabolism, TCA cycle, and starch and sucrose metabolism. Upon delving through glycolysis pathway in KEGG pathway all the phases studied, it was found that hexokinase [EC:2.7.1.1] that catalyzes the reaction to produce α-D glucose was found in phase 0 and phase 4 (**Figure 5.13**). However, through MapMan server identification, hexokinase was not found in phase 0 but all the subsequent phases. Studies have suggested that in leaf embryogenesis abundant protein has a protective effect on enzyme dehydration, especially for mitochondrial enzymes, while many enzymes are degraded due to water loss and shrinkage of seed size (Grelet et al., 2005). Another enzyme, 6-phosphofructokinase [EC:2.7.1.11], helps in the biosynthesis of β-D-glucose-6P, absent in phase 0. Fructose-1,6-bisphosphatase I [EC:3.1.3.11] and 6-phosphofructokinase [EC:2.7.1.11] are the enzymes that produce β-D-fructose-1,6 bisphosphate, which was absent in phases III and 0, respectively. Through this study, it was found that fructose-1,6-bisphosphatase I is preserved and expressed in the seedlings, and the supply β-D-fructose-1,6 bisphosphate that is indispensable for ATP production through glycolysis can be synthesized during all the phases of seed germination in presence of either of the enzymes (**Figure 5.13**). With regard to the pyruvate metabolism pathway, pyruvate decarboxylase [EC:4.1.1.1] and pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) [EC:2.3.1.12] that catalyzes the reaction to synthesize 2-(alpha-Hydroxyethyl)thiamine diphosphate or acetaldehyde and acetyl CoA, respectively, were found in both phases III and IV (**Figure 5.13; Figure 5.14; Appendix I)**. Hence, until phase II, there was no production of acetyl CoA which is the starting biomolecule for TCA cycle. In phase II the presence of acetyl-CoA synthetase [EC:6.2.1.1] that catalyze acetyl-CoA production was found, but no pyruvate decarboxylase [EC:4.1.1.1] or pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) [EC:2.3.1.12] or 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit alpha [EC:1.2.7.3, 1.2.7.11] or pyruvate ferredoxin oxidoreductase alpha subunit [EC:1.2.7.1] was found – these that are also required in the pathway of acetyl CoA production (**Figure 5.14**). In addition, succinate dehydrogenase (ubiquinone) flavoprotein subunit [EC:1.3.5.1] that catalyzes the conversion between fumarate and succinate was not found in phase II. This study shows that TCA cycle is less likely to initiate until mid-phase II (**Figure 5.14**). In

the starch and sucrose metabolism pathway of the control seeds, beta-glucosidase [EC:3.2.1.21] for synthesizing D-glucose was absent, while endoglucanase [EC:3.2.1.4] that helps in cellobiose formation was present (**Appendix I**). This result indicates that during the seed developmental stage, starch formation, and cell wall formation occurred due to the absence of these enzymes and after imbibition, the formation of beta-glucosidase and degradation of endoglucanase helps to produce more monomer of carbohydrates to be utilized directly by energy and other metabolic processes. However, an enzyme for starch glycogen biosynthesis, 1,4-alpha-glucan branching enzyme [EC:2.4.1.18] was found in phase 0 and phase II.

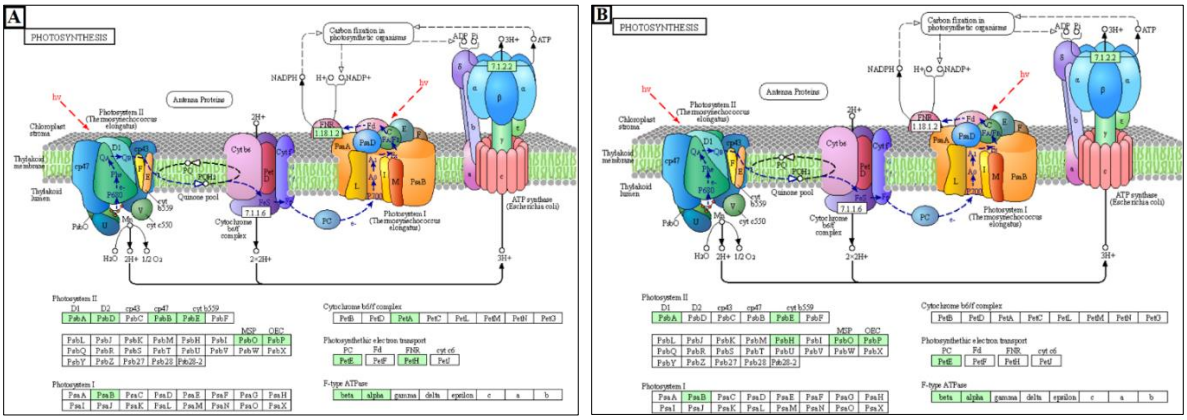


Figure 5.12 Representation of protein sequences involved in photosynthesis process; (A) phase 0 or control seeds, (B) phase IV. The protein similarity was analyzed thorugh KEGG annotation.

Proteins that are produced in response to stress showed variability. While many proteins were found to be present in all the phases of seed germination, including control, some proteins uniquely appeared in the control or in different phases of seed germination. For instance, spliceosome-associated non-snRNP MOS4-associated complex (MAC) was found in phase 0 of control seeds, while Eif-iso4f unwinding complex component eif-iso4g was found in the other phases of seed germination studied (**Table 5.2.5**). In addition, among the non-assigned proteins, ABA-responsive protein ABR18 and disease resistance response protein DRRG49-C were identified only among the seed germination phases (phase II, phase III, and phase IV) (**Appendix H**). However, among the unassigned proteins, the stress related proteins found among all the phases studied were Probable CCR4-associated factor 1 homolog 5, Tropinone reductase-like 2, Cyclase-like protein 1, Plastoglobulin-1, dehydrin DHN1, and plastid-lipid-associated protein, heat shock 70, universal stress protein PHOS32, NAD(P)H dehydrogenase (quinone) FQR1, and COP1-interactive protein

1 (**Appendix H**). There is also evidence that during the process of seed germination redox reaction is prompted that generates free radicle oxygen and for redox homeostasis several enzymes are expressed (El-Maarouf-Bouteau and Bailly, 2008; Chakrabarty et al., 2019; Adhikari et al., 2020). In this study 4-hydroxyphenylpyruvate dioxygenase (HPPD) and gamma-glutamyl: cysteine ligase was found in phase II, phase III, and phase IV **(Table 5.2.5)**. In addition, lambda glutathione S-transferase and methionine S-enantiomer sulfoxide reductase (msra) were found in phase III and phase IV, respectively which corroborate the antioxidant activity in germinating seeds to balance oxidative stress (**Table 5.2.3; Table 5.2.4**).

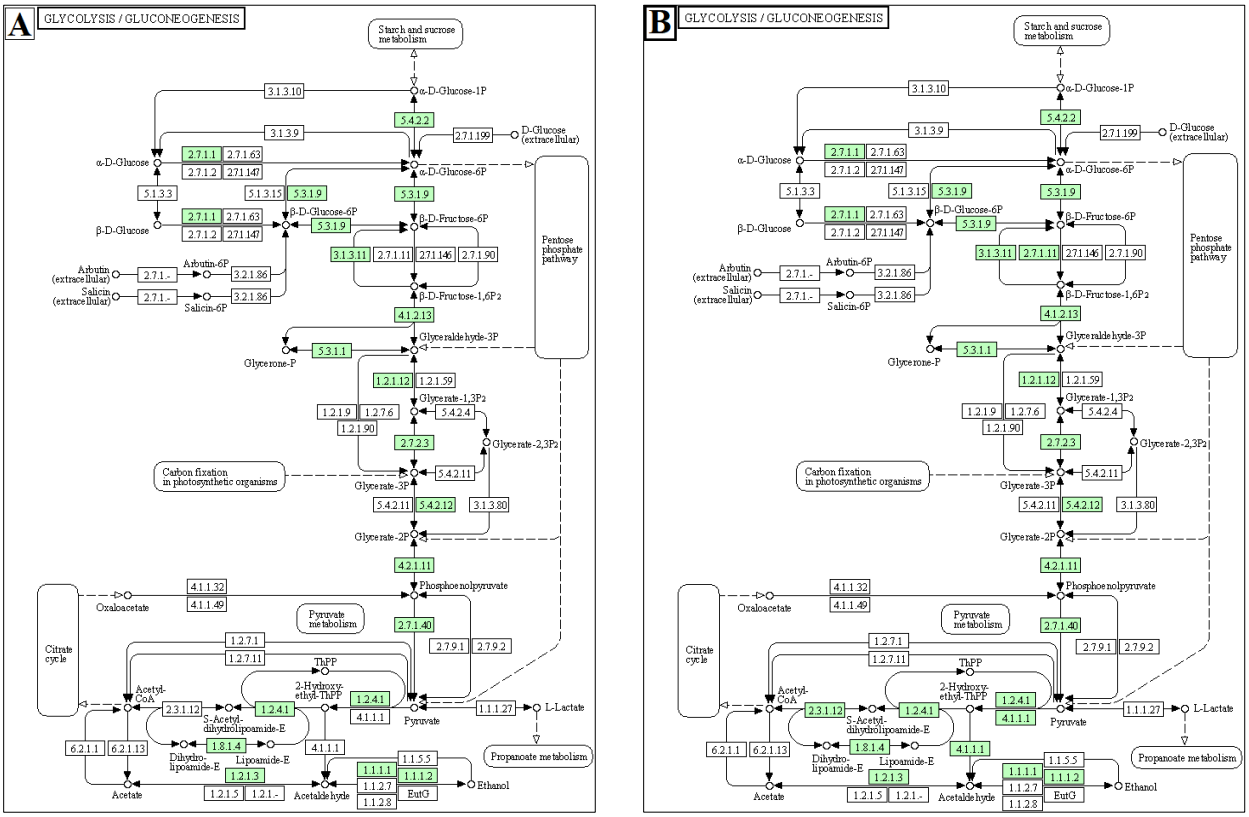


Figure 5.13 Representation of protein sequences involved in glycolysis pathway; (A) phase 0 or control seeds, (B) phase IV. The protein similarity was analyzed thorugh KEGG annotation server.

There are a few proteins found in control seeds that are ascribed from the seed developmental stage (**Table 5.2.1)**. One cell wall organization related protein that regulates the activity of polygalacturonase that is involved in depolymerization and solubilization of cell wall was present only in control seeds. Two multi-pathway trafficking protein homologs involved in regulating membrane fusion at the tonoplast and trans-Golgi network namely HOPS/CORVET tethering

component VPS50 and AP-4 vacuole cargo adaptor subunit mu (AP4M) were also found only in phase 0 or control seeds. Another protein homolog involved in protein homeostasis Hsp90/Hsp70-co-chaperone was also uniquely found in control seeds that are involved in chaperone-mediated protein complex assembly (**Table 5.2.1)**.

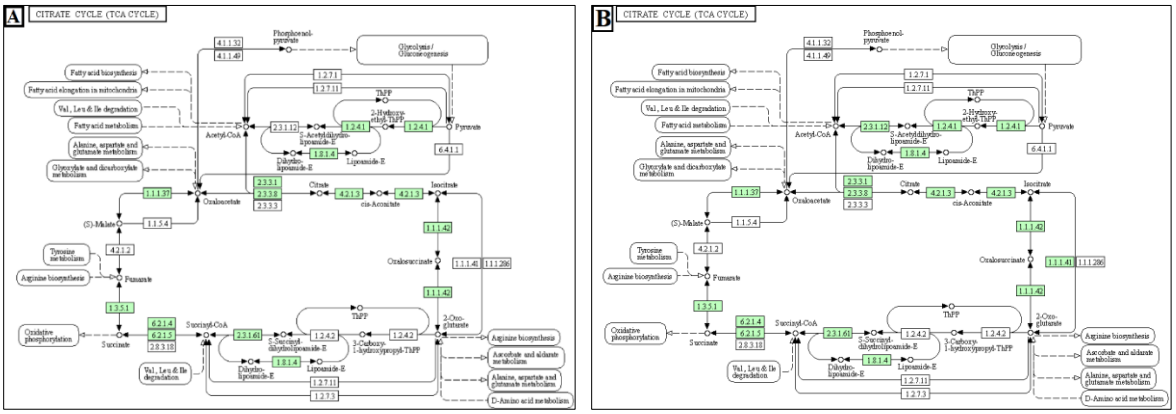


Figure 5.14 Representation of protein sequences involved in TCA (citrate) cycle; (A) phase 0 or control seeds, (B) phase IV. The protein similarity was analyzed thorugh KEGG annotation.

Some protein homologs found only in phase III may help in the germination process (**Table 5.2.3**). WEB1-PMI2 actin filament reorganization complex component PMI2 is indispensable for proper movement of plastid in response to blue light. This movement ensures absorbance of light energy through photosynthesis. HOPS/CORVET tethering component VPS41/VAM2 is a protein that was identified in phase III – it helps in transportation of biosynthesized proteins by forming and sorting of endosomal cargo proteins into MVBs (multivesicular bodies). RAM signalling scaffold component (MO25) homolog found that sense various environmental conditions, phytohormones, and external factors. Protein homologs that were uniqely identified during phase IV have tremendous influence on initiating metabolic processes during seed germination and seedling establishment process (**Table 5.2.4)**. According to the annotation through MapMan, multifunctional enzyme (MFP) as well as monofunctionial hydroxyacyl-CoA dehydrogenase homolog was identified in phase IV that participate in fatty acid catabolism processes. Isocitrate lyase and ATP-dependent citrate lyase alpha chain are the two enzymes found in phase IV that aids in storage lipid mobilization in plant seedling and biosynthesis of cytosolic acetyl-CoA, isoprenoids, and flavonoids, respectively. One methionine biosynthesis enzyme homolog namely methionine S-methyltransferase that produces S-methylmethionine (SMM) were also found. S-adenosyl methionine synthetase (MAT) is another enzyme that also help in formation of S-

adenosylmethionine as well as in the hydrolysis of the triphosphate for initiating lignin biosynthesis. A nucleotide sugar biosynthesis enzyme, UDP-L-arabinose kinase, which converts and transports arabinose was also found in phase IV. A previous study reported that arabinans arguably represent an abundance of non-cellulosic polysaccharide in embryos of seed in various plant species, and accounts for around 40% in Arabidopsis seeds, which declines roughly 15% during seedling establishment, and that the mobilization of arabinans plays an important role in seedling establishment (Gomez et al., 2009). Assembly phase scaffold protein (NBP35) and iron-sulfur cluster ISC transfer scaffold protein (NFU4/5) are two proteins that were present in phase IV – these are components of plastids, mitochondria, cytosol, and nucleus. Studies on Iron–sulphur (FeS) has revealed that FeS clusters are synthesized by the assembly of proteins that take part in biogenesis of plastids, mitochondria and cytosol for electron transfer (Bastow et al., 2017; Lu, 2018). Genesis of ribonucleotide from basic molecules is also indispensable for DNA replication and protein synthesis. Ribonucleotide (RN) purine synthetase (PUR5) is the protein that helps in formation of inosine 5'-monophosphate biosynthesis was found in phase IV. Two protein translocation related proteins namely TIC translocation channel (Tic110) and nucleoporin (NUP58) that helps in transportation of within plastid and nucleus were also found in phase IV only. A cell wall organization related protein homolog 1,4-beta-glucan synthase (CSLC) was found in phase IV is also an indicative of synthesis of the xyloglucan backbone of new cell.

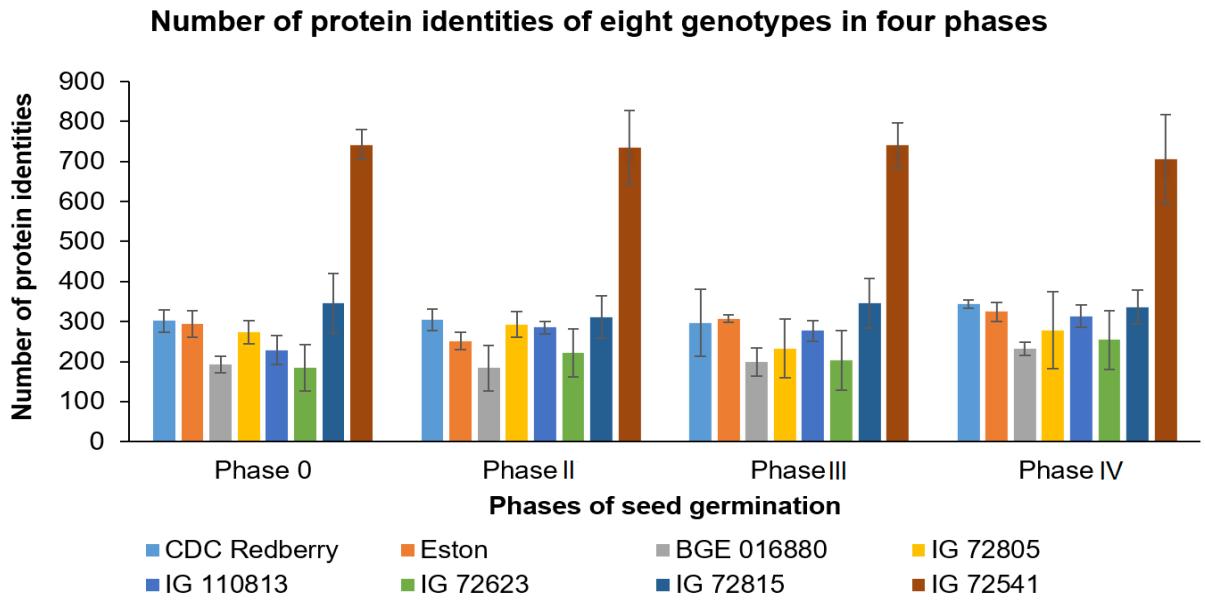


Figure 5.15 The number of identified proteins in different phases of seed germination in eight genotypes from seven lentil species.

OPC-8:CoA oxidase (ACX1) that is involved in JA biosynthesis was only found in phase IV. In the experiment, there were 6 replicates in each phase and the average number of identified protein sequences were around 760. However, in phase IV in the first three batches the number of identified protein sequences were 803, 790, and 807, while in the last three batches the number of identified protein sequences were 652, 537, and 645. Therefore, there is some possibility of missing proteins in the last batch of phase IV. In the phytohormone biosynthesis gibberellin perception and signal transduction protein DELLA, gibberellin-abscisic acid signal pathways crosstalk, cytokinin dehydrogenase, cytokinin perception and signal transduction. However, auxin transporter was present in all the phases during seed germination including control seeds. Alpha-amylase protein homolog comes after GA signal, which was present in all the phases studied (Damaris, et al., 2019). Through this study, β-Amylase was found in all the phases studied, while through KEGG pathway study beta-glucosidase was found in phase II, phase III, and phase IV. β-amylase degrade larger carbohydrates like starch while beta-glucosidase produce D-glucose from smaller carbohydrates (Yamasaki, 2003). Which reinstates the fact that carbohydrate degradation starts after the uptake of water by seeds, and glucose that is the key molecule for performing cellular activities being available during these phases. Starch-debranching isoamylase (ISA3) that reduces side branches of starch was also evident until phase III.

The quantitative analysis showed there was significant variation in 14 proteins among the studied phases (**Figure 5.6, Table 5.3, Supplementary Table 2**). All of them have direct or indirect roles in the cellular metabolism lentil seed germination. Among them, the thioredoxin-like protein (a thiol-disulfide oxidoreductase) which helps in chloroplast development by regulating plastid-encoded polymerase (PEP) dependent chloroplast transcription, was reduced 2.42 fold after reaching phase IV from phase 0. Mitochondrial-processing peptidase subunit alpha proteolysis of mitochondrial sequence from imported precursor's proteins was detected, and the quantity of this protein was reduced about 2.59 fold in phase IV compared to phase 0. According to a previous seed germination study, the mitochondria becomes fully functional during phase II (Ali and Elozeiri, 2017). Presumably in subsequent stages a major portion of mitochondrial-processing peptidase may become degraded. Among the defense related proteins, the amounts of resistance response protein Pi49 and disease resistance response protein DRRG49-C were reduced 2.613 and 10.681 times, respectively, in phase IV compared to control. Additionally, ABA-responsive protein ABR18 that transduces defense response to biotic stimulus, was decreased 2.38 times.

Conversely, vestitone reductase that induces defense reactions in plant was increased 1.22 times in phase IV compared to control seeds. A study on proteome of *Brassica napus* has revealed that biotic and abiotic stress-related proteins can be activated or inhibited during seed germination (Gu et al., 2019). The expression of some germination-related proteins were found to change significantly among the phases of seed germination studied in this experiment. Seed biotin-containing protein SBP65, which is a growth-limiting enzyme as well as a source of biotin was decreased 4.57 times in phase IV. Glyoxysomal enzyme malate synthase that participates in TCA cycle pathway was increased by 2.4 times.

This study initially was conducted with eight genotypes from seven lentil species (**Figure 5.15**). But the obtained number of protein homologs were limited to approximately between 250 to 350 proteins. As the number of protein homologs obtained was below average for plant seed protein counts, the experiment was repeated for verifying sample preparation or chemical preparation. Upon verification and multiple injections, the protein number was not improved. It is presumed that due to non-ideal conditions of where the instruments are located, factors such as the air flow and dust accumulation may retard proper spraying and charging of ions through ESI by formation of bubbles. Another possibility is that the load concentration was excessively higher (1ug/ul) than the column used, resulting in contamination, leading to poor recovery by the solvent and washing method. However, the samples from *L. nigricans* showed higher number of protein identities compared to the others. Therefore, only IG 72541 was accounted for further analysis.

The in-gel digestion study was also affected due to the instrumental limitations. There are some gel bands between 75 kDa and 10 kDa (at pH 8.5, **Figrue 5.11**) was taken for identification where not peptides could be identified from plant origin. This could possibly happened either due to instrumental limitations or the lower aboundance of low molecular weight proteins. Therefore, further study is required with and without filtration to identify proteins sequences that are hydrolyzed due to seed germination.

5.5 Conclusions

To conclude, the protein concentration was reduced as seed germination progressed. Few development related proteins were present in the control seeds. Some vital enzymes and proteins

for seed germination are ascribed from the seed developmental stage. The production of acetyl-CoA was limited to until germination phase II and therefore, the anaerobic respiration was the only source of energy supply for cellular activity. General energy metabolism was initiated in phase III while cellular multiplication and biosynthesis of biomolecules and cell wall components were more visible in phase IV. From the protein quantitation data, a significant change in germination and defense related proteins can be observed.

Chapter 6

General iscussion and future work

The first objective was to define phases of seed germination and post-germination in lentil species. This is the first report that defines the seed germination phases of the wild lentil species. The duration of the germination phases varies among the species. Species with smaller seed size imbibed faster than the species with larger seed size. This is similar to what occurs with seeds of other plant species that exhibit a wide range of seed size (Nakayama and Komatsu, 2008). With regard to individual seeds, the interim period between germination phases are similar for seed belong to genotypes, while not all seeds within same genotype germinated at period after soaking. The pattern of seed germination varied among lentil species in terms period and percentage to radicle and plumule emergence.

The second objective of the study was to compare the protease activity and free amino acid profiles of the lentil species during different phases of seed germination. Protease activity was increased in germination phases III and IV. Therefore, it can be suggested that storage protein hydrolysis, peptide and aminoacid release will be higher in phase III and phase IV. As a result, the lentil protein fraction after radicle and plumule emergence could be more digestible and contain higher amounts of bioavailable peptides and aminoacids. With regards to the free amino acids, overall, the level of both essential and non-essential amino acids increased significantly in phases III and IV, although variations were observed in the percentage of increase among the genotypes studied. IG 72623, IG 72815, and IG 72541 genotypes showed an overall increase of NEFFAs (non-essential free amino acids); while in BGE 016880, IG 110813 and IG 72623, the increase EFFAs (essential free amino acids) was higher. Individual free amino acid analysis showed that asparagine (ASN) and histidine (HIS) had the highest abundance in all lentil species studied. Cysteine (CYS) was present in dry intact seeds of CDC Redberry, Eston, IG 72805, IG 72623, and IG 72815 and was no longer detectable from phase II of seed germination. Based on the study of overall FEAAs, isoleucine (ILE), lysine (LYS), leucine (LEU), valine (VAL), and threonine (THR) were increased through the process of seed germination and BGE 016880, IG 72623 and IG 110813 was found as the best genotypes among all the genotypes studied in terms of increase in FEAAs in phase IV

compared to control. The FEAAs increased siginificantly in BGE 016880, IG 110813, and IG 72623, and the increase was 1344%, 752%, and 723%, respectively from dry seed to phase IV.

The third objective was to determine what changes occur in the proteome due to seed germination of the seven lentil species as a way to elucidate the major metabolic events during seed germination. It was determined that proteins involved in photosynthesis were mostly present in all the phases studied, including the control seeds. A few components were found to be present only in specific germination phases. Further KEGG analysis showed not all components involved photosynthesis were present in all the phases studied. Enzymes involved in carbohydrate degradation were abundant in phase 0, phase II, and phase III compared to phase IV, which shows that carbohydrate utilization initiates right after phase 0. From this study it was also found that the acetyl CoA production was initiated after phase II, therefore TCA cycle activity was limited up to phase II, while some activity might have been performed by the cell from seed reserves, as the enzymes involved in the TCA cycle waere present in all the phases.

Some of the proteins involved in protein biosynthesis process were found in phase 0, while others started synthesizing within cells during phase II. The biosynthesis of new proteins involved in organizing cellular components through external stimuli response, and in vesicle trafficking were initiated during phase III. While some proteins involved in redox homeostasis pathway were present from the initial phase, some that started synthesizing during phase III were also found from this study. While JA biosynthesis proteins were found in phase IV, some auxin, cytokinin biosynthesis related proteins were not shown to be consistently present due to their being observed in 25% of the best four selected samples. Some neosynthesized proteins involved in lipid catabolism, lipid biosynthesis, amino acid biosynthesis, protein modification, and protein translocation were found in phase IV. While there are studies that report on proteome identification at specific intervals of seed germination period in other plant species, proteome identification following phase determination has not been reported in lentil, or in other plant species (Catusse et al., 2008; Sreenivasulu et al., 2008; Howell et al., 2009; Jiménez-López et al., 2011; He and Yang, 2013; Galland et al., 2014; Yu et al., 2014; Dong et al., 2015).

The other aims of this project were to identify the released bioactive peptides during distinct phases of seed germination post simulated gastrointestinal digestion, which was attempted but could not

be accomplished. For this experiment, seed sample collection from distinct phases and simulated gastrointestinal digestion methods was attempted but could not proceed further due to instrumental limitations as well as the limitation of timeline as the repetition of proteome identification experiment with all the genotypes took an extended period of time. As the novelty of the study is to conduct experiments with samples collected from distinct phases, an attempt was taken to identify naturally released peptides through germination process but this could not be accomplished due to instrumental limitations as well. The limitation of this study was overabundance of storage protein during extraction. Therefore, it could be suggested to develop a method for screening low molecular weight protein filtration techniques to identify the naturally released peptide. Identification of naturally released peptides would help with determining about category of peptides that are being released during seed germination. Protein digestibility score determination is also essential. On the other hand, categorization of peptide identification followed by simulated gastrointestinal systems will allow us to get a clear understanding of the categories of peptides that arevreleased through human consumption, and if the intake of sprouted lentils are beneficial over the not germinated seed or not.

Overall, this study has found differences in amino acid profile, protease activity, and in proteome profile among distinct phases of seed germination. The aforementioned further studies will provide a transparent overview of protein quality with regards to human consumption, and could lead to determination of which germination phase is most beneficial for human consumption.

References

Adenekan, M. K., Fadimu, G. J., Odunmbaku, L. A., and Oke, E. K. (2018). Effect of isolation techniques on the characteristics of pigeon pea (*Cajanus cajan*) protein isolates. *Food Science and Nutrition*, *6*(1), 146–152. https://doi.org/10.1002/fsn3.539.

Adhikari, B., Adhikari, M., Ghimire, B., Adhikari, B. C., Park, G., and Choi, E. H. (2020). Cold plasma seed priming modulates growth, redox homeostasis and stress response by inducing reactive species in tomato (*Solanum lycopersicum*). *Free Radical Biology and Medicine*, *156*, 57-69. https://doi.org/10.1016/j.freeradbiomed.2020.06.003.

Afify, A. E. M. M., El-Beltagi, H. S., Abd El-Salam, S. M., and Omran, A. A. (2012). Protein solubility, digestibility and fractionation after germination of sorghum varieties. *Plos One*, *7*(2), e31154. https://doi.org/10.1371/journal.pone.0031154.

Aggarwal, A., and Drewnowski, A. (2019). Plant- and animal-protein diets in relation to sociodemographic drivers, quality, and cost: Findings from the Seattle Obesity Study. *The American Journal of Clinical Nutrition*, *110*(2), 451–460. https://doi.org/10.1093/ajcn/nqz064.

Aguilar, E. G., Albarracín, G. de J., Uñates, M. A., Piola, H. D., Camiña, J. M., and Escudero, N. L. (2015). Evaluation of the nutritional quality of the grain protein of new amaranths varieties. *Plant Foods for Human Nutrition*, *70*(1), 21–26. https://doi.org/10.1007/s11130-014-0456-3.

Alberta Pulse Growers. Lentils. [(accessed on 10 November 2022)]; Available online: https://albertapulse.com/marketing-pulses/lentils.

Ali, A. S., and Elozeiri, A. A. (2017). Metabolic processes during seed germination. *Advances in Seed Biology*, 141-166. https://doi.org/10.5772/intechopen.70653.

Al-Islam, M., Rabah, H., and Yousef, A. (2012). Role of lentils (*Lens culinaris* L.) in human health and nutrition: a review. *Mediterranean Journal of Nutrition and Metabolism*, *6*, 3-16. https://doi.org/10.1007/s12349-012-0109-8.

Al-Karaki, G. N. (1998). Seed size and water potential effects on water uptake, germination and growth of lentil. *Journal of Agronomy and Crop Science*, *181*(4), 237-242. https://doi.org/10.1111/j.1439-037X.1998.tb00423.x.

Annor, G. A., Tyl, C., Marcone, M., Ragaee, S., and Marti, A. (2017). Why do millets have slower starch and protein digestibility than other cereals? *Trends in Food Science and Technology*, *66*(2017), 73–83. https://doi.org/10.1016/j.tifs.2017.05.012.

Argos, P., Narayana, S. V., and Nielsen, N. C. (1985). Structural similarity between legumin and vicilin storage proteins from legumes. *The EMBO Journal*, *4*(5), 1111-1117. PMCID: PMC554311.

Arribas, C., Cabellos, B., Sánchez, C., Cuadrado, C., Guillamón, E., and Pedrosa, M. M. (2017). The impact of extrusion on the nutritional composition, dietary fiber and in vitro digestibility of gluten-free snacks based on rice, pea and carob flour blends. *Food and*

Function, *8*(10), 3654–3663. https://doi.org/10.1039/C7FO00910K.

Avilés-Gaxiola, S., Chuck-Hernández, C., and Serna Saldívar, S. O. (2018). Inactivation methods of trypsin inhibitor in legumes: a review. *Journal of Food Science*, *83*(1), 17-29. https://doi.org/10.1111/1750-3841.13985.

Bai, B., Van der Horst, N., Cordewener, J. H., America, A. H., Nijveen, H., and Bentsink, L. (2021). Delayed protein changes during seed germination. *Frontiers in Plant Science*, *12*, 735719. https://doi.org/10.3389/fpls.2021.735719.

Bamdad, F., Dokhani, S., Keramat, J. (2009a). Functional assessment and subunit constitution of lentil (*Lens culinaris*) proteins during Germination. *International Journal of Agriculture and Biology*, 11(6), 690-694. ISSN: 1560-8530.

Bamdad, F., Dokhani, S., Keramat, J., Zareie, R. (2009b). 'The Impact of Germination and In Vitro Digestion on the Formation of Angiotensin Converting Enzyme (ACE) Inhibitory Peptides from Lentil Proteins Compared to Whey Proteins'. *World Academy of Science, Engineering and Technology, Open Science Index 25, International Journal of Nutrition and Food Engineering*, *3*(1), 109 - 119. https://doi.org/10.5281/zenodo.1327470.

Barbana, C., and Boye, J. I. (2011). Angiotensin I-converting enzyme inhibitory properties of lentil protein hydrolysates: Determination of the kinetics of inhibition. *Food Chemistry*, *127*(1), 94-101. https://doi.org/10.1016/j.foodchem.2010.12.093.

Bastow, E. L., Bych, K., Crack, J. C., Le Brun, N. E., and Balk, J. (2017). NBP 35 interacts with DRE 2 in the maturation of cytosolic iron-sulphur proteins in *Arabidopsis thaliana*. *The Plant Journal*, *89*(3), 590-600. https://doi.org/10.1111/tpj.13409.

Bautista-Expósito, S., Peñas, E., Vanderberg, A., Frias, J., Martínez-Villaluenga, C. (2020). Effect of time and legume type on germination-induced proteolysis of lentils and faba beans. *Multidisciplinary Digital Publishing Institute Proceedings*, *70*(1), 4. https://doi.org/10.3390/foods_2020-07823.

Benincasa, P., Falcinelli, B., Lutts, S., Stagnari, F., and Galieni, A. (2019). Sprouted grains: A comprehensive review. *Nutrients*, *11*(2), 421. https://doi.org/10.3390/nu11020421.

Berno, L. I., Guimarães-lopes, T. G., and Canniatti-Brazaca, S. G. (2007). Avaliação da composição centesimal, digestibilidade e atividade inibitória de tripsina em produtos derivados de soja (*Glycine max*). *Alimentos e Nutricão*, *18*(3), 277–282. ISSN 0103-4235.

Bewley, J. D. (1997). Seed germination and dormancy. *The Plant Cell*, *9*(7), 1055. https://doi.org/10.1105/tpc.9.7.1055.

Bewley, J. D., Bradford, K., & Hilhorst, H. (2012). *Seeds: physiology of development, germination and dormancy*. Springer Science and Business Media. https://doi.org/10.1007/978-1-4614-4693-4.

Bhatty, R. S. (1986). Protein subunits and amino acid composition of wild lentil. *Phytochemistry*, *25*(3), 641–644. https://doi.org/10.1016/0031-9422(86)88015-3.

Bhatty, R. S. (1988). Composition and quality of lentil (*Lens culinaris* Medik): A review.

Canadian Institute of Food Science and Technology Journal, *21*(2), 144–160. https://doi.org/10.1016/S0315-5463(88)70770-1.

Blalogoe, J. S., Odindo, A. O., Sogbohossou, E. D., Sibiya, J., and Achigan-Dako, E. G. (2020). Origin-dependence of variation in seed morphology, mineral composition and germination percentage in *Gynandropsis gynandra* (L.) Briq. accessions from Africa and Asia. *BMC Plant Biology*, *20*(1), 1-14. https://doi.org/10.1186/s12870-020-02364-w.

Bourgeois, M., Jacquin, F., Savois, V., Sommerer, N., Labas, V., Henry, C., and Burstin, J. (2009). Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics*, *9*(2), 254–271. https://doi.org/10.1002/pmic.200700903.

Boye, J., Zare, F., and Pletch, A. (2010). Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Research International*, *43*(2), 414-431. https://doi.org/10.1016/j.foodres.2009.09.003.

Buckeridge, M. S. (2010). Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. *Plant Physiology*, *154*(3), 1017-1023. https://doi.org/10.1104/pp.110.158642.

Carrera, E., Holman, T., Medhurst, A., Dietrich, D., Footitt, S., Theodoulou, F. L., and Holdsworth, M. J. (2008). Seed after-ripening is a discrete developmental pathway associated with specific gene networks in *Arabidopsis*. *The Plant Journal*, *53*(2), 214-224. https://doi.org/10.1111/j.1365-313X.2007.03331.x.

Catusse, J., Job, C., & Job, D. (2008). Transcriptome-and proteome-wide analyses of seed germination. *Comptes Rendus Biologies*, *331*(10), 815-822. https://doi.org/10.1016/j.crvi.2008.07.023.

Chakrabarty, A., Banik, N., and Bhattacharjee, S. (2019). Redox-regulation of germination during imbibitional oxidative and chilling stress in an indica rice cultivar (*Oryza sativa* L., Cultivar Ratna). *Physiology and Molecular Biology of Plants*, *25*(3), 649-665. https://doi.org/10.1007/s12298-019-00656-6.

Chitra, U., Singh, U., and Venkateswara Rao, P. (1996). Phytic acid, in vitro protein digestibility, dietary fiber, and minerals of pulses as influenced by processing methods. *Plant Foods for Human Nutrition*, *49*(4), 307-316. https://doi.org/10.1007/BF01091980.

Cideciyan, M. A., and Malloch, A. J. (1982). Effects of seed size on the germination, growth and competitive ability of *Rumex crispus* and *Rumex obtusifolius*. *The Journal of Ecology*, 227-232. https://doi.org/10.2307/2259875.

Córdoba-Cañero, D., Roldán-Arjona, T., and Ariza, R. R. (2014). *Arabidopsis* ZDP DNA 3′-phosphatase and ARP endonuclease function in 8-oxoG repair initiated by FPG and OGG 1 DNA glycosylases. *The Plant Journal*, *79*(5), 824-834. https://doi.org/10.1111/tpj.12588.

Czarna, M., Kolodziejczak, M., and Janska, H. (2016). Mitochondrial proteome studies in seeds during germination. *Proteomes*, *4*(2), 19. https://doi.org/10.3390/proteomes4020019.

Damaris, R. N., Lin, Z., Yang, P., and He, D. (2019). The rice alpha-amylase, conserved regulator of seed maturation and germination. *International Journal of Molecular Sciences*, *20*(2), 450. https://doi.org/10.3390/ijms20020450.

De Souza Rocha, T., Hernandez, L. M. R., Mojica, L., Johnson, M. H., Chang, Y. K., and González de Mejía, E. (2015). Germination of *Phaseolus vulgaris* and alcalase hydrolysis of its proteins produced bioactive peptides capable of improving markers related to type-2 diabetes in vitro. *Food Research International*, *76*(P1), 150–159. https://doi.org/10.1016/j.foodres.2015.04.041.

Díaz-Camino, C., Conde, R., Ovsenek, N., and Villanueva, M. A. (2005). Actin expression is induced and three isoforms are differentially expressed during germination in *Zea mays*. *Journal of Experimental Botany*, *56*(412), 557-565. https://doi.org/10.1093/jxb/eri034.

Dogra, V., Ahuja, P. S., and Sreenivasulu, Y. (2013). Change in protein content during seed germination of a high altitude plant *Podophyllum hexandrum* Royle. *Journal of Proteomics*, *78*, 26-38. https://doi.org/10.1016/j.jprot.2012.10.025.

Dong, K., Zhen, S., Cheng, Z., Cao, H., Ge, P., & Yan, Y. (2015). Proteomic analysis reveals key proteins and phosphoproteins upon seed germination of wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*, *6*, 1017. https://doi.org/10.3389/fpls.2015.01017.

Duodu, K. ., Taylor, J. R. ., Belton, P. ., and Hamaker, B. . (2003). Factors affecting sorghum protein digestibility. *Journal of Cereal Science*, *38*(2), 117–131. https://doi.org/10.1016/S0733-5210(03)00016-X.

Elemo, G. N., Elemo, B. O., Okafor, J. N. C. (2011). Preparation and nutritional compostion of a weaning food formulated from germinated sorghum (*Sorghum bicolor*) and steamed cooked cowpea (*Vigna unguiculata* Walp.). *American Journal of Food Science and Technology*, *6*(5), 413–421. https://doi.org/10.3923/ajft.2011.413.421.

El-Maarouf-Bouteau, H., and Bailly, C. (2008). Oxidative signaling in seed germination and dormancy. *Plant Signaling and Behavior*, *3*(3), 175-182. https://doi.org/10.4161/psb.3.3.5539.

El-Mahdy, A. R., El-Sebaiy, L. A. (1985). Proteolytic activity, amino acid composition and protein quality of germinating fenugreek seeds (*Trigonella foenum-graecum* L.). *Food Chemistry*, *18*(1), 19-33. https://doi.org/10.1016/0308-8146(85)90100-1.

EL-Suhaibani, M., Ahmed, M. A., and Osman, M. A. (2020). Study of germination, soaking and cooking effects on the nutritional quality of goat pea (*Securigera securidaca* L.). *Journal of King Saud University - Science*, *32*(3), 2029–2033. https://doi.org/10.1016/j.jksus.2020.02.021.

Erskine, W, Muehlbauer, F. J., Sarker, A., Sharma, B., and Erskine, William, Muehlbauer, Fred J., Sarker, Ashutosh, Sharma, B. (2009). The lentil Botany, production and uses. In M. F. J. Sharma Balram, Erskine William, Sarker Ashutosh (Ed.), *The Lentil Botany, Production and Uses*. https://doi.org/10.1079/9781845934873.0000.

Erskine, W., Sarker, A., and Kumar, S. (2016). Lentil Breeding. *Reference Module in Food Science* (2nd ed., pp. 1–8). Elsevier. https://doi.org/10.1016/B978-0-08-100596-5.00210-9.

Fait, A., Nesi, A. N., Angelovici, R., Lehmann, M., Pham, P. A., Song, L., Haslam, R. P., Napier, J. A., Galili, G., and Fernie, A. R. (2011). Targeted enhancement of glutamate-to-γ-aminobutyrate conversion in *Arabidopsis* seeds affects carbon-nitrogen balance and storage reserves in a development-dependent manner. *Plant Physiology*, *157*(3), 1026–1042. https://doi.org/10.1104/pp.111.179986.

Feng, H., Nemzer, B., and Devries, J. W. (Eds.). (2018). Sprouted grains: Nutritional value, production, and applications. *Elsevier*, eBook ISBN: 9780128115268.

Fenner, M. K., Fenner, M., and Thompson, K. (2005). The ecology of seeds. *Cambridge University Press*. (pp. V-Vii). https://doi.org/10.1017/CBO9780511614101.

Ferguson, M. E., Fias, N. M., Slageren, M. V. A. N., & S, L. D. R. (2000). A re-assessment of the taxonomy of Lens Mill . *New York*, 41–59. https://doi.org/10.1111/j.1095-8339.2000.tb01536.x.

Fernie, A. R., Cavalcanti, J. H. F., and Nunes-Nesi, A. (2020). Metabolic roles of plant mitochondrial carriers. *Biomolecules*, *10*(7), 1013. https://doi.org/10.3390/biom10071013.

Ferreira, C. D. S., Piedade, M. T. F., Tiné, M. A. S., Rossatto, D. R., Parolin, P., and Buckeridge, M. S. (2009). The role of carbohydrates in seed germination and seedling establishment of *Himatanthus sucuuba*, an Amazonian tree with populations adapted to flooded and non-flooded conditions. *Annals of Botany*, *104*(6), 1111-1119. https://doi.org/10.1093/aob/mcp212.

Flikweert, M. T., van der Zanden, L., Janssen, W. M. T. M., Yde Steensma, H., van Dijken, J. P., and Pronk, J. T. (1996). Pyruvate decarboxylase: an indispensable enzyme for growth of *Saccharomyces cerevisiae* on glucose. *Yeast*, *12*(3), 247-257. PMCID: PMC106289.

Food and Agriculture Organisation. Pulses and Derived Products. [(accessed on 24 February 2022)]; Available online: https://www.fao.org/faostat/en/#data/QCL.

Fouad, A. A., and Rehab, F. M. (2015). Effect of germination time on proximate analysis, bioactive compounds and antioxidant activity of lentil (*Lens culinaris* Medik.) sprouts. *Acta Scientiarum Polonorum Technologia Alimentaria*, *14*(3), 233-246. https://doi.org/10.17306/J.AFS.2015.3.25.

Fratini, R., Ruiz, M. L., and de la Vega, M. P. (2004). Intra-specific and inter-sub-specific crossing in lentil (*Lens culinaris* Medik.). *Canadian Journal of Plant Science, 84*(4), 981-986. https://doi.org/10.4141/P03-201.

Galili, G., Avin-Wittenberg, T., Angelovici, R., and Fernie, A. R. (2014). The role of photosynthesis and amino acid metabolism in the energy status during seed development. *Frontiers in Plant Science, 5*, 447. https://doi.org/10.3389/fpls.2014.00447.

Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D., and Rajjou, L. (2014). Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during *Arabidopsis* seed germination. *Molecular and Cellular Proteomics*, *13*(1), 252-268. https://doi.org/10.1074/mcp.M113.032227.

Ganesan, K., & Xu, B. (2017). Polyphenol-rich lentils and their health promoting

effects. *International Journal of Molecular Sciences*, *18*(11), 2390. https://doi.org/10.3390/ijms18112390.

Gendreau, E., Cayla, T., and Corbineau, F. (2012). S phase of the cell cycle: a key phase for the regulation of thermodormancy in barley grain. *Journal of Experimental Botany*, *63*(15), 5535-5543. https://doi.org/10.1093/jxb/ers204.

Ghavidel, R. A., and Prakash, J. (2007). The impact of germination and dehulling on nutrients, antinutrients, in vitro iron and calcium bioavailability and in vitro starch and protein digestibility of some legume seeds. *LWT-Food Science and Technology*, 40(7), 1292-1299. https://doi.org/10.1016/j.lwt.2006.08.002.

Girolamo, D., G., and Barbanti, L. (2012). Treatment conditions and biochemical processes influencing seed priming effectiveness. *Italian Journal of Agronomy*, *7*(2), e25-e25. https://doi.org/10.4081/ija.2012.e25.

Gomez, L. D., Steele-King, C. G., Jones, L., Foster, J. M., Vuttipongchaikij, S., and McQueen-Mason, S. J. (2009). Arabinan metabolism during seed development and germination in *Arabidopsis*. *Molecular Plant*, *2*(5), 966-976. https://doi.org/10.1093/mp/ssp050.

Grelet, J., Benamar, A., Teyssier, E., Avelange-Macherel, M. H., Grunwald, D., and Macherel, D. (2005). Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiology*, *137*(1), 157-167. https://doi.org/10.1104/pp.104.052480.

Gu, J., Chao, H., Gan, L., Guo, L., Zhang, K., Li, Y., and Li, M. (2016). Proteomic dissection of seed germination and seedling establishment in *Brassica napus*. *Frontiers in Plant Science*, *7* (1482). https://doi.org/10.3389/fpls.2016.01482.

Gu, J., Hou, D., Li, Y., Chao, H., Zhang, K., Wang, H., and Li, M. (2019). Integration of proteomic and genomic approaches to dissect seed germination vigor in *Brassica napus* seeds differing in oil content. *BMC Plant Biology*, *19*(1), 1-20. https://doi.org/10.1186/s12870-018-1624-7.

Guo, B., Chen, Y., Zhang, G., Xing, J., Hu, Z., Feng, W, and Sun, Q. (2013). Comparative proteomic analysis of embryos between a maize hybrid and its parental lines during early stages of seed germination. *PLoS One*, *8*(6), e65867. https://doi.org/10.1371/journal.pone.0065867.

He, D., and Yang, P. (2013). Proteomics of rice seed germination. *Frontiers in Plant Science*, *4*, 246. https://doi.org/10.3389/fpls.2013.00246.

Hefnawy, T. H. (2011). Effect of processing methods on nutritional composition and anti-nutritional factors in lentils (*Lens culinaris*). *Annals of Agricultural Sciences*, *56*(2), 57-61. Dey, P. M., and Harborne, J. B. (Eds.). (1997). *Plant Biochemistry*. Elsevier. https://doi.org/10.1016/j.aoas.2011.07.001.

Herman, E. M., and Larkins, B. A. (1999). Protein storage bodies and vacuoles. *The Plant Cell*, *11*(4), 601-613. https://doi.org/10.1105/tpc.11.4.601.

Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. *Cell*, *136*(4), 763-776. https://doi.org/10.1016/j.cell.2009.01.019.

Howell, K. A., Narsai, R., Carroll, A., Ivanova, A., Lohse, M., Usadel, B., & Whelan, J. (2009). Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. *Plant Physiology*, *149*(2), 961-980. https://doi.org/10.1104/pp.108.129874.

Hu, Y., Zou, W., Wang, Z., Zhang, Y., Hu, Y., Qian, J., and Zhao, J. (2019). Translocase of the outer mitochondrial membrane 40 is required for mitochondrial biogenesis and embryo development in *Arabidopsis*. *Frontiers in Plant Science*, *10*, 389. https://doi.org/10.3389/fpls.2019.00389.

Hulse, J. H., Rachie, K. O., and Billingsley, L. W. (2003). *Nutritional standards and methods of evaluation for food legume breeders prepared* (Issue 1). International Development Research Centre, Ottawa, Canada. https://doi.org/10.16309/j.cnki.issn.1007-1776.2003.03.004.

Jarpa-Parra, M. (2018). Lentil protein: A review of functional properties and food application. An overview of lentil protein functionality. *International Journal of Food Science and Technology*, *53*(4), 892-903. https://doi.org/10.1111/ijfs.13685.

Jiménez-López, S., Mancera-Martínez, E., Donayre-Torres, A., Rangel, C., Uribe, L., March, S., & Sánchez de Jiménez, E. (2011). Expression profile of maize (*Zea mays* L.) embryonic axes during germination: translational regulation of ribosomal protein mRNAs. *Plant and Cell Physiology*, *52*(10), 1719-1733. https://doi.org/10.1093/pcp/pcr114.

Kasai, T., Ishikawa, Y., Obata, Y., Tsukamoto, T. (1966). Changes in amino acid composition during germination of soybean: Part I. Changes in free amino acids, several nitrogen compounds and total amino acids part II. Identification of Two γ-Glutamyl peptides and their Change during Germination. *Agricultural and Biological Chemistry*, *30*(10), 973-981. https://doi.org/10.1080/00021369.1966.10858716.

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*. https://doi.org/10.1093/aobpla/plr026.

Kestring, D., Klein, J., de Menezes, L. C. C. R., and Rossi, M. N. (2009). Imbibition phases and germination response of *Mimosa bimucronata* (Fabaceae: Mimosoideae) to water submersion. *Aquatic Botany*, *91*(2), 105-109. https://doi.org/10.1016/j.aquabot.2009.03.004.

Khazaei, H., Podder, R., Caron, C. T., Kundu, S. S., Diapari, M., Vandenberg, A., and Bett, K. E. (2017). Marker-trait association analysis of iron and zinc concentration in lentil (*Lens culinaris* Medik.) seeds. *The Plant Genome*, *10*(2), plantgenome2017.02.0007. https://doi.org/10.3835/plantgenome2017.02.0007.

Khazaei, H., Subedi, M., Nickerson, M., Martínez-Villaluenga, C., Frias, J., Vandenberg, A. (2019). Seed protein of lentils: Current status, progress, and food applications. *Foods*, *8*(9), 391. https://doi.org/10.3390/foods8090391.

Kiran, K. R., Deepika, V. B., Swathy, P. S., Prasad, K., Kabekkodu, S. P., Murali, T. S., and Muthusamy, A. (2020). ROS-dependent DNA damage and repair during germination of NaCl primed seeds. *Journal of Photochemistry and Photobiology B: Biology*, *213*,

112050. https://doi.org/10.1016/j.jphotobiol.2020.112050.

Kitajima, K. and Myers J. A. (2008). Seedling ecophysiology; strategies toward achievement of positive net carbon balance. In: Leck MA, Parker TV, Simpson RL, editors. Seedling ecology and evolution. Cambridge: Cambridge University Press; 2008. pp. 172–188. ISBN : 9780521694667.

Kniskern, M. A., and Johnston, C. S. (2011). Protein dietary reference intakes may be inadequate for vegetarians if low amounts of animal protein are consumed. *Nutrition*, *27*(6), 727–730. https://doi.org/10.1016/j.nut.2010.08.024.

Koizumi, M., Kikuchi, K., Isobe, S., Ishida, N., Naito, S., and Kano, H. (2008). Role of seed coat in imbibing soybean seeds observed by micro-magnetic resonance imaging. *Annals of Botany*, *102*(3), 343-352.

Kumar, J., Singh, J., Kanaujia, R., and Gupta, S. (2016). Protein content in wild and cultivated taxa of lentil (*Lens culinaris* ssp. *culinaris* Medikus). *Indian Journal of Genetics and Plant Breeding (The)*, *76*(4), 631. https://doi.org/10.5958/0975-6906.2016.00078.X.

Kuo, Y. H., Rozan, P., Lambein, F., Frias, J., Vidal-Valverde, C. (2004). Effects of different germination conditions on the contents of free protein and non-protein amino acids of commercial legumes. *Food Chemistry*, *86*(4), 537-545. https://doi.org/10.1016/j.foodchem.2003.09.042.

Lechowska, K., Kubala, S., Wojtyla, Ł., Nowaczyk, G., Quinet, M., Lutts, S., and Garnczarska, M. (2019). New insight on water status in germinating *Brassica napus* seeds in relation to priming-improved germination. *International Journal of Molecular Sciences*, *20*(3), 540. https://doi.org/10.3390/ijms20030540.

Lehmann, T., and Ratajczak, L. (2008). The pivotal role of glutamate dehydrogenase (GDH) in the mobilization of N and C from storage material to asparagine in germinating seeds of yellow lupine. *Journal of Plant Physiology*, *165*(2), 149-158. https://doi.org/10.1016/j.jplph.2006.12.010.

Liu, H., Liu, B., Zhao, C., Pepper, M., and Lin, C. (2011). The action mechanisms of plant cryptochromes. *Trends in Plant Science*, *16*(12), 684-691. https://doi.org/10.1016/j.tplants.2011.09.002.

Liu, Y., Ragaee, S., Marcone, M. F., and Abdel-Aal, E. S. M. (2020). Composition of phenolic acids and antioxidant properties of selected pulses cooked with different heating conditions. *Foods*, *9*(7), 908. https://doi.org/10.3390/foods9070908.

López, D. N., Galante, M., Robson, M., Boeris, V., and Spelzini, D. (2018). Amaranth, quinoa and chia protein isolates: Physicochemical and structural properties. *International Journal of Biological Macromolecules*, *109*(2018), 152–159. https://doi.org/10.1016/j.ijbiomac.2017.12.080.

López-Millán, A. F., Duy, D., and Philippar, K. (2016). Chloroplast iron transport proteins–function and impact on plant physiology. *Frontiers in Plant Science*, *7*, 178. https://doi.org/10.3389/fpls.2016.00178.

Lu, Y. (2018). Assembly and transfer of iron–sulfur clusters in the plastid. *Frontiers in Plant Science, 9*, 336. https://doi.org/10.3389/fpls.2018.00336.

Lutts, S., Benincasa, P., Wojtyla, L., Kubala, S., Pace, R., Lechowska, K, and Garnczarska, M. (2016). Seed priming: new comprehensive approaches for an old empirical technique. *New Challenges in seed Biology- Basic and Translational Research Driving Seed Technology*, 1-46. https://doi.org/10.5772/64420.

Mandal, S. M., Dipjyoti, C., and Kajal, G. (2010). Seed size variation: influence on germination and subsequent seedling performance in *Hyptis suaveolens* (Lamiaceae). *Research Journal of Seed Science, 3*(3), 185-192. https://doi.org/10.3923/rjss.2008.26.33.

Manz, B., Muller, K., Kucera, B., Volke, F., and Leubner-Metzger, G. (2005). Water uptake and distribution in germinating tobacco seeds investigated in vivo by nuclear magnetic resonance imaging. *Plant Physiology, 138*(3), 1538-1551. https://doi.org/10.1104/pp.105.061663.

Marambe, P. W. M. L. H. K., and Wanasundara, J. P. D. (2014). Seed storage proteins as sources of bioactive peptides. In 2012 Marambe and Wanasundara (Ed.), *Bioactive Molecules in Plant Foods* (Issue January 2012).

Marcela, G.-M., Eva, R.-G., del Carmen, R.-R. M., and Rosalva, M.-E. (2016). Evaluation of the antioxidant and antiproliferative effects of three peptide fractions of germinated soybeans on breast and cervical cancer cell lines. *Plant Foods for Human Nutrition, 71*(4), 368–374. https://doi.org/10.1007/s11130-016-0568-z.

Marcone, M. F., Kakuda, Y. (1999). A comparative study of the functional properties of amaranth and soybean globulin isolates. *Food/Nahrung, 43*(6), 368-373. https://doi.org/10.1002/(SICI)1521-3803(19991201)43:6%3C368::AID-FOOD368%3E3.0.CO;2-R.

Mbithi-Mwikya, S., Van Camp, J., Yiru, Y., and Huyghebaert, A. (2000). Nutrient and antinutrient changes in finger millet (*Eleusine coracan*) during sprouting. *LWT-Food Science and Technology, 33*(1), 9-14. https://doi.org/10.1006/fstl.1999.0605.

Miernyk, J. A., and Hajduch, M. (2011). Seed proteomics. *Journal of Proteomics, 74*(4), 389-400. https://doi.org/10.1016/j.jprot.2010.12.004.

Müntz, K., Belozersky, M. A., Dunaevsky, Y. E., Schlereth, A., and Tiedemann, J. (2001). Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. *Journal of Experimental Botany, 52*(362), 1741-1752. https://doi.org/10.1093/jexbot/52.362.1741.

Nakayama, N., and Komatsu, S. (2008). Water uptake by seeds in yellow-seeded soybean (*Glycine max* (L.) Merrill) cultivars with contrasting imbibition behaviors. *Plant Production Science, 11*(4), 415-422. https://doi.org/10.1626/pps.11.415.

Nkhata, S. G., Ayua, E., Kamau, E. H., and Shingiro, J.-B. (2018). Fermentation and germination improve nutritional value of cereals and legumes through activation of endogenous enzymes. *Food Science and Nutrition, 6*(8), 2446–2458. https://doi.org/10.1002/fsn3.846.

Nonogaki, H., Bassel, G. W., and Bewley, J. D. (2010). Germination—still a mystery. *Plant*

Science, *179*(6), 574-581. https://doi.org/10.1016/j.plantsci.2010.02.010.

Ogata, Y., Iizuka, M., Nakayama, D., Ikeda, M., Kamada, H., and Koshiba, T. (2005). Possible involvement of abscisic acid in the induction of secondary somatic embryogenesis on seed-coat-derived carrot somatic embryos. *Planta*, *221*(3), 417-423. https://doi.org/10.1007/s00425-004-1449-5.

Ogawa, T., Kimura, A., Sakuyama, H., Tamoi, M., Ishikawa, T., and Shigeoka, S. (2015). Identification and characterization of cytosolic fructose-1, 6-bisphosphatase in *Euglena gracilis*. *Bioscience, Biotechnology, and Biochemistry*, *79*(12), 1957-1964. https://doi.org/10.1080/09168451.2015.1069694.

Ohanenye, I. C., Ekezie, F. G. C., Sarteshnizi, R. A., Boachie, R. T., Emenike, C. U., Sun, X., & Udenigwe, C. C. (2022). Legume seed protein digestibility as influenced by traditional and emerging physical processing technologies. *Foods*, *11*(15), 2299. https://doi.org/10.3390/foods11152299.

Orozco-Segovia, A., Márquez-Guzmán, J., Sánchez-Coronado, M. E., Gamboa de Buen, A., Baskin, J. M., and Baskin, C. C. (2007). Seed anatomy and water uptake in relation to seed dormancy in *Opuntia tomentosa* (Cactaceae, Opuntioideae). *Annals of Botany*, *99*(4), 581-592. https://doi.org/10.1093/aob/mcm001.

Osborne T. B. (1924). The vegetable proteins second edition. Longmans, green and Co. *London Plant Physiology*. *154*,1842–1854.

Patil, S. B., and Khan, M. (2011). Germinated brown rice as a value-added rice product: A review. *Journal of Food Science and Technology*, *48*(6), 661-667. https://doi.org/10.1007/s13197-011-0232-4.

Penfield, S., Pinfield-Wells, H. M., and Graham, I. A. (2006). Storage reserve mobilization and seedling establishment in *Arabidopsis*. *The Arabidopsis Book*. *American Society of Plant Biologists*, *4*. https://doi.org/10.1199/tab.0100.

Petrova, T., Marinova, M., Tchorbanov, B. (2010). Dynamics of some hydrolytic enzymes during the sprouts production from lentil seeds (*Lens culinaris*). *Biotechnology Biotechnological Equipment*, *24*(4), 2102-2107. https://doi.org/10.2478/V10133-010-0095-2.

Pirovani, C. P., Macêdo, J. N. A., Contim, L. A. S., Matrangolo, F. S., Loureiro, M. E., and Fontes, E. P. (2002). A sucrose binding protein homologue from soybean exhibits GTP-binding activity that functions independently of sucrose transport activity. *European Journal of Biochemistry*, *269*(16), 3998-4008. https://doi.org/10.1046/j.1432-1033.2002.03089.x.

Pogson, B. J., Ganguly, D., and Albrecht-Borth, V. (2015). Insights into chloroplast biogenesis and development. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1847*(9), 1017-1024. https://doi.org/10.1016/j.bbabio.2015.02.003.

Polak, R., Phillips, E. M., and Campbell, A. (2015). Legumes: Health benefits and culinary approaches to increase intake. *Clinical Diabetes*, 33(4), 198–205. https://doi.org/10.2337/diaclin.33.4.198.

Potier, M., and Tomé, D. (2008). Comparison of digestibility and quality of intact proteins with their respective hydrolysates. *Journal of AOAC International*, *91*(4), 1002–1006. https://doi.org/10.1093/jaoac/91.4.1002.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of α-amanitin on the *Arabidopsis* seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiology*, *134*(4), 1598-1613. https://doi.org/10.1104/pp.103.036293.

Ramsay, L., Koh, C. S., Kagale, S., Gao, D., Kaur, S., Haile, T. and Bett, K. E. (2021). Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies of wild and cultivated lentil species. *BioRxiv*. 2021 Jul 24Retrieved from https://knowpulse. usask. ca/genome-assembly. *Lcu. 2RBY*. https://doi.org/10.1101/2021.07.23.453237.

Rosental, L., Nonogaki, H., and Fait, A. (2014). Activation and regulation of primary metabolism during seed germination. *Seed Science Research*, *24*(1), 1-15.

Rospert, S., Dubaquie, Y., and Gautschi, M. (2002). Nascent-polypeptide-associated complex. *Cellular and Molecular Life Sciences CMLS*, *59*(10), 1632-1639. https://doi.org/10.1007/PL00012490.

Rozan, P., Kuo, Y. H., & Lambein, F. (2000). Free amino acids present in commercially available seedlings sold for human consumption. A potential hazard for consumers. Journal of Agricultural and Food Chemistry, 48(3), 716-723. https://doi.org/10.1021/jf990729v.

Rozan, P., Kuo, Y.-H. H., and Lambein, F. (2001). Amino acids in seeds and seedlings of the genus *Lens*. *Phytochemistry*, *58*(2), 281–289. https://doi.org/10.1016/S0031-9422(01)00200-X.

Sá, A. G. A., Moreno, Y. M. F., and Carciofi, B. A. M. (2019). Food processing for the improvement of plant proteins digestibility. *Critical Reviews in Food Science and Nutrition*, *0*(0), 1–20. https://doi.org/10.1080/10408398.2019.1688249.

Saeed, S., and Shaukat, S. S. (2000). Effect of seed size on germination, emergence, growth and seedling survival of *Senna occidentalis* Link. *Pakistan Journal of Biological Sciences*, *3*(2), 292-295. https://doi.org/10.3923/pjbs.2000.292.295.

Schaafsma, G. (2012). Advantages and limitations of the protein digestibility-corrected amino acid score (PDCAAS) as a method for evaluating protein quality in human diets. *British Journal of Nutrition*, *108*(S2), S333–S336. https://doi.org/10.1017/S0007114512002541.

Scippa, G. S., Rocco, M., Ialicicco, M., Trupiano, D., Viscosi, V., Di Michele, M., & Scaloni, A. (2010). The proteome of lentil (*Lens culinaris* Medik.) seeds: discriminating between landraces. Electrophoresis, *31*(3), 497-506. https://doi.org/10.1002/elps.200900459.

Shahwar, D., Bhat, T. M., Ansari, M. Y. K., Chaudhary, S., and Aslam, R. (2017). Health functional compounds of lentil (*Lens culinaris* Medik): A review. *International Journal of Food Properties*. 20, S1–S15. doi:10.1080/10942912.2017.1287192.

Shewry, P. R., Napier, J. A., and Tatham, A. S. (1995). Seed storage proteins: structures and

biosynthesis. *The Plant Cell*, *7*(7), 945–956. https://doi.org/10.1105/tpc.7.7.945.

Shi, C., Zhang, Y., Lu, Z., and Wang, Y. (2017). Solid-state fermentation of corn-soybean meal mixed feed with *Bacillus subtilis* and *Enterococcus faecium* for degrading antinutritional factors and enhancing nutritional value. *Journal of Animal Science and Biotechnology*, *8*(1), 50. https://doi.org/10.1186/s40104-017-0184-2.

Singh, A. K., Rehal, J., Kaur, A., and Jyot, G. (2015). Enhancement of attributes of cereals by germination and fermentation: A review. *Critical Reviews in Food Science and Nutrition*, *55*(11), 1575–1589. https://doi.org/10.1080/10408398.2012.706661.

Soriano, D., Orozco-Segovia, A., Márquez-Guzmán, J., Kitajima, K., Gamboa-de Buen, A., and Huante, P. (2011). Seed reserve composition in 19 tree species of a tropical deciduous forest in Mexico and its relationship to seed germination and seedling growth. *Annals of Botany*, *107*(6), 939-951. https://doi.org/10.1093/aob/mcr041.

Sosulski, F. W., and Holt, N. W. (2020). Amino acid composition, protein content and accurate nitrogen-to-protein conversion factor for sheepgrass (*Leymus chinensis*). *Botany*, *98*(3), 137–146. https://doi.org/10.1139/cjb-2019-0082.

Souza, M. L., and Fagundes, M. (2014). Seed size as key factor in germination and seedling development of *Copaifera langsdorffii* (Fabaceae). *American Journal of Plant Sciences*, 2014. https://doi.org/10.4236/ajps.2014.517270.

Sreenivasulu, N., Usadel, B., Winter, A., Radchuk, V., Scholz, U., Stein, N., & Wobus, U. (2008). Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant Physiology*, *146*(4), 1738-1758. https://doi.org/10.1104/pp.107.111781.

Suffo Kamela, A. L., Mouokeu, R. S., Ashish, R., Maffo Tazoho, G., Glory Moh, L., Pamo Tedonkeng, E., and Kuiate, J.-R. (2016). Influence of processing methods on proximate composition and dieting of two Amaranthus species from West Cameroon. *International Journal of Food Science*, *2016*, 1–8. https://doi.org/10.1155/2016/6707313.

Sulieman, M. A., Eltayeb, M. M., Babiker, E. E., Mustafa, A. I., El Tinay, A. H. (2008). Effect of sprouting on chemical composition and amino acid content of Sudanese lentil cultivars. *Journal of Applied Sciences*, *8*(12), 2337-2340. https://doi.org/10.3923/jas.2008.2337.2340.

Tahir, M., Lindeboom, N., Båga, M., Vandenberg, A., and Chibbar, R. (2011). Composition and correlation between major seed constituents in selected lentil (*Lens culinaris*. Medik) genotypes. *Canadian Journal of Plant Science*, *91*(5), 825-835. https://doi.org/10.4141/cjps2011-010.

Tegeder, M. (2014). Transporters involved in source to sink partitioning of amino acids and ureides: opportunities for crop improvement. *Journal of Experimental Botany*, *65*(7), 1865-1878. https://doi.org/10.1093/jxb/eru012.

Thavarajah, D., Vandenberg, A., George, G. N., & Pickering, I. J. (2007). Chemical form of selenium in naturally selenium-rich lentils (*Lens culinaris* L.) from Saskatchewan. *Journal*

of Agricultural and Food Chemistry, *55*(18), 7337-7341. https://doi.org/10.1021/jf070681i.

Toyomasu, T., Kawaide, H., Mitsuhashi, W., Inoue, Y., and Kamiya, Y. (1998). Phytochrome regulates gibberellin biosynthesis during germination of photoblastic lettuce seeds. *Plant Physiology*, *118*(4), 1517-1523. https://doi.org/10.1104/pp.118.4.1517.

Tuan, P. A., Sun, M., Nguyen, T. N., Park, S., and Ayele, B. T. (2019). Molecular mechanisms of seed germination. In *Sprouted Grains* (pp. 1-24). *AACC International Press*. https://doi.org/10.1016/B978-0-12-811525-1.00001-4.

Tuśnio, A., Taciak, M., Barszcz, M., Święch, E., Bachanek, I., and Skomiał, J. (2017). Effect of replacing soybean meal by raw or extruded pea seeds on growth performance and selected physiological parameters of the ileum and distal colon of pigs. *PloS One*, *12*(1), e0169467. https://doi.org/10.1371/journal.pone.0169467.

Urbano, G., Porres, J. M., Frías, J., and Vidal-Valverde, C. (2007). Nutritional value. In *Lentil* (pp. 47-93). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6313-8_5.

Van Oss, H., Aron, Y., & Ladizinsky, G. (1997). Chloroplast DNA variation and evolution in the genus Lens Mill. *Theoretical and Applied Genetics*, *94*(3–4), 452–457. https://doi.org/10.1007/s001220050436.

Vandenberg, A., and Slinkard, A. E. (1990). Genetics of seed coat color and pattern in lentil. *Journal of Heredity*, *81*(6), 484–488. https://doi.org/10.1093/oxfordjournals.jhered.a111030.

Vandenberg, A., Banniza, S., Warkentin, T. D., Ife, S., Barlow, B., McHale, S, and Dueck, S. (2006). CDC Redberry lentil. *Canadian Journal of Plant Science*, *86*(2), 497-498. https://doi.org/10.4141/P05-071.

Vázquez-Yanes, C., and Orozco-Segovia, A. (1993). Patterns of seed longevity and germination in the tropical rainforest. *Annual Review of Ecology and Systematics*, *24*(1), 69-87. https://doi.org/10.1146/annurev.es.24.110193.000441.

Wong, M. M. L., Gujaria-Verma, N., Ramsay, L., Yuan, H. Y., Caron, C., Diapari, M., Vandenberg, A., and Bett, K. E. (2015). Classification and characterization of species within the genus *Lens* using genotyping-by-sequencing (GBS). *PloS One*, *10*(3), e0122025. https://doi.org/10.1371/journal.pone.0122025.

Xu, J., Li, W., Zhang, C., Liu, W., and Du, G. (2014). Variation in seed germination of 134 common species on the eastern Tibetan Plateau: phylogenetic, life history and environmental correlates. *Plos One*, *9*(6), e98601. https://doi.org/10.1371/journal.pone.0098601.

Yamasaki, Y. (2003). β-Amylase in germinating millet seeds. *Phytochemistry*, 64(5), 935-939. https://doi.org/10.1016/S0031-9422(03)00430-8.

Yu, Y., Guo, G., Lv, D., Hu, Y., Li, J., Li, X., & Yan, Y. (2014). Transcriptome analysis during seed germination of elite Chinese bread wheat cultivar Jimai 20. *BMC Plant Biology*, *14*(1), 1-19. https://doi.org/10.1186/1471-2229-14-20.

Zambrowicz, A., Timmer, M., Polanowski, A., Lubec, G., and Trziszka, T. (2013). Manufacturing of peptides exhibiting biological activity. *Amino Acids*, *44*(2), 315–320. https://doi.org/10.1007/s00726-012-1379-7.

Zhang, H., De Silva, D., Dissanayaka, D., Warkentin, T. D., and Vandenberg, A. (2021). Validated B vitamin quantification from lentils by selected reaction monitoring mass spectrometry. *Food Chemistry*, *359*, 129810. https://doi.org/10.1016/j.foodchem.2021.129810.

Žilić, S., Basić, Z., Hadži-Tašković Šukalović, V., Maksimović, V., Janković, M., and Filipović, M. (2014). Can the sprouting process applied to wheat improve the contents of vitamins and phenolic compounds and antioxidant capacity of the flour? *International Journal of Food Science and Technology*, *49*(4), 1040-1047. https://doi.org/10.1111/ijfs.12397.

Appendices

Appendix A Percentage of water uptake in six lentil species during the imbibition period. The uppercase letters indicate the varitation among the genotypes and the lowercase letters indicate the variation among the time periods measured

Genotype	DS-1h	1-2h	2-3h	3-4h	4-5h
CDC Redberry	$68.64\pm1.12^{A,a}$	$12.05\pm1.11^{A,b}$	$8.40\pm1.08^{A,c}$	$4.16\pm0.62^{A,d}$	$1.64\pm0.13^{A,e}$
BGE 016880	$62.90\pm1.56^{AB,a}$	$12.71\pm1.17^{AB,b}$	$4.32\pm0.63^{AB,c}$	$13.02\pm1.90^{AB,d}$	$0.79\pm0.08^{AB,e}$
IG 72805	$65.28\pm1.19^{BC,a}$	$14.57\pm0.75^{BC,b}$	$2.40\pm0.01^{BC,c}$	$3.20\pm0.84^{BC,d}$	$1.86\pm0.46^{BC,e}$
IG 110813	$47.58\pm3.76^{CD,a}$	$21.42\pm1.48^{CD,b}$	$4.77\pm0.97^{CD,c}$	$4.34\pm1.05^{CD,d}$	$2.89\pm0.66^{CD,e}$
IG 72623	$52.88\pm3.48^{CD,a}$	$9.59\pm5.96^{CD,b}$	$17.28\pm3.53^{CD,c}$	$2.71\pm0.23^{CD,d}$	$1.40\pm0.12^{CD,e}$
IG 72815	$70.04\pm1.30^{D,a}$	$3.31\pm0.77^{D,b}$	$0.97\pm0.90^{D,c}$	$0.73\pm1.48^{D,d}$	$1.71\pm0.79^{D,e}$

Appendix B The average diameter of dry seeds and multiple correlation test (MCT) of six lentil species. The letter indicate the variation among seed diameter in mm unit of measurement

Genotype	Species	Dry seed diameter (mm)
CDC Redberry	*L. culinaris*	4.81 ± 0.025^{A}
BGE 016880	*L. orientalis*	3.28 ± 0.013^{C}
IG 72805	*L. tomentosus*	3.24 ± 0.025^{BC}
IG 110813	*L. lamottei*	3.36 ± 0.042^{B}
IG 72623	*L. odemensis*	3.08 ± 0.014^{BC}
IG 72815	*L. ervoides*	2.58 ± 0.018^{D}
MCT effect	**F Value**	**Pr > F**
Genotype	84.87	<.0001

Appendix C Protease activity of eight lentil genotypes from seven species of lentil. Different lowercase letters within column indicate statistical differences among different phase's endpoints ($p \leq 0.05$ post hoc Duncan test). Different uppercase letter denotes statistical differences among samples at the same phase ($p \leq 0.05$, post hoc Duncan test). Values of one biological replicate (n = 1)

Genotypes	Control	Phase II	Phase III	Phase IV
IG 110813	0.72±0.06 a, A	1.02±0.08 b, B	1.36±0.13 c, B	2.68±0.10 d, B
IG 72815	0.91±0.09 a, B	1.38±0.13 b, D	1.99±0.10 c, C	1.8±0.18 c, A
IG 72805	1.41±0.09 a, C	1.27±0.10 a, CD	4.35±0.13 c, F	3.47±0.25 b, C
BGE 016880	0.92±0.09 a, B	1.08±0.09 b, BC	1.01±0.01 ab, A	2.17±0.13 c, A
IG 72541	1.01±0.02 b, B	0.55±0.04 a, A	0.83±0.07 b, A	2.01±0.32 c, A
IG 72623	0.89±0.02 b, AB	0.72±0.03 a, A	1.78±0.10 c, C	2.18±0.09 d, A
CDC Redberry	2.33±0.18 a, D	2.23±0.17 a, E	3.18±0.10 b, D	4.55±0.36 c, D
Eston	2.84±0.16 a, E	3.02±0.22 a, F	3.68±0.24 b, E	4.37±0.32 c, D

Appendix D Total free amino acids, total free essential amino acids and total free non-essential amino acids of eight lentil genotypes from seven species of lentil. Different lowercase letters indicate statistical differences among different phase's endpoints ($p \leq 0.05$ post hoc Duncan test). Different uppercase letters denotes statistical differences among samples at the same phase ($p \leq 0.05$, post hoc Duncan test). Values of one biological replicate (n = 1)

Genotype	Species	Phase	EFAA (mg/g flour)	NEFAA (mg/g flour)	TFAA (mg/g flour)
CDC Redberry	*L. culinaris*	Control	1.12 ± 0.18 $^{a, C}$	8.43 ± 0.14 $^{b, CD}$	9.55 ± 0.03 $^{b, C}$
		Phase II	2.03 ± 0.20 $^{b, C}$	4.06 ± 0.30 $^{a, A}$	5.98 ± 0.27 $^{a, A}$
		Phase III	3.06 ± 0.07 $^{c, D}$	7.96 ± 0.51 $^{b, A}$	11.03 ± 0.54 $^{c, A}$
		Phase IV	4.63 ± 0.25 $^{d, A}$	15.83 ± 0.68 $^{c, A}$	20.46 ± 0.93 $^{d, A}$
Eston	*L. culinaris*	Control	1.25 ± 0.11 $^{a, C}$	8.87 ± 0.50 $^{a, D}$	10.12 ± 0.60 $^{a, CD}$
		Phase II	1.95 ± 0.08 $^{b, C}$	8.37 ± 0.40 $^{a, C}$	10.33 ± 0.46 $^{a, BC}$
		Phase III	2.95 ± 0.14 $^{c, D}$	12.81 ± 1.27 $^{b, CD}$	13.93 ± 1.09 $^{b, ABC}$
		Phase IV	5.11 ± 0.16 $^{d, B}$	22.91 ± 0.61 $^{c, C}$	28.02 ± 0.76 $^{c, C}$
BGE 016880	*L. orientalis*	Control	0.45 ± 0.01 $^{a, A}$	9.86 ± 0.33 $^{a, E}$	10.31 ± 0.33 $^{a, CD}$
		Phase II	0.72 ± 0.01 $^{a, A}$	9.71 ± 0.60 $^{a, D}$	10.37 ± 0.61 $^{a, BC}$
		Phase III	1.78 ± 0.20 $^{b, A}$	12.92 ± 0.84 $^{b, CD}$	14.43 ± 1.30 $^{b, BC}$
		Phase IV	6.50 ± 0.39 $^{c, C}$	25.66 ± 2.43 $^{c, D}$	31.55 ± 3.50 $^{c, D}$
IG 72805	*L. tomentosus*	Control	0.82 ± 0.02 $^{a, B}$	10.11 ± 0.11 $^{a, E}$	10.93 ± 0.11 $^{a, D}$
		Phase II	0.95 ± 0.09 $^{a, A}$	10.39 ± 0.75 $^{a, D}$	11.49 ± 1.03 $^{a, C}$
		Phase III	2.20 ± 0.03 $^{b, BC}$	14.71 ± 0.61 $^{b, D}$	16.57 ± 1.13 $^{b, C}$
		Phase IV	4.50 ± 0.20 $^{c, A}$	22.87 ± 1.41 $^{c, C}$	27.37 ± 1.60 $^{c, BC}$
IG 110813	*L. lamottei*	Control	0.81 ± 0.05 $^{a, B}$	13.71 ± 0.86 $^{a, F}$	14.52 ± 0.91 $^{a, E}$
		Phase II	1.44 ± 0.02 $^{b, B}$	13.66 ± 0.04 $^{a, E}$	14.12 ± 1.68 $^{a, D}$
		Phase III	1.96 ± 0.16 $^{c, AB}$	13.85 ± 1.63 $^{a, D}$	15.80 ± 1.86 $^{a, C}$
		Phase IV	6.90 ± 0.19 $^{d, CD}$	30.81 ± 0.88 $^{b, E}$	37.81 ± 0.93 $^{b, E}$
IG 72623	*L. odemensis*	Control	0.87 ± 0.02 $^{a, B}$	7.79 ± 0.23 $^{a, BC}$	8.66 ± 0.26 $^{a, B}$
		Phase II	1.54 ± 0.04 $^{b, B}$	8.68 ± 0.31 $^{a, C}$	10.22 ± 0.32 $^{a, BC}$
		Phase III	2.36 ± 0.01 $^{c, BC}$	11.24 ± 1.90 $^{b, BC}$	13.60 ± 1.91 $^{b, ABC}$
		Phase IV	7.16 ± 0.22 $^{d, D}$	28.73 ± 0.79 $^{c, E}$	35.89 ± 1.01 $^{c, E}$
IG 72815	*L. ervoides*	Control	0.58 ± 0.03 $^{a, A}$	7.14 ± 0.12 $^{a, B}$	7.72 ± 0.15 $^{a, A}$
		Phase II	1.64 ± 0.09 $^{b, BC}$	7.15 ± 0.45 $^{a, B}$	8.80 ± 0.54 $^{a, B}$
		Phase III	2.16 ± 0.03 $^{c, ABC}$	12.33 ± 1.03 $^{b, CD}$	14.50 ± 1.07 $^{b, BC}$
		Phase IV	4.42 ± 0.09 $^{d, A}$	24.16 ± 0.40 $^{c, CD}$	28.58 ± 0.46 $^{c, CD}$
IG 72541	*L. nigricans*	Control	1.66 ± 0.19 $^{a, D}$	6.25 ± 0.46 $^{a, A}$	7.91 ± 0.28 $^{a, AB}$
		Phase II	1.78 ± 0.56 $^{a, BC}$	4.57 ± 0.02 $^{a, A}$	6.36 ± 0.54 $^{a, A}$
		Phase III	2.45 ± 0.56 $^{a, C}$	9.84 ± 0.97 $^{b, AB}$	12.28 ± 1.53 $^{b, AB}$
		Phase IV	4.61 ± ___ $^{b, A}$	19.62 ± ___ $^{c, B}$	24.23 ± ___ $^{c, B}$

Appendix E Mutiple correlation test result from protein concentrations obtained from the BCA assay

Effect	Num DF	Den DF	F Value	Pr > F
Genotype	7	64	70.21	<.0001
Phase	3	64	16.62	<.0001
Genotype*Phase	21	64	5.36	<.0001

Appendix F Number of proteins found in each category and subcategory in phase 0, phase II, phase III, and phase IV of lentil germination

Category	Sub-category	Phase 0	Phase II	Phase III	Phase IV
Photosynthesis	Photophosphorylation	19	18	14	15
	Calvin cycle	12	10	10	10
	Photorespiration	3	4	3	4
	CAM/C4 photosynthesis	2	2	2	2
	Total	36	33	29	30
Cellular respiration	Glycolysis (cytosolic)	5	5	5	6
	Glycolysis (methylglyoxal degradation)	1	1	1	1
	Pyruvate oxidation (mitochondrial)	3	3	3	3
	Tricarboxylic acid cycle	8	7	6	7
	Oxidative phosphorylation (NADH dehydrogenase complex)	9	9	8	10
	Oxidative phosphorylation (cytochrome c complex)	3	3	4	6
	Oxidative phosphorylation (ATP synthase complex)	6	6	5	6
	Total	35	34	32	39
Carbohydrate metabolism	Sucrose metabolism (transport)	1	1	1	1
	Sucrose metabolism (biosynthesis-cytosolic)	5	4	4	4
	Sucrose metabolism (sucrose-phosphate)	2	2	2	2
	Sucrose metabolism (degradation-invertase activities, sucrose synthase, hexokinase, fructose kinase)	3	3	3	3
	Starch metabolism (biosynthesis)	5	7	6	6
	Starch metabolism (degradation-phosphorylation)	2	1	1	1
	Starch metabolism (degradation-hydrolysis and phosphorolysis)	3	2	3	2

	Starch metabolism (degradation-maltose metabolism)	1	1	1	1
	Oligosaccharide metabolism	2	2	2	2
	Galactose metabolism	1	0	0	0
	Oxidative pentose phosphate pathway	3	2	2	3
	Gluconeogenesis	1	0	0	2
	Carbohydrate metabolism (non-oxidative)	1	1	0	0
	Fermentation (acetic acid biosynthesis)	2	2	2	1
	Plastidial glycolysis	4	4	4	4
	Nucleotide sugar biosynthesis	6	6	5	8
	Total	42	40	37	41
Amino acid metabolism	Glutamate group amino acid biosynthesis	5	5	3	4
	Aspartate group amino acid biosynthesis	5	7	6	5
	Pyruvate group amino acid biosynthesis	3	4	1	1
	Serine group amino acid biosynthesis	3	3	3	3
	Shikimate group amino acid biosynthesis	0	0.5	0	1.5
	Amino acid degradation	6	6	8	8
	Total	23	26	21	23
Lipid metabolism	Fatty acid metabolism (citrate shuttle)	4	4	4	5
	Fatty acid metabolism (acetyl-CoA)	1	2	2	2
	Fatty acid metabolism (FAE system)	1	0	0	1
	Fatty acid metabolism (ptfas system)	1	2	1	1
	Fatty acid metabolism (mtfas system)	1	1	0	0
	Fatty acid metabolism (desaturation)	1	1	1	1
	Fatty acid metabolism (degradation-alpha-beta-oxidation)	2	2	2	3
	Fatty acid metabolism (degradation-glyoxylate cycle)	3	2	4	4
	Glycerolipid metabolism (biosynthesis)	1	2	0	1

	Glycerolipid metabolism (degradation)	2	0	0	0
	Galactolipid and sulfolipid biosynthesis	1	0	0	0
	Phytosterol metabolism	1	0	0	1
	Lipid trafficking	1	1	1	1
	Lipid droplet-associated activities	2	2	2	2
	Sphingolipid metabolism (biosynthesis)	0	0	0	0
	Total	20	20	18	22
Nucleotide metabolism	Ribonucleotide (RN) anabolism	1	1	1	1
	Purine metabolism	4	4	3	4
	Pyrimidines (De novo biosynthesis)	2	2	1	2
	Pyrimidine catabolism	1	1	1	2
	Deoxynucleotide metabolism (biosynthesis)	1	0	0	0
	Deoxynucleotide metabolism (salvage pathway)	1	1	1	1
	Deoxynucleotides catabolism	0	0	0	0
	Total	10	9	8	10
Coenzyme metabolism	SAM cycle	2	2	2	2
	Tetrahydrofolate metabolism (THF)	2	1	1	2
	Pyridoxalphosphate biosynthesis	1	1	1	1
	Prenylquinone biosynthesis	1	1	1	1
	NAD/NADP biosynthesis	1	1	1	1
	Iron-sulfur cluster assembly machinery	3	2	2	3
	Tetrapyrrole biosynthesis	2	0	0	0
	Chlorophyll metabolism (biosynthesis)	0	0	0	0
	Total	11	9	9	10
Polyamine metabolism	Spermidine biosynthesis	1	1	1	1
	Total	1	1	1	1
Secondary metabolism	Terpenoids (mevalonate pathway)	2	2	1	2
	Terpenoids (terpene biosynthesis)	1	2	2	2
	Phenolics-p-coumaroyl-CoA biosynthesis	1	1	1	1
	Phenolics-flavonoid biosynthesis	3	3	2	3
	Phenolics- lignan biosynthesis	0	0	0	0
	Betaine biosynthesis	1	1	1	1
	Terpenoids(MEP pathway)	0	0	0	1

	Terpenoids (isoprenyl diphosphate biosynthesis)	0	0	0	0
	Coenzyme A biosynthesis	0	0	0	0
	Total	9	8	8	10
Redox homeostasis	Enzymatic reactive oxygen scavenging	4	4	4	4
	Glutathione-based redox regulation	1	2	2	2
	Thiol-based redox regulation	5	6	4	5
	Ascorbate-based redox regulation	5	5	5	5
	Tocopherol biosynthesis	1	2	2	2
	Singlet oxygen-induced signalling	0	0	0	0
	Reactive carbonyl scavenging (plastidial)	0	0	0	0
	Total	17	18	16	19
'Phytohormone action	Abscisic acid	1	0	1	0
	Jasmonic acid	0	1	1	1
	Auxin	1	1	1	0
	Cytokinin	0	0	1	0
	Gibberellin	0	0	0	1
	Signalling peptides	0	0	0	0
	Ethylene biosynthesis	1	0	0	0
	Strigolactone (perception and signal transduction)	0	0	0	0
	Total	3	3	4	3
Chromatin organisation	Chromatin structure (DNA wrapping)	4	4	4	4
	Chromatin structure (DNA bridging)	1	1	1	1
	Histone chaperone activities	1	2	2	1.5
	Nucleosome remodeling (chromatin)	1	1	1	1.5
	DNA methylation	0	0	0	1.5
	DNA homeostasis	1	0	0	0
	Post-translational histone modification (methylation)	0	0	0	0
	Organelle machinery	0	0	0	0
	Post-translational histone modification (histone methylation)	0	0	0	0
	Total	9	9	10	6
Cell division	DNA replication	2	1	1	2
	Cell cycle organisation	1	1	1	0
	Cell cycle control	0	0	0	0
	Meiotic recombination	1	2	2	1

	Cytokinesis	2	2	3	4
	Plastid division	1	0	0	0
	Mitochondrion/peroxisome binary fission	1	1	0	0
	Total	8	7	7	7
DNA damage response	Nonhomologous end-joining (NHEJ) repair	1	1	1	1
	Nucleotide excision repair (NER)	2	1	1	1
	DNA damage sensing and signalling (ATR pathway)	0	1	1	0
	Total	3	3	3	2
RNA biosynthesis	RNA polymerase II-dependent transcription	2	1	3	1
	Transcriptional regulation	3	5	3	3
	Organelle machinery (RNA polymerase activities)	1	2	0	1
	Total	6	8	6	5
RNA processing	Pre-RNA splicing	2	2	1	2
	RNA modification	1	0	2	1
	RNA export	1	1	1	1
	RNA homeostasis (mRNA stress granule formation, mRNA deadenylation-dependent decay)	2	3	3	1
	mRNA silencing	2	2	2	2
	Organelle machinery	3	3	2	3
	Total	12	10	11	10
'Protein biosynthesis	Ribosome biogenesis	70	69	63	70
	Aminoacyl-tRNA synthetase activities	7	6	7	6
	Translation initiation	11	10	10	13
	Translation elongation	9	5	5	6
	Translation termination	0	0	0	1
	Organelle machinery	3	3	2	1
	Total	100	93	87	97
Protein modification	Glycosylation	2	3	3	2
	Acetylation	0	1	0	0
	Phosphorylation	5	7	5	6
	ADP-ribosylation	1	1	1	1
	S-nitrosylation	2	2	1	2
	S-glutathionylation	1	1	1	1
	Hydroxylation	1	0	0	0
	Targeting peptide maturation	3	2	1	2
	Cysteine disulfide formation	3	3	3	3
	Protein folding	2	2	2	2

	Total	20	22	16	19
Protein homeostasis	Er quality control (erqc)	4	3	3	4
	NAC chaperone heterodimer	1	1	1	1
	Mitochondrial Hsp70 chaperone system	2	2	1	2
	Plastidial Hsp70 chaperone system	2	1	1	1
	Cytosolic Hsp70 chaperone system	3	2	2	2
	Hsp90 chaperone system	2	1	1	1
	Hsp60 chaperone system	2	1	1	2
	Smallhsp holdase chaperone activities	4	4	4	4
	Ubiquitin-proteasome system (conjugation and deconjugation)	7	5	3	7
	Ubiquitin-proteasome system	3	2	1	2
	26S proteasome (20S proteosome)	12	12	11	12
	26S proteasome (19S proteosome)	17	17	15	16
	Proteolysis (cysteine-type peptidase activities)	1	1	1	1
	Proteolysis (serine-type peptidase activities)	6	5	4	4
	Proteolysis (aspartic-type peptidase activities)	1	1	1	1
	Proteolysis (threonine-type peptidase activities)	0	0	0	0
	Proteolysis (metallopeptidase activities)	5	4	4	3
	Proteolysis (protease inhibitor activities)	5	4	4	4
	Seed storage protein activities	3	2	2	1
	Vegetative storage protein activities	0	0	0	0
	Total	79	67	62	68
Cytoskeleton organisation	Microtubular network (alpha-beta-Tubulin)	2	2	2	2
	Microtubular network (Kinesin)	5	4	2	2
	Microtubular network (Dynein microtubule-based motor protein)	1	0	0	1
	Microfilament network	5	2	3	3
	Actin and tubulin folding	7	8	7	7
	Cytoskeleton-plasma membrane	2	1	0	1
	Plastid movement	1	0	1	0
	Total	22	18	15	17
Cell wall organisation	Cellulose (cellulose synthase complex)	2	2	1	1

	Hemicellulose (modification and degradation)	2	1	1	1
	Pectin (homogalacturonan modification and degradation)	1	0	1	1
	Pectin (rhamnogalacturonan biosynthesis)	1	1	0	0
	Pectin (rhamnogalacturonan-modification and degradation)	2	1	0	0
	Pectin (modification and degradation)	1	0	0	0
	Cell wall proteins	0	0	0	0
	Lignin (monolignol biosynthesis)	2	2	2	2
	Cuticular lipid formation (cutin and suberin)	1	0	0	1
	Callose synthesis	1	0	1	1
	Hemicellulose biosynthesis (xyloglucan)	0	0	0	0
	Cell wall proteins (hydroxyproline-rich glycoprotein activities)	0	0	0	0
	Cell wall proteins (expansin activities)	0	0	0	0
	Total	13	7	8	8
Vesicle trafficking	Anterograde trafficking	4	0	3	4
	Retrograde trafficking	8	7	6	7
	Endocytic trafficking	1	2	1	1
	Exocytic trafficking	2	2	2	3
	Multi-pathway trafficking regulation	14	12	12	10
	Total	29	26	24	25
Protein translocation	Chloroplast outer envelope (TOC)	2	2	1	1
	Chloroplast inner envelope (TIC)	4	2	2	3
	Inner mitochondrion membrane (TIM)	3	1	1	1
	Nucleocytoplasmic transport	6	4	4	4
	Nucleus (nuclear pore complex NPC)	0	0	1	1
	Endoplasmic reticulum	0	0	0	1
	*Translocation system accessory	0	0	0	0
	Total	14	10	10	10
Solute transport	Primary active transport (V-type)	4	4	4	5
	Primary active transport (P-type)	2	1	1	1
	Primary active transport (ABC superfamily)	1	1	2	2

	Primary active transport (VHP ppase family)	2	1	1	1
	Carrier-mediated transport (DMT superfamily)	2	2	1	1
	Carrier-mediated transport (MFS superfamily)	2	1	0	0
	Carrier-mediated transport (APC superfamily)	2	1	1	2
	Carrier-mediated transport (IT superfamily)	1	0	0	0
	CPA superfamily	1	0	0	0
	Carrier-mediated transport (VIT family)	1	0	0	0
	Carrier-mediated transport (MTCC)	1	1	1	1
	Carrier-mediated transport (potassium/sodium)	3	0	0	0
	Maltose transporter	0	0	0	0
	Transport channels (MIP family)	2	1	2	1
	Transport channels (calcium-permeable)	3	0	0	0
	Transport channels (voltage gated)	1	0	0	0
	Prions	1	1	1	1
	Transport channels (mechanosensitive ion channel)	0	0	0	0
	Transport channels (cora family)	0	0	0	0
	Total	29	15	14	17
Nutrient uptake	Nitrogen assimilation (systemic nitrogen signalling)	0	0	0	0
	Nitrogen assimilation (ammonium assimilation)	2	1	1	1
	Nitrogen assimilation (glutamate deamination)	1	1	1	1
	Nitrogen assimilation (aspartate aminotransferase)	1	1	1	1
	Sulfur assimilation (sulfate, sulfite)	2	1	1	1
	Phosphorus assimilation	1	0	0	0
	Transition metal homeostasis	3	2	1	1
	Total	9	7	5	5
External stimuli response	External stimuli response	3	3	4	3
	Total	3	3	4	3
Multi-process regulation	Programmed Cell Death (PCD) system	2	0	0	0

	14-3-3 regulatory system	3	1	1	1
	ROP-gtpase regulatory system	3	0	0	1
	Pyrophosphate homeostasis	3	1	1	1
	Phosphatidylinositol and inositol phosphate system	3	2	1	2
	Calcium-dependent signalling	3	2	3	3
	RAM (Regulation and Morphogenesis)	0	0	0	0
	Snrk1-kinase regulatory system	0	0	0	0
	Total	16	7	7	9
Plant reproduction	Flowering modulation	1	0	1	1
	Flowering (autonomous floral-promotion pathway)	0	1	0	0
	Total	2	1	1	1
Clade-specific metabolism	Glucosinolate biosynthesis	2	2	1	1
	Total	2	2	1	1
Enzyme classification	EC_1.1 oxidoreductase acting on CH-OH group of donor	1	1	1	1
	EC_1.2 oxidoreductase acting on aldehyde or oxo group of donor	1	1	1	1
	EC_1.3 oxidoreductase acting on CH-CH group of donor	1	1	1	1
	EC_1.10 oxidoreductase acting on diphenol or related substance as donor'	0	0	1	0
	EC_1.13 oxidoreductase acting on single donor with incorporation of molecular oxygen (oxygenase)	1	1	1	1
	EC_1.14 oxidoreductase acting on paired donor with incorporation or reduction of molecular oxygen	1	1	0	1
	EC_1.17 oxidoreductase acting on CH or CH2 group	1	1	1	1
	EC_1.18 oxidoreductase acting on iron-sulfur protein as donor	0	0	0	0
	EC_2.3 acyltransferase	1	1	0	0
	EC_2.4 glycosyltransferase	1	1	1	1
	EC_2.5 transferase transferring alkyl or aryl group, other than methyl group	0	0	1	1
	EC_2.6 transferase transferring nitrogenous group	0	0	0	0
	EC_2.7 transferase transferring phosphorus-containing group	1	1	1	1

	EC_2.8 transferase transferring sulfur-containing group	0	0	0	0
	EC_3.2 glycosylase	1	1	1	1
	EC_3.3 hydrolase acting on ether bond	0	0	0	0
	EC_3.4 hydrolase acting on peptide bond (peptidase)'	0	0	1	1
	EC_3.6 hydrolase acting on acid anhydride	1	0	1	1
	EC_4.2 carbon-oxygen lyase	1	1	1	1
	EC_4.3 carbon-nitrogen lyase	0	0	0	0
	EC_5.3 intramolecular oxidoreductase	1	1	1	1
	EC_5.4 intramolecular transferase	1	1	1	1
	EC_5.5 intramolecular lyase	0	0	0	0
	Total	14	13	14	15

Appendix G List of common proteins identified through MapMan with their category and subcategories in phases 0, II, III, and IV of lentil germination

Category	Sub-category	Protein
Amino acid metabolism	Amino acid degradation (arginine)	Arginase
	Amino acid degradation (GABA)	Gamma-aminobutyric acid pyruvate transaminase
		Succinate formation.NAD-dependent succinic semialdehyde dehydrogenase
	Amino acid degradation (lysine)	Piperideine-6-carboxylate dehydrogenase
	Aspartate group biosynthesis (methionine)	Methyl-tetrahydrofolate-dependent methionine synthase
	Aspartate group biosynthesis (asparagine)	Glutamine-dependent asparagine synthetase
	Glutamate group biosynthesis (Arginine)	Argininosuccinate lyase
		Argininosuccinate synthetase
	Glutamate group biosynthesis (ornithine)	N2-acetylornithine deacetylase
	Pyruvate group amino acid biosynthesis	Ketol-acid reductoisomerase
	Serine group amino acid biosynthesis	Phosphoglycerate dehydrogenase
Carbohydrate metabolism	Fermentation (acetic acid biosynthesis)	Alcohol dehydrogenase (ADH)
		Aldehyde dehydrogenase (ALDH2B)
	Nucleotide sugar biosynthesis	UDP-L-rhamnose synthase (RHM)
		UDP-D-glucose 4-epimerase
		UDP-D-xylose biosynthesis.UDP-D-apiose / UPD-D-xylose synthetase
		UDP-L-arabinose biosynthesis.UDP-L-arabinose mutase
	Oligosaccharide metabolism	Alpha-galactosidase (AGAL)
		Stachyose synthase
	Oxidative pentose phosphate pathway	Non-oxidative phase.transketolase
		Oxidative phase.6-phosphogluconate dehydrogenase
	Plastidial glycolysis	Glyceraldehyde 3-phosphate dehydrogenase
	Starch biosynthesis	ADP-glucose pyrophosphorylase complex.large subunit APL
		ADP-glucose pyrophosphorylase complex.small subunit APS
	Starch biosynthesis	Plastidial phosphoglucomutase

		Starch synthase activities.granule-bound starch (amylose) synthase
		Starch synthase activities.starch synthase (SSII)
	Starch degradation	Carbohydrate metabolism.starch metabolism.degradation.hydrolysis and phosphorolysis.amylase activities.beta amylase
		Hydrolysis and phosphorolysis.plastidial alpha-glucan phosphorylase (PHS1)
		Maltose metabolism.cytosolic alpha-glucan phosphorylase
	Sucrose biosynthesis	Cytosolic phosphoglucomutase
		Cytosolic UDP-glucose pyrophosphorylase
		Sucrose-phosphate phosphatase
	Sucrose degradation	Fructose kinase
		Sucrose synthase
Cell division	DNA elongation	DNA-tracking platform.sliding clamp protein (PCNA)
Cell wall organisation	Xyloglucan degradation	1,6-alpha-xylosidase
		Cell wall organisation.lignin.monolignol biosynthesis.cinnamyl-alcohol dehydrogenase
Cellular respiration	Cytosolic glycolysis	Fructose-bisphosphate aldolase (FBA)
		Glyceraldehyde 3-phosphate dehydrogenase activities.NAD-dependent glyceraldehyde 3-phosphate dehydrogenase
		Phosphoglycerate kinase
	Methylglyoxal degradation	Lactoyl-glutathione lyase (GLX1)
	Oxidative phosphorylation	ATP synthase complex.peripheral MF1 subcomplex.subunit alpha
		Cytochrome c reductase complex.peptidase component MPP heterodimer.subunit beta
		Cytochrome c.cytochrome c protein
		NADH dehydrogenase complex.non-core modules.alpha subcomplex.component NDUFA6/B14
		NADH dehydrogenase complex.module N (electron input).component NQO2/24kda
		NADH dehydrogenase complex.module Q (electron output).component NQO4/ND7

	Pyruvate oxidation	Mitochondrial pyruvate dehydrogenase complex.dihydrolipoamide dehydrogenase component E3
		Mitochondrial pyruvate dehydrogenase complex.component E1 heterodimer.subunit beta
		2-oxoglutarate dehydrogenase complex.component E2
	Tricarboxylic acid cycle	Aconitase (ACO)
		Succinyl-CoA ligase heterodimer.subunit alpha and beta
Chromatin organisation	Chromatin structure	DNA bridging.linker histone (H1)
		DNA wrapping.histone (H2B)
		DNA wrapping.histone (H2A)
		DNA wrapping.histone (H4)
	Chromatin remodeling activities	SSO1653-like group.chromatin remodeling factor (ERCC6)
Coenzyme metabolism	NAD/NADP biosynthesis	De-novo pathway.quinolinate phosphoribosyl transferase
	Pyridoxalphosphate biosynthesis	Pyridoxal 5-phosphate synthase complex.synthase component
	S-adenosyl methionine (SAM) cycle	S-adenosyl homocysteine hydrolase (SAHH)
	Actin and tubulin folding	CCT chaperonin folding complex.subunit zeta (CCT6)
		CCT chaperonin folding complex.subunit beta (CCT2)
		CCT chaperonin folding complex.subunit alpha (CCT1)
		CCT chaperonin folding complex.subunit theta (CCT8)
		CCT chaperonin folding complex.subunit delta (CCT4)
		CCT chaperonin folding complex.subunit eta (CCT7)
Cytoskeleton organisation	Microfilament network	Actin filament protein
		Actin polymerisation.profilin actin nucleation protein
		Alpha-beta-Tubulin heterodimer.component beta-Tubulin
		Alpha-beta-Tubulin heterodimer.component alpha-Tubulin
		Kinesin microtubule-based motor protein activities.motor protein (Kinesin-5)

DNA damage response	Nonhomologous end-joining (NHEJ) repair	Protein ADP-ribosyltransferase (PARP3)
Enzyme classification	EC_1 oxidoreductases	EC_1.1 oxidoreductase acting on CH-OH group of donor
		EC_1.13 oxidoreductase acting on single donor with incorporation of molecular oxygen (oxygenase)
		EC_1.17 oxidoreductase acting on CH or CH2 group
		EC_1.2 oxidoreductase acting on aldehyde or oxo group of donor
		EC_1.3 oxidoreductase acting on CH-CH group of donor
	EC_2 transferases	EC_2.4 glycosyltransferase
		EC_2.7 transferase transferring phosphorus-containing group
	EC_3 hydrolases	EC_3.2 glycosylase
	EC_4 lyases	EC_4.2 carbon-oxygen lyase
	EC_5 isomerases	EC_5.3 intramolecular oxidoreductase
		EC_5.4 intramolecular transferase
External stimuli response	heat response	Cytosolic chaperone (Hsp101)
Lipid metabolism	Fatty acid metabolism (acetyl-CoA carboxylation)	Polymeric acetyl-CoA carboxylase complex.BC-BCCP-BADC subcomplex.biotin carboxylase subunit (BC)
	Fatty acid metabolism (acetyl-CoA generation)	Plastidial pyruvate dehydrogenase complex.dihydrolipoamide dehydrogenase component E3
	Fatty acid metabolism (citrate shuttle)	Cytosolic NADP-dependent malic enzyme
	Fatty acid metabolism (fatty acid degradation)	Alpha-oxidation.alpha dioxygenase
		Core beta-oxidation.3-ketoacyl-CoA thiolase (KAT)
		Glyoxylate cycle.citrate synthase
	Fatty acid metabolism (plastidial fatty acid synthase ptFAS system)	Enoyl-ACP reductase (ER)
	Lipid droplet-associated activities	Associated factor (OBAP)
		Peroxygenase (CALEOSIN/CLO/PXG)
Multi-process regulation	14-3-3 regulatory system	14-3-3 phosphoprotein-binding protein (GRF)
	Calcium-dependent signalling	Calcium sensor (cam)

Nucleotide metabolism	Salvage pathway (deoxynucleotides)	Nucleoside diphosphate kinase
	Purines (phosphotransfers)	Adenylate kinase
	Salvage pathway (purines)	Adenosine kinase (ADK)
	Pyrimidines (catabolism)	.dihydropyrimidine dehydrogenase (PYD1)
	Pyrimidines (de novo biosynthesis)	UMP synthase (UMPS)
Nutrient uptake	Nitrogen assimilation	Ammonium assimilation.glutamine synthetase activities.cytosolic glutamine synthetase (GLN1)
		Aspartate aminotransferase
	Sulfate assimilation	Fd-dependent sulfite reductase (SIR)
	Metallochaperone activities	Iron storage protein (FER)
Photosynthesis.calvin cycle.	CAM/C4	Fructose 1,6-bisphosphate aldolase
		Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) activity.glyceraldehyde 3-phosphate dehydrogenase (GAPDH)
		Phosphoglycerate kinase
		Ribulose-1,5-bisphosphat carboxylase/oxygenase (rubisco) activity.rubisco heterodimer.small subunit
		Ribulose-1,5-bisphosphat carboxylase/oxygenase (rubisco) activity.rubisco assembly.CPN60 chaperonin heterodimer.subunit beta
		NAD-dependent malate dehydrogenase
	Photophosphorylation	Phosphoenolpyruvate (PEP) carboxylase activity.PEP carboxylase (PPC)
		ATP synthase complex.peripheral CF1 subcomplex.subunit beta
		Cpn60 chaperonin heterodimer.subunit alpha
	Photosystem I	Cytb6/f to PS-I, plastocyanin copper-protein (pete)
	Photosystem II	Thioredoxin (TRX-M)
		LHC-II complex.component lhcb5
		Oxygen-evolving center (OEC) component OEC23/psbp
	Photorespiration	Non-peroxisomal hydroxypyruvate reductase (HPR)
		Serine hydroxymethyltransferase (SHM)
Polyamine metabolism	Spermidine biosynthesis	Spermidine synthase

Protein biosynthesis	Aminoacyl-tRNA synthetase activities	Serine-tRNA ligase
	Organelle machinery	Elongation factor (EF-Tu)
	Large ribosomal subunit (LSU) proteome	Component ul10, el27, ul6, ul16, el22, ul4, ul18, el13, el8, el34, ul30, ul14, el24, ul13, el14, el37, ul15, ul22, ul10, el18, ul1, el43, ul2, el32, el19, ul3, el8, ul3, ul23, el6, ul5, ul29, el36
	LSU processome	Ribosome assembly factor (eif6)
	small ribosomal subunit (SSU) proteome	Component us2, us7, es10, es4, us15, us11, us13, es24, us17, es6, es17, RACK1, us15, us17, us7, es25, es1, es7, us9, us3, us19, us15, es7, us15, es28, es25, us12
	eEF1 aminoacyl-tRNA binding factor activity	Aminoacyl-tRNA binding factor (eef1a)
		Component eef1b-beta/-delta
		Component eef1b-gamma
		Component eef1b-beta/-delta
		Aminoacyl-tRNA binding factor (eef1a)
	eEF2 mRNA-translocation factor activity	mRNA-translocation factor (eef2)
		Poly-P/G elongation factor (eef5/eif5a)
	mRNA loading	mRNA poly-A-tail binding factor (PABP)
	eIF1 PIC assembly factor activity	Assembly factor (eif1)
		Assembly factor (eif1a)
	eIF3 mRNA-to-PIC binding complex	Component eif3b
		Component eif3c
Protein homeostasis	cytosolic Hsp70 chaperone system	Co-chaperone (Hsp40)
		Chaperone (Hsp70)
	BiP chaperone system.	Erdj3b-bip-SDF2 chaperone complex.chaperone component bip
	calnexin/calreticulin chaperone system	ER luminal lectin chaperone (CRT)
		Membrane-anchored lectin chaperone (CNX)
	Hsp60 chaperone system	Chaperone (Hsp60)
	Hsp90 chaperone system	Chaperone (Hsp90)
	mitochondrial Hsp70 chaperone system	Chaperone (mthsc70)
	plastidial Hsp70 chaperone system	Chaperone (cphsc70)
	ribosome-associated chaperone activities	NAC chaperone heterodimer subunit alpha
	smallHsp holdase chaperone activities	Class-C-I protein
		Class-C-II protein

	seed storage protein activities	7S globulin family.globulin seed storage protein (PAP85)
		11S globulin family.globulin seed storage protein (CR)
	aspartic-type peptidase activities	A1-class protease (Pepsin)
	aminopeptidase activities	M17-class leucyl aminopeptidase (LAP)
	metallopeptidase activities	M3-class metalloprotease (Thimet)
	protease inhibitor activities	Bowman-Birk protease inhibitor
		Kunitz protease inhibitor
		PR6 protease inhibitor
	serine-type peptidase activities	S10-class serine carboxypeptidase (SCPL)
		Prolyl aminopeptidase (PAP1)
	ubiquitin-proteasome system (26S/19S regulatory particle)	Component RPT3, RPN2, RPT6, RPN1, RPN2, RPN5, RPT4, RPT1, RPT2, RPN12, RPN8, RPN9, RPN11, RPN7, RPT5
	ubiquitin-proteasome system (26S/20S core particle)	Component alpha type-3, alpha type-1, beta type-2, beta type-1, alpha type-2, beta type-3, beta type-4, beta type-5
	CDC48-NPL4-UFD1 chaperone complex	Platform atpase component CDC48
Protein modification	cysteine disulfide formation	Protein disulfide isomerase (PDI-A)
		Protein disulfide isomerase (PDI-L)
	glycosylation	N-linked glycosylation.oligosaccharyl transferase (OST) complex.component DGL1
	phosphorylation	CK protein kinase superfamily.protein kinase (CKL)
		Protein serine/threonine phosphatase superfamily.PPM/PP2C Mn/Mg-dependent phosphatase families.clade F phosphatase
	protein folding	Protein folding catalyst (Cyclophilin)
		Protein folding catalyst (FKBP)
	S-glutathionylation	Glutaredoxin
		S-nitrosoglutathione reductase (GSNOR)
Protein translocation	nucleus.nucleocytoplasmic transport	Chloroplast.outer envelope TOC translocation system.translocation channel (Toc75-III)
		Cargo adaptor protein (IMP-alpha)
		RAN gtpase cycle gtpase (Ran)
		RAN gtpase cycle.Ran-activation accessory protein (ranbp1)

Redox homeostasis	ascorbate-based redox regulation	Ascorbate metabolism.L-galactose biosynthesis pathway.GDP-D-mannose-epimerase (GME)
		Cytosolic ascorbate peroxidase (APX)
		Dehydroascorbate reductase (DHAR)
		Glutathione reductase (GR)
	enzymatic reactive oxygen scavenging	Catalase
		Glutathione peroxidase
		Superoxide dismutase activities.copper/zinc superoxide dismutase (CSD)
		Superoxide dismutase activities.manganese superoxide dismutase (MSD)
		Glutathione-based redox regulation.glutathione S-transferase activities.class tau glutathione S-transferase
	thiol-based redox regulation	Typical 2-Cys peroxiredoxin (2-cysprx)
		Type-2 peroxiredoxin (prxii)
		1-Cys peroxiredoxin (1-cysprx)
		NADPH-dependent thioredoxin reductase (NTRC)
	tocopherol biosynthesis	MPBQ-methyltransferase (VTE3)
RNA biosynthesis		RNA polymerase II-dependent transcription.transcription termination.R-loop removal.R-loop reader protein (ALBA1/2)
RNA processing		RNA-induced silencing complex (RISC) assembly miRNA recruiting factor (AGO1/AGO10)
		Organelle machinery.pre-RNA splicing.plastidial RNA splicing.maturase (matk)
		Organelle machinery.RNA modification.C-to-U RNA editing.RNA editing factor (ORRM)
		RNA export.TREX/THO ribonucleoparticle (RNP) trafficking complex.helicase (UAP56)
Secondary metabolism	glycinebetaine biosynthesis	Betaine-aldehyde dehydrogenase
	flavonoid biosynthesis	Chalcone isomerase
		Polyketide/acetate pathway.3-ketoacyl-CoA thiolase (KAT)
	mevalonate pathway (terpenoids)	Acetyl-CoA C-acyltransferase
		Isopentenyl diphosphate isomerase
		Terpene biosynthesis.mono-/sesquiterpene-/diterpene synthase

Solute transport	carrier-mediated transport	APC superfamily.NCS-2 family.solute transporter (NAT)
		DMT superfamily.NST-TPT group.phosphometabolite transporter (TPT/PPT/GPT/XPT)
		Solute transporter (MTCC)
	MIP family Channels	Plasma membrane intrinsic protein (PIP)
		Tonoplast intrinsic protein (TIP)
	porins	Voltage-gated anion channel (VDAC)
	primary active transport (V-type ATPase complex)	Peripheral V1 subcomplex.subunit A
		Atpase complex.membrane V0 subcomplex.subunit c
Vesicle trafficking	anterograde trafficking	Coat protein recruiting.small gtpase (Sar1)
		Coat protein complex.scaffolding component Sec31
	vacuolar protein sorting	Lytic vacuole sorting receptor (VSR)
	endocytic trafficking	ESCRT-III complex.component VPS32/SNF7
	exocytic trafficking	Exocyst complex.component SEC5
	multi-pathway trafficking regulation	Clathrin triskelion structure heavy chain
		Vesicle tethering B-class RAB gtpase
		Vesicle tethering D-class RAB gtpase
		Vesicle tethering A-class RAB gtpase
		Vesicle tethering C-class RAB gtpase
		Vesicle tethering G-class RAB gtpase
	retrograde trafficking	Coat protein I (COPI) ccomplex subunit beta2
		Coat protein I (COPI) subunit alpha
		Coat protein I (COPI) class-I ARF-gtpase (ARFA1)

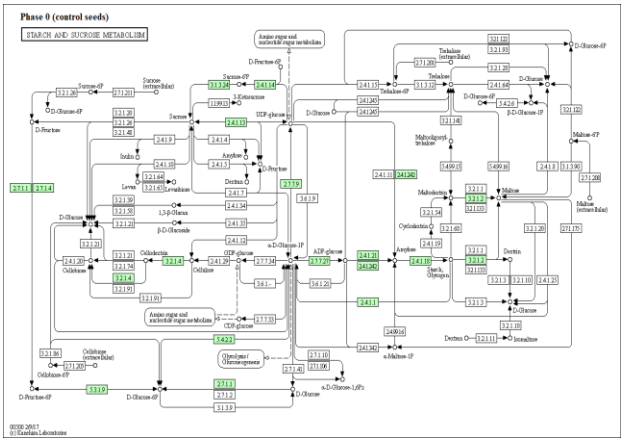
Appendix H List of not assigned but annotated proteins obtained through MapMan. The functions were collected from uniprot.org

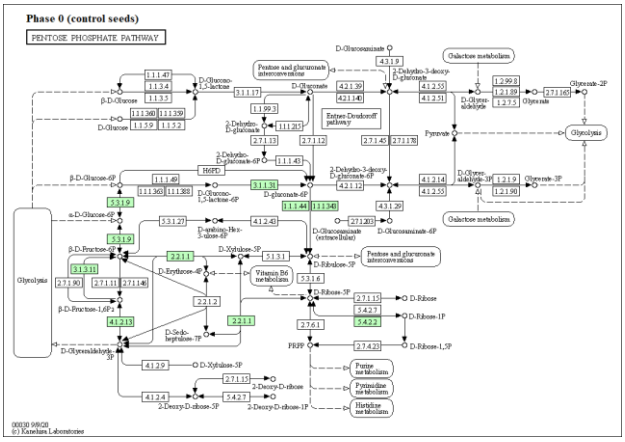
Protein	Phase 0	Phase II	Phase III	Phase IV	Function
Provicilin	100	100	100	100	Seed storage protein
Vicilin	100	100	100	100	Seed storage protein
Seed biotin-containing protein SBP65	100	100	100	100	Sourcefor biotin
Luminal-binding protein (2, 5)	100	100	100	100	Chaperone protein folding at ER
Lectin (alpha-1, beta-1, and beta-2)	100	100	100	100	Lectin which binds carbohydrates, abscisic acid-induced stomatal closure, disease resistance
Albumin (1, 2)	100	100	100	100	Seed storage protein
Prohibitin-1	100	100	100	100	Mitochondrion organization
Plastoglobulin-1	100	100	100	100	Abiotic stress response, light/cold stress-related jasmonate (JA) biosynthesis, PS I and II protection
Late embryogenesis abundant protein 47	100	100	100	100	Seed embryo protein
Non-specific lipid-transfer protein 4	100	100	100	100	Seed and ovule maturation and development
ABA-responsive protein ABR18	0	100	100	100	Biotic stress response
Disease resistance response protein DRRG49-C	0	100	100	100	Biotic stress response
Dehydrin DHN1	100	100	100	75	Abiotic stress response
Carrot ABA-induced in somatic embryos 3	100	100	100	100	Embryo protein
Translationally-controlled tumor protein homolog	100	100	100	100	Development (embroy to dormancy, root growth)
Hypersensitive-induced response protein 1	100	100	75	100	Induced response
Non-specific lipid-transfer protein 3	100	100	100	100	Transfer phospholipids as well as galactolipids across membranes, in wax or cutin deposition
Plastid-lipid-associated protein	100	100	100	100	Light/cold stress-related jasmonate (JA) biosynthesis. Response to osmotic stress, Contributes to the protection of photosystem II (PSII) against light stress.

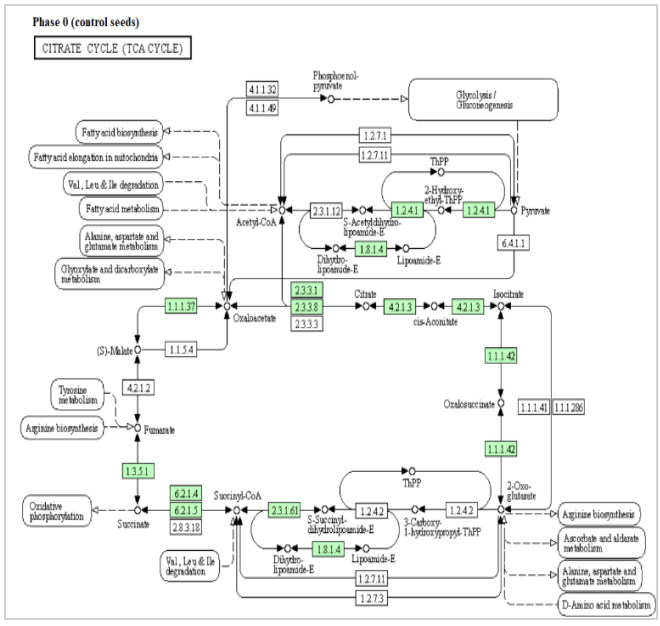
Protein	Phase 0	Phase II	Phase III	Phase IV	Function
Heat shock 70	100	100	100	100	Chaperone, protein folding, transcription regulation, stress response
Probable CCR4-associated factor 1 homolog 5	100	50	100	100	Defense response
Selenium-binding protein 1	100	0	0	0	Sulfate assimilation, response to metal ions
Calreticulin	100	100	100	100	Chaperone promoting folding, oligomeric assembly and quality control
Proton pump-interactor 1	100	75	50	0	Regulation of proton transport, mRNA binding
Tropinone reductase-like 2	100	75	50	75	Defense respons
Cyclase-like protein 1	100	50	75	100	Defense response, negative regulation of programmed cell death
Protein (SLE2, SLE1)	100	100	100	75	Late embryogenesis abundant (LEA) protein interacting with non-reducing sugars and phospholipids.
NADPH-dependent aldehyde reductase 1	100	100	100	100	Folic acid-containing compound metabolic process
Glucose and ribitol dehydrogenase	100	100	100	100	Alcohol dehydrogenase (NADP+) activity
Obg-like ATPase 1	100	100	100	100	Ribosome, ATP, gtpase, and metal ion binding activity
Outer envelope pore protein 16	100	100	100	100	Voltage-dependent cation channel
Peroxidase 17, C3	100	100	100	100	Removal of H2O2, auxin catabolism, and degradation of lignin
Universal stress protein PHOS32	75	100	100	100	ATP binding, response to biotic stress
ATP-dependent (S)-NAD(P)H-hydrate dehydratase	50	100	50	100	Na
Probable voltage-gated potassium channel subunit beta	100	100	100	100	Voltage-gated potassium channel
Transmembrane 9 superfamily member 11	0	50	75	0	Protein localization to membrane, trans-Golgi network
Nudix hydrolase 3	100	100	100	100	Hydrolysis of some nucleoside diphosphate derivatives
FAM10 family protein (At4g22670)	100	50	100	100	Hsp70 protein binding, chaperone cofactor-dependent protein refolding

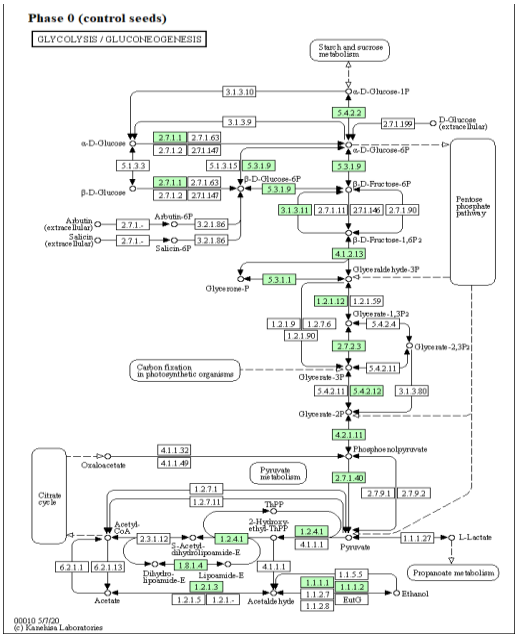
Protein	Phase 0	Phase II	Phase III	Phase IV	Function
NAD(P)H dehydrogenase (quinone) FQR1	100	100	100	100	Response to oxidative stress, cellular response to auxin stimulus
Probable bifunctional TENA-E protein	100	100	100	100	Involved in thiamine biosynthetic process
Transcription factor Pur-alpha 1	75	50	100	75	Regulation of transcription by RNA polymerase II
Cilia- and flagella-associated protein 43	50	75	25	0	Repressor of photomorphogenesis
Jacalin-related lectin 16	100	100	100	100	Carbohydrate binding
Probable methyltransferase PMT24	100	100	100	100	Methylation, trans-Golgi network
Phytohormone-binding protein	0	50	50	50	Ubiquitin-dependent protein catabolic process
Protein RETICULATA-RELATED 3	0	100	100	100	May play a role in leaf development. Required for leaf mesophyll cell division in the early stages of leaf organogenesis.
Desiccation protectant protein Lea14 homolog	100	100	100	100	Response to desiccation
Pumilio homolog	50	75	75	100	Translation and mRNA stability by binding the 3'-UTR of target mRNAs
ABA-responsive protein ABR17	0	75	100	100	Response to biotic stimulus
4-coumarate--CoA ligase-like 7	100	100	100	100	Beta-oxidative chain, jasmonic acid biosynthesis
Purple acid phosphatase 4	50	75	0	0	
COP1-interactive protein 1	75	75	0	0	(ABA)-mediated signaling, abiotic stress responses
Protein SHOOT GRAVITROPISM 6	25	50	50	0	Gravitropism
Defensin-2	100	100	100	100	Plant defense
Albumin-2	100	100	100	100	Storage protein
20 kDa chaperonin	100	100	100	100	Protein folding and tranportation
Defensin-1	100	100	100	100	Plant defense
11 kDa late embryogenesis abundant protein	100	100	100	75	Involved dehydration tolerance. Aids in seed germination
Wound-induced basic protein	100	100	100	100	Response to stress

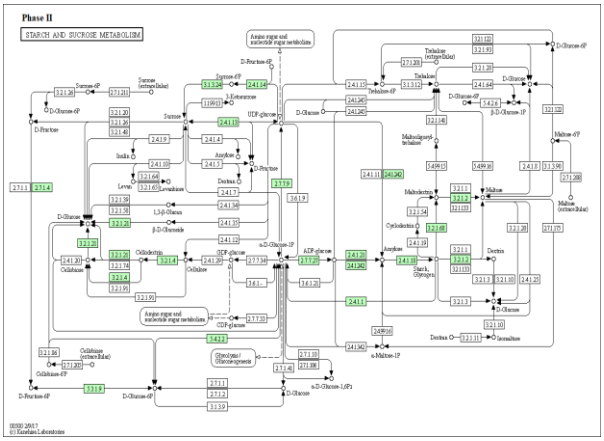
Appendix I Major metabolic pathways during seed germination in phase 0, phase II, phase III, and phase IV generated from KEGG. The pathways include glycolysis/gluconeogenesis, TCA cycle, starch and sucrose metabolism, and pentose phosphate pathway

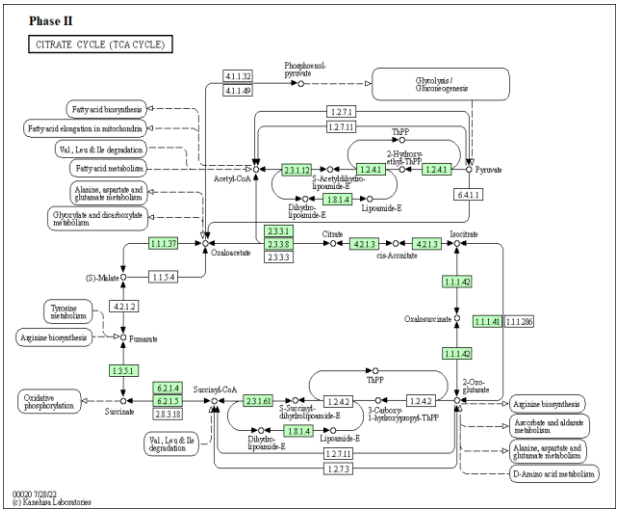


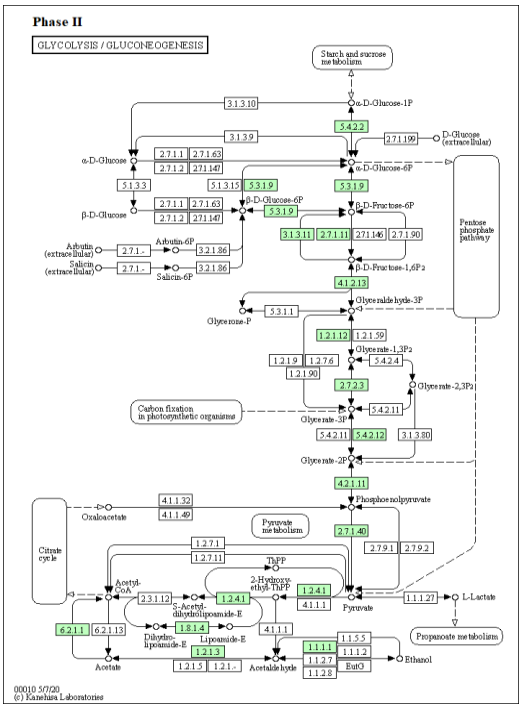


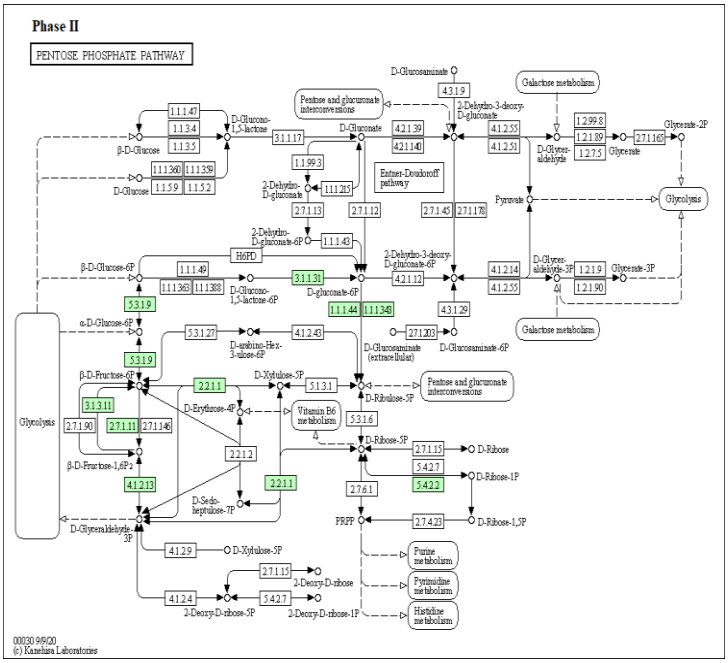


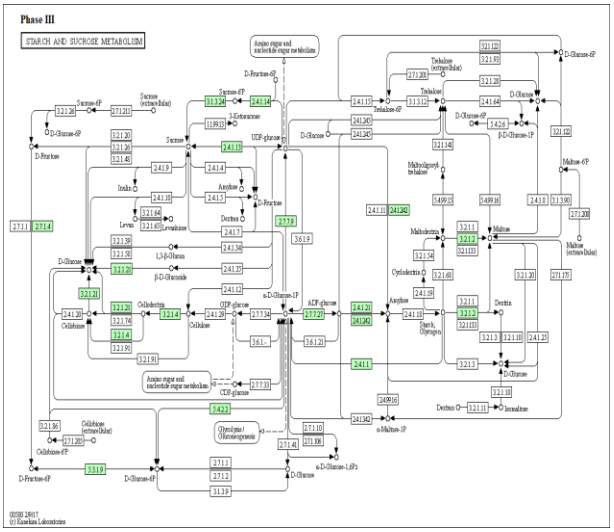


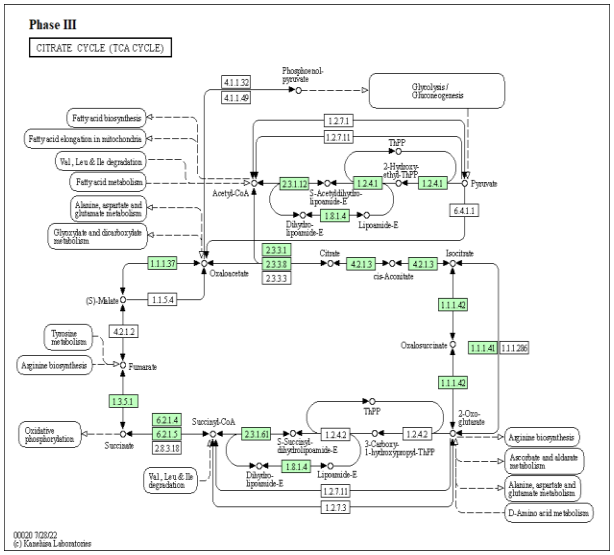

Phase II
STARCH AND SUCROSE METABOLISM

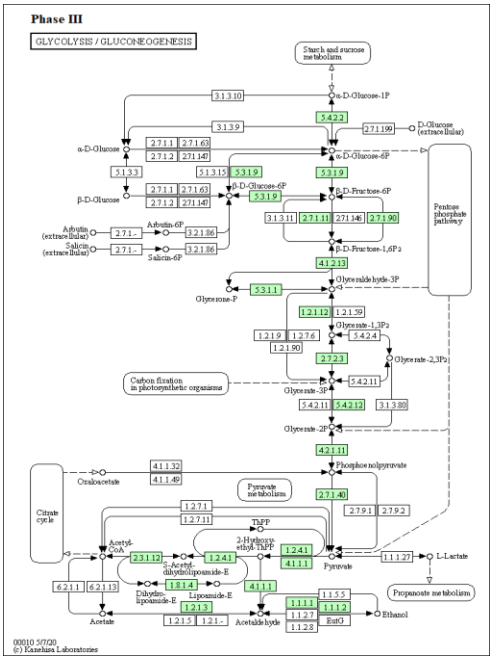

Phase II
CITRATE CYCLE (TCA CYCLE)

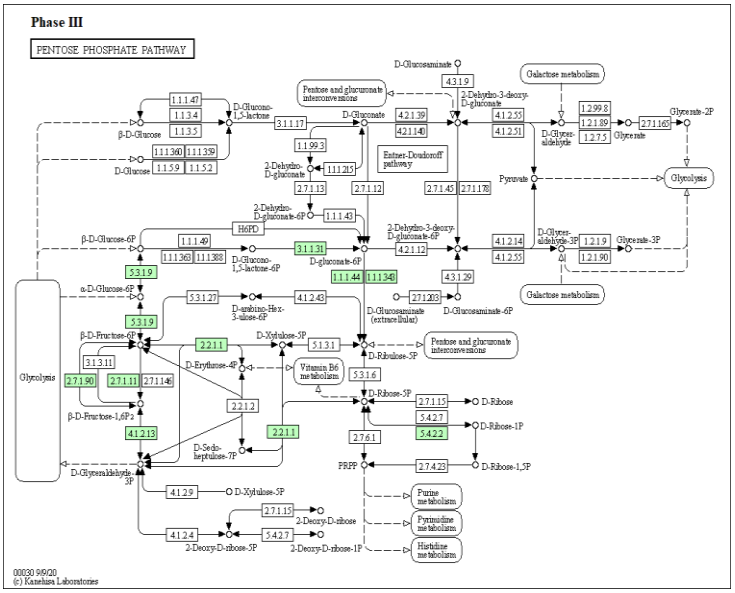

Phase II
GLYCOLYSIS / GLUCONEOGENESIS

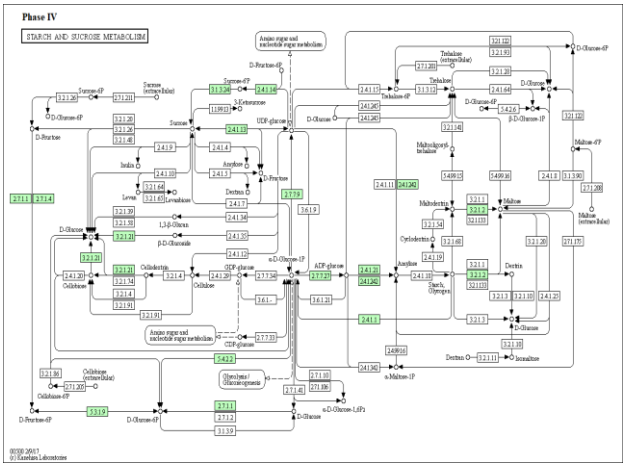

Phase II
PENTOSE PHOSPHATE PATHWAY

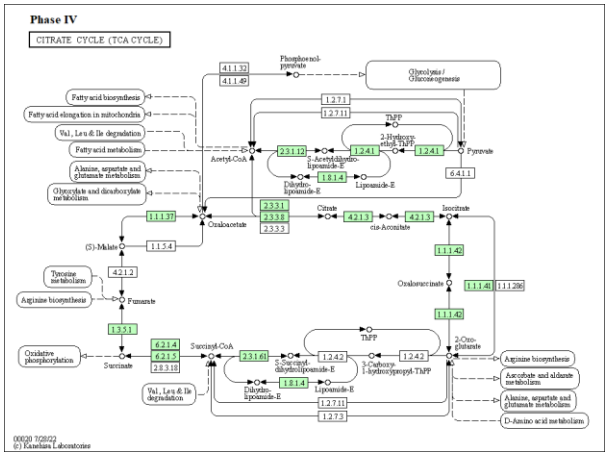

Phase III
STARCH AND SUCROSE METABOLISM

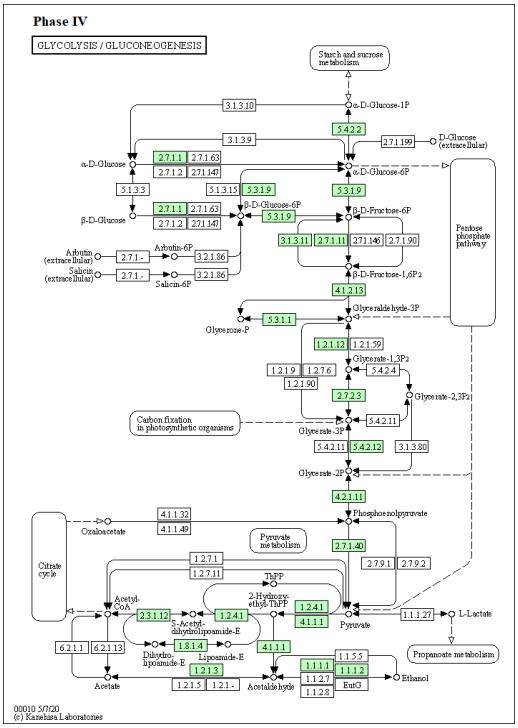

Phase III
CITRATE CYCLE (TCA CYCLE)

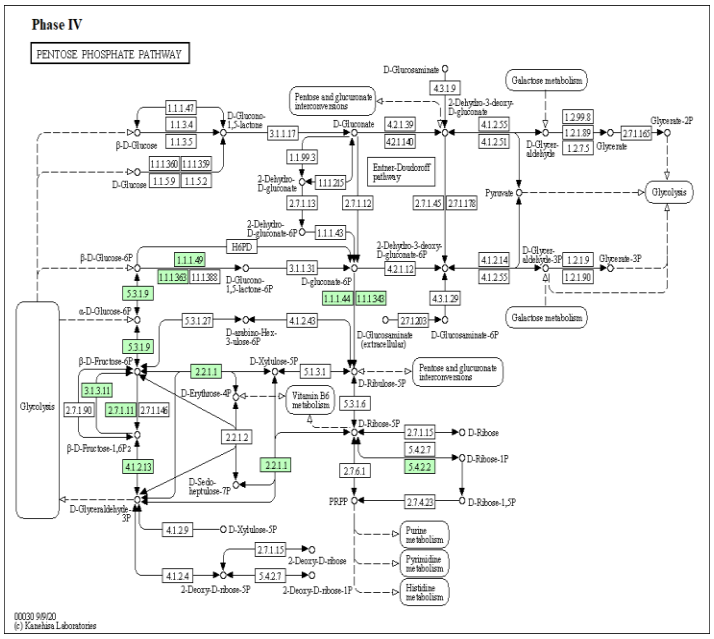

Phase III
GLYCOLYSIS / GLUCONEOGENESIS


Phase III
PENTOSE PHOSPHATE PATHWAY


Phase IV
STARCH AND SUCROSE METABOLISM


Phase IV
CITRATE CYCLE (TCA CYCLE)


Phase IV
GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
α-D-Glucose-1P
D-Glucose (extracellular)
α-D-Glucose
α-D-Glucose-6P
β-D-Glucose
β-D-Glucose-6P
β-D-Fructose-6P
Pentose phosphate pathway
Arbutin (extracellular)
Arbutin-6P
Salicin (extracellular)
Salicin-6P
β-D-Fructose-1,6P2
Glyceraldehyde-3P
Glycerose-P
Glycerate-1,3P2
Glycerate-2,3P2
Carbon fixation in photosynthetic organisms
Glycerate-3P
Glycerate-2P
Phosphoenolpyruvate
Oxaloacetate
Pyruvate metabolism
Citrate cycle
ThPP
Acetyl-CoA
2-Hydroxyethyl-ThPP
2-Acetyldihydrolipoamide-E
Pyruvate
L-Lactate
Propanoate metabolism
Dihydrolipoamide-E
Liposamide-E
Acetaldehyde
Ethanol
Acetate
00010 5/7/20
(c) Kanehisa Laboratories


Phase IV
PENTOSE PHOSPHATE PATHWAY
D-Glucosaminate
Galactose metabolism
Pentose and glucuronate interconversions
2-Dehydro-3-deoxy-D-gluconate
D-Glucono-1,5-lactone
D-Gluconate
β-D-Glucose
Entner-Doudoroff pathway
D-Glucose
2-Dehydro-D-gluconate
Glycerate-2P
Glycerate
Pyruvate
Glycolysis
2-Dehydro-D-gluconate-6P
H6PD
β-D-Glucose-6P
2-Dehydro-3-deoxy-D-gluconate-6P
D-Glucono-1,5-lactone-6P
D-gluconate-6P
Glycerate-3P
α-D-Glucose-6P
D-arabino-Hex-3-ulose-6P
D-Glucosaminate (extracellular)
D-Glucosaminate-6P
β-D-Fructose-6P
D-Xylulose-5P
D-Ribulose-5P
Vitamin B6 metabolism
D-Erythrose-4P
D-Ribose-5P
D-Ribose
D-Ribose-1P
β-D-Fructose-1,6P2
D-Sedo-heptulose-7P
PRPP
D-Ribose-1,5P
D-Glyceraldehyde-3P
D-Xylulose-5P
Purine metabolism
Pyrimidine metabolism
2-Deoxy-D-ribose
2-Deoxy-D-ribose-5P
2-Deoxy-D-ribose-1P
Histidine metabolism
00030 9/9/20
(c) Kanehisa Laboratories

Appendix J List of identified peptides from three gel band of SDS-PAGE experiment

Protein name	Mascot Score	P-value (Expect)	Total number of peptides	Number of unique peptides	Peptide sequences
Vicilin	27-106	1.8e-06 to .00011	26	19	K.AILTVLKPDDR.N, K.AILTVLKPDDR.N, K.LPAGTIAYLVNR.D, K.LPAGTIAYLVNR.D, R.QQSQEENVIVK.L, K.SVSSESEPFNLR.S, K.EGSLLLPHYNSR.A, R.NSFNLERGDTIK.L, K.ELAFPGSAQEVDR.I, K.RQQSQEENVIVK.L, K.RQQSQEENVIVK.L, K.KSVSSESEPFNLR.S, K.KSVSSESEPFNLR.S, R.GDTIKLPAGTIAYLVNR.D, K.NILEASFNTDYEEIEK.V, K.ASSNLDLLGFGINAENNQR.N, K.ASSNLDLLGFGINAENNQR.N, K.ELAFPGSAQEVDRILENQK.Q, K.ELAFPGSAQEVDRILENQK.Q
Seed linoleate 9S-lipoxygenase	32-83	2.3e-06 to .04	26	26	R.YSMEMSSK.V, R.YSMEMSSK.V, + Oxidation (M), R.YSMEMSSK.V, + Oxidation (M), R.YSMEMSSK.V, + 2 Oxidation (M), R.SDYVYLPR.D, R.EEELNNLR.G, K.IKGTVVLMR.K, + Oxidation (M), K.ELVEVGHGDK.K, K.GSAEFEELVK.S, K.SSDFLTYGLK.A, R.DTMNINSLAR.L, R.DTMNINSLAR.L, + Oxidation (M), K.ELVEVGHGDKK.N, R.LSLVNDGGIIEK.T, R.HASDELYLGER.D, R.SGWMTDEEFAR.E, R.EEELNNLRGDGTGER.K, K.LPTNILSQISPLPVLK.E, R.FMPEKGSAEFEELVK.S, + Oxidation (M), K.EHLEPNLEGLTVEEAIQNK.K, K.ATFLEGIISSLPTLGAGQSAFK.I, K.ATFLEGIISSLPTLGAGQSAFK.I, K.EHLEPNLEGLTVEEAIQNKK.L, K.EHLEPNLEGLTVEEAIQNKK.L, R.IFFANQTYLPSETPAPLVHYR.E, R.IYDYDVYNDLGNPDSGENHARPVLGGSETYPYPR.R

Protein name	Mascot Score	P-value (Expect)	Total number of peptides	Number of unique peptides	Peptide sequences
Convicilin	32-73	1.5e-05 to 0.033	14	9	R.NPIYSNK.F, R.NPIYSNK.F, R.SRNPIYSNK.F , R.SDQENPFIFK.S, K.FGKFFEITPEK.N, K.ELAFPGSSHEVDR.-, K.ELAFPGSSHEVDR.-, R.FQTLYENENGHIR.L, R.FQTLYENENGHIR.L
Legumin type B	27-78	7.4e-06 to 0.035	5	5	R.AAVSHVQQVFR.A, R.ATPADVLANAFGLR.Q, R.ATPADVLANAFGLR.Q, R.VFYLGGNPEVEFPETQEEQQER.H, R.VFYLGGNPEVEFPETQEEQQER.H