BIOCHEMICAL AND MOLECULAR CHARACTERIZATION OF RAFFINOSE FAMILY OLIGOSACCHARIDE (RFO) BIOSYNTHESIS IN LENTIL (*Lens culinaris* Medik.)

A Thesis Submitted to the College of Graduate Studies and Research

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

in the Department of Plant Sciences,

University of Saskatchewan,

Saskatoon, Saskatchewan

By

Udhaya Kannan

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate 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 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 use of material in thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences

51 Campus Drive

University of Saskatchewan

Saskatoon, Saskatchewan, Canada

S7N 5A8

ABSTRACT

Lentil (*Lens culinaris* Medik.) is an important pulse crop and a balanced source of carbohydrates, proteins, minerals and vitamins. Among the carbohydrates, lentils contain raffinose family oligosaccharides (RFO) or sucrosyl oligosaccharides that upon consumption remain undigested in the lower bowel of humans and monogastric animals causing flatulence and stomach discomfort. The high concentration of RFO in lentils limits human consumption and acceptance worldwide. To develop strategies to reduce RFO concentration in lentils, it is imperative to understand the biochemical and molecular mechanism of RFO deposition in lentil seeds. The main objective of this study was to isolate and characterize RFO biosynthetic genes galactinol synthase (GS or GolS; EC 2.4.1.123), raffinose synthase (RS; EC 2.4.1.82) and stachyose synthase (STS; EC 2.4.1.67). The expression of RFO biosynthetic genes, corresponding enzyme activity and RFO accumulation were also analyzed during lentil seed development.

Gene expression studies in developing seeds necessitated an optimized method to extract RNA with high purity in adequate quantities. Previously used RNA isolation procedures, namely (i) Trizol based RNA isolation; (ii) Phenol guanidine based RNA isolation; and (iii) Lysis buffer; were tested with modifications. All three methods did not yield RNA with the required purity and/or in adequate quantity. A CTAB (hexadecyltrimethylammonium bromide) based RNA isolation procedure with lithium chloride precipitation yielded RNA in adequate quantity (210 - 260 μg from 200 mg of seeds) and required purity (A260 / 280 ratio of about 2.2). The optimized RNA isolation procedure was used to isolate mRNA to construct a cDNA library from developing lentil seeds and in gene expression studies during seed development.

Galactinol synthase (GS or *GolS*) catalyses the synthesis of galactinol, the first committed step in RFO biosynthesis. Screening a cDNA library prepared from developing lentil seeds identified two cDNA clones *LcGolS1* (1336 bp, ORF - 1002 bp, 334 amino acids) and *LcGolS2* (1324 bp, ORF - 975 bp, 325 amino acids). Predicted molecular weights of *LcGolS1* and *LcGolS2* were 38.7 kDa and 37.6 kDa, respectively. Real time quantitative PCR studies revealed that the two *GolS* alleles were differentially expressed. *LcGolS1* increased from 20 days after flowering (DAF), with rapid increase from 24 DAF, and maximum expression was obtained at 32 DAF. The expression of *LcGolS2* was observed at 24 DAF and decreased at 26 DAF. Due

to similarity in nucleotide sequences, RNA gel blot analyses could not distinguish transcript accumulation due to two *LcGolS* alleles. *LcGolS* transcript accumulation corresponded to the GS enzyme activity that was highest at 26 and 28 DAF and gradually decreased as the seed matured.

For raffinose synthase (RS), a full length *LcRS* sequence of 2379 bp, coding for 792 amino acids and predicted molecular weight of 88.2 kDa was isolated. Gene expression of *LcRS* during seed development showed maximum expression at 24 - 26 DAF. Raffinose synthase enzyme activity correlated with gene expression with maximum activity obtained at 24 DAF. Raffinose concentration during seed development showed rapid increase from 20 DAF with maximum accumulation at 28 DAF. Raffinose accumulation correlated with accumulation of its precursors, sucrose and galactinol during seed development. *LcRS* gene expression and RS enzyme activity correlated with oligosaccharide accumulation, suggesting that the isolated gene contributed to raffinose accumulation in seeds.

In lentil seeds the predominant RFO is stachyose, synthesized by stachyose synthase (STS) which transfers a galactosyl residue from galactinol to raffinose. A cDNA clone *LcSTS* was isolated that is 2669 bp long (5'UTR - 30bp, ORF - 2559, 3'UTR 80 bp followed by polyA) coding for 853 amino acids with predicted molecular weight of 96 kDa. RNA gel blot and qPCR analyses showed maximum accumulation of *LcSTS* transcript at 26 and 24 DAF respectively. Stachyose and verbascose concentrations increased during seed development with maximum concentration at 32 DAF. STS enzyme activity correlated with *LcSTS* expression resulting in maximum activity at 26 DAF. The gene coding for verbascose synthase (VS) was not isolated in this study, however enzyme activity for VS was studied during seed development. Maximum VS enzyme activity was observed at 26 DAF. RFO can also be synthesized by galactinol independent enzyme (Galactan:galactan galactosyl transferase, GGT) and its activity was also detected in developing lentil seeds. Maximum GGT activity was observed from 26 - 32 DAF. Throughout the seed developmental stages, oligosaccharide accumulation correlated well with transcript accumulation and enzyme activity of RFO biosynthetic genes.

In conclusion, genes for three key enzymes in RFO biosynthesis were identified and characterized by DNA sequencing. During lentil seed development, transcript accumulation of the three genes correlated with respective enzyme activity and corresponding RFO accumulation. The gene sequences for these three key RFO biosynthetic genes can be used in the near future to

study allelic variation to understand their role in influencing total RFO or its constituent concentrations in lentils.

ACKNOWLEDGEMENTS

I would like to express my gratitude and sincere thanks to my PhD supervisor Prof. Ravindra N Chibbar for providing me the opportunity to pursue my PhD under his supervision. His continued guidance, trust and support has made my PhD thesis possible. I would also like to thank my committee members Prof. Yuguang Bai / Prof. Bruce Coulman, Prof. Albert Vandenberg, Prof. Nicholas Low and Prof. Rosalind Bueckert for their valued comments in my research. I would also like to thank Dr. Ian Tetlow (external examiner) for his suggestions in my thesis.

My sincere thanks to Dr. Pooba Ganeshan for his continued help and guidance throughout my graduate program. I deeply appreciate his time and effort all along my study. I would also like to express my sincere gratitude to Dr. Monica Baga for her help with my experiments and her suggestions. My special thanks to Dr. Roopam Sharma, Dr. Yogendra Khedikar and Dr. Manu Gangola for their help with my experiments. I would also like to thank Dr. Sarita Jaiswal and all members of molecular crop quality lab for their continued help and encouragement. Their concern and affection has helped me overcome many hurdles professionally and personally that came my way.

Funding support from Plant sciences graduate student scholarship, Rene Vandevald postgraduate scholarship, education enhancement award (College of Agriculture and Bioresources), Natural Sciences and Engineering Research Council (NSERC) Collaborative Research Program (CRD) with Saskatchewan Pulse Growers, Canada Research Chairs and Agriculture and Agri-Food Canada Internationalization of Research and Development (IRD) program is kindly acknowledged. I would like to thank all my friends in Saskatoon for their continued support and encouragement all along my stay in Saskatoon. Last but not least, I would like to thank my parents, sister and brother for their love and without their support this would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

bp	:	Base pair
cDNA	:	Complementary Deoxyribonucleic Acid
CTAB	:	Cetyltrimethyl ammonium bromide
DAF	:	Days after flowering
DEPC	:	Diethyl pyrocarbonate
DM	:	Dry matter
DTT	:	Dithiothreitol
EDTA	:	Ethylenediaminetetraacetic acid
GS	:	Galactinol synthase
HEPES	:	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid, N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)
HPAEC-PAD	:	High performance anion exchange chromatography with pulse amperometric detection
kDa	:	Kilodalton
LiCl	:	Lithium chloride
$MgSO_4$	:	Magnesium sulphate
NaCl	:	Sodium chloride
Na_2HPO_4	:	Disodium hydrogen phosphate
NaOH	:	Sodium hydroxide
NH_4Cl	:	Ammonium chloride
ORF	:	Open reading Frame
PVP	:	Polyvinylpyrrolidone
RFO	:	Raffinose family oligosaccharides
RPM	:	Revolutions per minute
RS	:	Raffinose synthase
SDS	:	Sodium dodecyl sulfate
STS	:	Stachyose synthase
VS	:	Verbascose synthase
GGT	:	Galactan:galactan galactosyl transferase

APPENDICES

1. INTRODUCTION

1.1 Research background

Lens culinaris Medik. is one of the eight founder crops which include emmer wheat, einkorn wheat, hulled barley, lentil, chickpea, bitter vetch and flax. Archeological evidence suggests that lentil originated at the fertile-crescent, Near East along with other founder crops (Zohary and Hopf 1973). Lentil is widely consumed in the Asian and Mediterranean countries for its high nutritional value. Lentil seeds are a good source of carbohydrates, proteins, vitamins and minerals. Compared to cereals, lentils have higher protein content and contain essential amino acids including lysine which is deficient in cereals. Lentil is deficient in sulphur containing amino acids, hence consumption of lentils with cereals contribute to a well balanced diet. Lentils carbohydrates are also a rich source of dietary fibre and starch with higher amylose concentration that makes it less digestible and as a consequence reduces risk of chronic diseases such as diabetes, obesity and coronary heart diseases (Leterme 2002; Chibbar et al 2010). However, presence of high concentrations of raffinose family oligosaccharides in seeds deter lentil consumption by humans in some parts of the world, especially in developed countries (Frias et al 1994).

Raffinose family oligosaccharides (RFO) are considered anti-nutritional as they cause stomach discomfort and flatulence in humans and monogastric animals. Alpha - galactosidase, which is required to hydrolyze α 1→6 glycosidic linkages is deficient in humans and monogastric animals, therefore RFO remain undigested in the upper gastrointestinal tract. In the human lower gut, undigested oligosaccharides are fermented by anaerobic bacteria causing stomach discomfort, bloating, flatulence and diarrhea (Price et al 1988). Several methods such as soaking in water, germination, partial fermentation and addition of α-galactosidases have been used to hydrolyze RFO to reduce their concentration in seeds (Vidal-Valverde and Frias 1992; Frias et al 2000; Frias et al 2003). These processes are time consuming and inefficient, therefore genetically improving lentils to decrease RFO in seeds will not only reduce processing costs but also promote human consumption of lentils.

RFO are ubiquitous in plant tissues and are involved in several plant processes (Obendorf and Górecki 2012). They play an important role in phloem transport in plants and are also reported to protect plants during abiotic (Hannah et al 2006; Nishizawa et al 2008) and biotic stress (Gil et al 2012). RFO are synthesized during later stages of seed development and are

postulated to confer desiccation tolerance (Martínez-Villaluenga et al 2008a). Concentration of total α-galactosides in lentil seeds was reported in the range 1.8 - 6.8 g/100g dry matter (Grusak 2009). In 11 cultivated varieties of lentil in Saskatchewan, total RFO concentration had limited variation (4.5 to 5.5 mmoles/100g seed meal), with stachyose being the predominant RFO member followed by raffinose and small quantities of verbascose (Tahir et al 2011a). Genotype by environment interaction revealed a high heritability for RFO concentration (Tahir et al 2011a). In another study that included 122 domesticated lentil genotypes originating from 56 different countries and 46 wild relatives showed a highly significant difference in concentrations of total RFO and its components (Tahir et al 2012). This suggested that in the genus *Lens*, natural genetic variation for RFO related traits can be used to develop lentil cultivars with desired RFO concentration. No correlation between RFO and other quality traits such as starch and protein, suggested that selection for low RFO concentration may not affect other significant seed quality traits such as proteins and starch (Tahir et al 2011b).

Detection and quantitative determination of seed quality traits such as RFO and its components are resource intensive and time consuming, requiring precise measurements in mature seeds. Therefore, availability of precise and perfect molecular markers will allow screening of plant material in early stages of crop improvement thus saving time and resources (Singh et al 2012). Large scale genome and transcriptome analyses have often been used to identify DNA markers for agronomic traits (Edwards and Batley 2010). However, for seed quality traits such as RFO, large scale genomic techniques have shown limited value. Screening of lentil transcriptome data (Kaur et al 2011; Sharpe et al 2013) showed only a very few transcripts related to RFO biosynthetic genes. Therefore, to develop perfect DNA markers for RFO related traits, it is imperative to identify and characterize RFO biosynthetic genes in lentils.

Raffinose, stachyose and verbascose are the major RFO members with some higher homologues also found in seeds of some species (Dey 1985). RFO in seeds are synthesized through galactinol dependent and independent pathways (Peterbauer et al 2001a). Synthesis of galactinol is the first committed step in the galactinol dependent RFO biosynthetic pathway. Galactinol synthase (EC 2.4.1.123) catalyzes the synthesis of galactinol from *myo*-inositol and UDP- galactose. Raffinose is synthesized by the transfer of a galactosyl residue from galactinol to sucrose catalyzed by the enzyme raffinose synthase (EC 2.4.1.82). Elongation of raffinose by the addition of a galactosyl residue from galactinol is catalyzed by the enzyme stachyose

synthase (EC 2.4.1.67) to yield stachyose. Verbascose is synthesized by addition of a galactosyl residue from galactinol to stachyose catalyzed by the enzyme verbascose synthase. Presence of the enzyme galactan: galactan galactosyl transferase (GGT) catalyzing galactinol independent RFO biosynthesis was first reported in *Ajuga reptans* (Bachmann et al 1994). Variation in the RFO biosynthetic enzyme activity and expression of their genes has been reported to explain the variation in RFO concentration and composition in seeds of pea and soybean (Peterbauer et al 2001b; Sebastian et al 2000; Dierking and Bilyeu 2008; Skoneczka et al 2009). These results indicate that RFO biosynthetic enzymes influence RFO accumulation in the seeds suggesting that, possible differences in gene expression and activity might lead to altered RFO concentration in seeds. To alter RFO concentration in seeds it is crucial to understand the biosynthetic pathway in developing lentil seeds.

1.2 Research hypothesis

Changes in expression of genes for major RFO biosynthetic enzymes and their activity contribute to lentil seed RFO concentration and composition.

1.3 Objectives

The main objective of this research was to isolate, characterize and analyze the expression of major RFO biosynthetic enzymes in lentil seeds. To achieve this main objective, the following sub-objectives were established.

a) To optimize an efficient protocol to isolate high quality RNA for gene expression analyses in developing seeds of lentils (*Lens culinaris* Medik.)
b) To isolate and characterize from *Lens culinaris* the following three RFO biosynthetic genes:
    i. galactinol synthase
    ii. raffinose synthase
    iii. stachyose synthase
c) To analyze the expression of the three genes during lentil seed development.

2. LITERATURE REVIEW

2.1 Introduction

2.1.1 Classification of legumes

Pulse crops, which belong to the family Fabaceae, are defined by the presence of unusual flowers, podded fruits and their ability to fix nitrogen in their root nodules (de Faria et al 1989). The family Fabacaeae is further divided into three subfamilies: Papilionoideae, Caesalpinioideae and Mimosoideae. The three subfamilies show distinct characteristics in their flowers: Papilionoideae has two partially fused petals, two wing petals and a banner like petal; Caesalpinioideae has irregular flowers with no distinct petals; and Mimisoideae is characterized by the presence of spikes. Most legume crops belong to the subfamily Papilionoideae.

Pulse crops cultivated for human and animal consumption include field pea (*Pisum sativum* L.), common bean (*Phaseolus vulgaris* L.), chickpea (*Cicer arietinum* L.), broad bean (*Vicia faba* L.), pigeon pea (*Cajanus cajan* L.), cowpea (*Vigna unguiculata* (L.) Walp.), and lentil (*Lens culinaris* Medik.). Two other legume crops grown primarily for oil production include soybean (*Glycine max* L.) and peanut (*Arachis hypogaea* L.).

2.1.2 Lentil origin and morphology

Lentil, one of the founder crops (emmer wheat, einkorn wheat, hulled barley, lentil, chickpea, bitter vetch and flax), has been cultivated since the Neolithic times in the Near East (Zohary and Hopf 1973). Archeological evidence suggests lentil cultivation started as early as the seventh millennium BC in the Near East (Zohary 1972; Zohary and Hopf 1973).

`Lens` a Latin word, indicates a disc shaped object, whereas, culinaris describes its culinary use (Robertson and Erskine 1997; Sarker and Erskine 2006). The taxonomic classification of lentils is Plantae (Kingdom), Magnoliophyta (Division), Magnoliopsida (Class), Fabales (Order), Fabaceae (Family), Papilionoideae (Subfamily), Lens (Genus). Cytoplasmic studies indicated that *Lens orientalis* was the progenitor of domesticated *Lens culinaris* (Muehlbauer et al 2006) with both having the same number of chromosomes (2n = 14). Lentil is a branched softly pubescent light green herbaceous plant (Figure 2.1; Erskine et al 1990). It is a self pollinating, indeterminate crop with soil pH requirements of 6.0 to 8.0. Lentil flowers vary in color and can be white, pink, purple, light purplish-blue or pale-blue (Muehlbauer et al 1985). Lentils were traditionally grouped into two classes based on seed sizes, (i) Microsperma or

small-seeded, and (ii) Macrosperma or large-seeded. Small-seeded types range from 3 - 6 mm in diameter whereas, large-seeded range from 6 - 9 mm in diameter. Lentil testa color can be yellow, green, brown, black, pink and grey. Lentil can be classified into three types based on cotyledon color; orange (also known as red), yellow and green (Saxena et al 2009).

2.1.3 Lentil utilization

2.1.3.1 Crop rotation

Legume crops including lentil have been used for crop rotation purposes. The nitrogen fixation ability of legume crops enrich soil nitrogen thereby reducing the need for addition of nitrogen fertilizer to soil. Nitrogen fixation by legume crops is facilitated by the presence of symbiotic micro-organisms species of *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium*, *Azorhizobium*, *Mesorhizobium* and *Allorhizobium* (Graham and Vance 2000). The microbes invade the root nodules of leguminous plants and fix the atmospheric nitrogen into ammonia, thus providing readily assimilable nitrogen. After the decomposition of plants, nitrogen is released from the roots and fixed in the soil (Graham and Vance 2003).

2.1.3.2 Human consumption

Lentils are widely consumed in Asian, Near East, and Mediterranean countries. Lentils cook quickly compared to other pulses making them more desirable for human consumption. Lentil consumption has been promoted widely and is recommended as a part of balanced diet (Leterme 2002). The nutritional profile of lentils makes them a good substitute of animal protein and can also be consumed by gluten sensitive persons. It has been reported that the inclusion of lentil in the diet reduced the risk of non-communicable chronic disease including, coronary heart disease, diabetes and obesity (Anderson and Major 2002; Ramdath and Tsao 2012). The beneficial health effect of lentil is attributed to its low glycemic index (GI) (Jenkins et al 1981). The glycemic index is defined as the *incremental area* under the *blood glucose* response curve of a *50g carbohydrate portion* of a test food expressed as percent of the response to the same amount of carbohydrate from a *standard food* taken by the *same subject"* (FAO/WHO 1998). Non-starch polysaccharides, resistant starch and non-sucrose oligosaccharides present in lentils contribute to its low glycemic index (Sievenpiper et al 2009; Chibbar et al 2010).

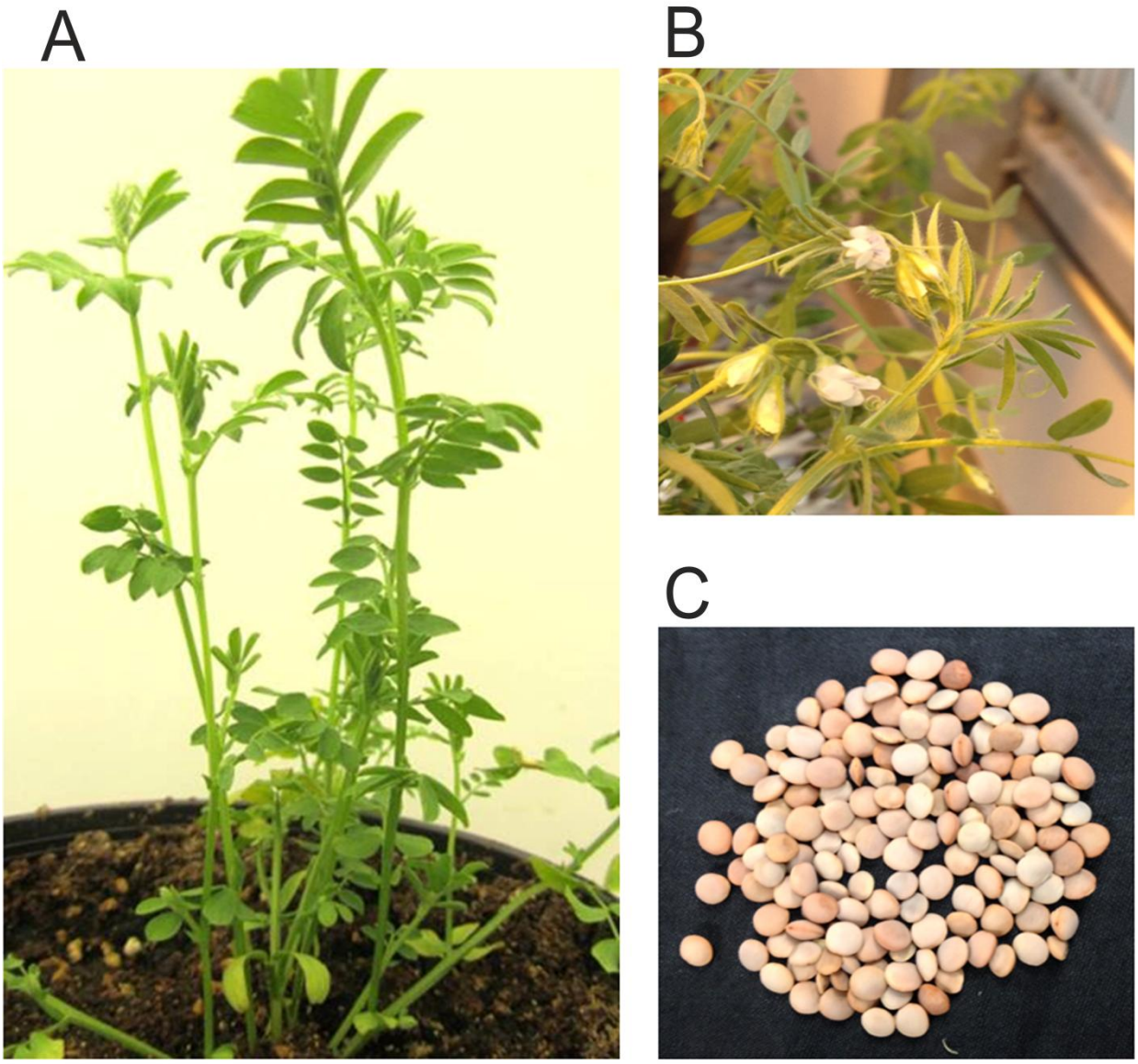


Figure 2.1: Morphology of *Lens culinaris* Medik. (A) plant (B) flowers (C) seeds

2.2 Lentil production and export

Canada, India, Turkey, Australia, USA, Syria and Nepal are the major producers of lentil. Canada is the largest producer and exporter of lentil (FAOSTAT 2014) cultivated for economic returns and to diversify and lengthen crop rotations. In the crop year 2013 in Canada, lentils were harvested from more than a million hectares and 1,750,000 tonnes of lentils were exported (Statistics Canada 2014). Production of lentil in Saskatchewan dates back to 1970, whereas it was introduced to Manitoba and Alberta in the 1980s (Bhatty 1988). The major share of lentil (approximately 95 %) produced in Canada is grown in Saskatchewan (Figure 2.2) where it is more suited to Saskatchewan's Brown and Dark Brown soils. Lentil production in Saskatchewan has increased significantly from 2000 hectares in 1970 to more than a million hectares in 2014 (Statistics Canada 2014). The two major market classes of lentils are the red and green types. They are further classified into large, medium and small seeds. Saskatchewan is the major producer of small red and large green lentils, the two largest market classes.

2.3 Nutritional composition of lentils

2.3.1 Carbohydrates

Carbohydrates can be classified into monosaccharides, disaccharides, oligosaccharides and polysaccharides based on their polymeric nature (BeMiller 2001). Carbohydrates can also be classified into available and unavailable based on nutritional properties. Carbohydrates that are easily broken down and the resultant glucose molecules absorbed by humans and animals are referred to as available carbohydrates. On the other hand, carbohydrates that are not absorbed in the digestive tract but are fermented by microbial flora are referred to as unavailable (Chibbar et al 2010). In lentil seeds sucrose is the major disaccharide with concentrations ranging from 1.1 - 3.0 g/100g dry matter (DM) (Table 2.1). Starch is the major polysaccharide with concentrations ranging from 34.7 - 65.0 g/100g DM. The amylose concentration of the starch fraction ranged from 26.1 - 34.0 g/100g of starch. An increase in starch concentration (2.3 - 9.7 g/100g DM) was observed after decortication of lentil seeds (Tahir et al 2012).

"Dietary fiber is the edible parts of plants or analogous carbohydrates that are resistant to digestion and absorption in the human small intestine with complete or partial fermentation in the large intestine. Dietary fiber includes polysaccharides, oligosaccharides, lignin, and associated plant substances. Dietary fiber promotes beneficial physiological effects including

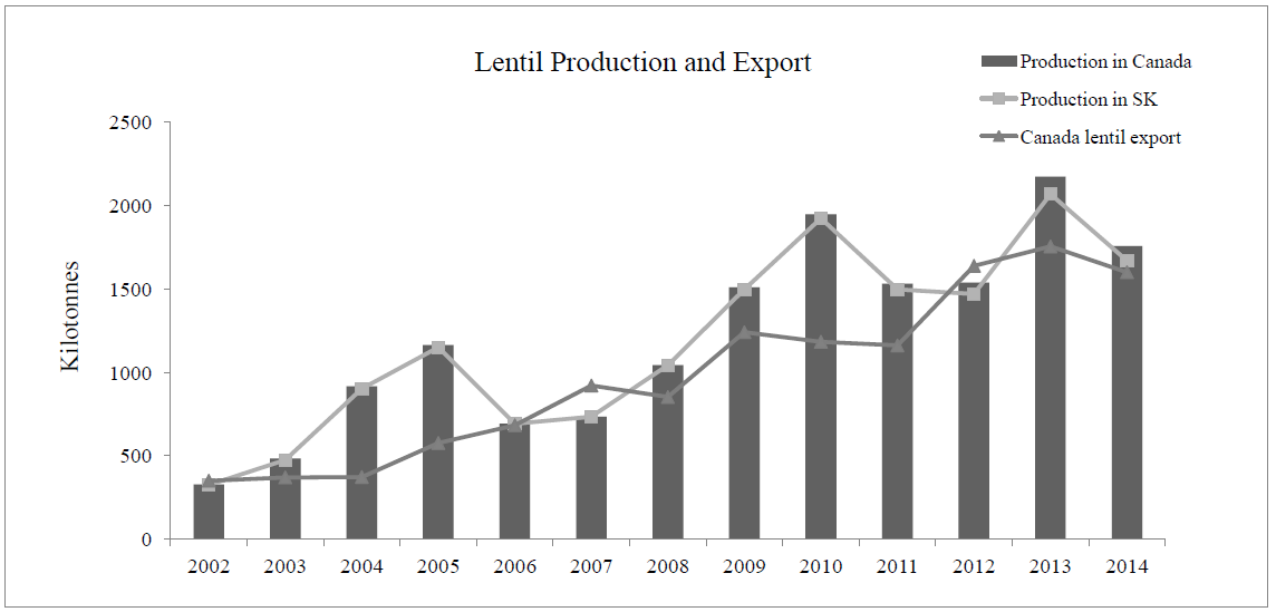


Figure 2.2: Canadian lentil production and export (Source: Statistics Canada 2014).

Table 2.1: Carbohydrate composition of lentils

Type of saccharides	Physiological classification	Site of digestion	Component	Concentration in Lentils (g / 100 g DM)
Mono- di saccharides	Available carbohydrates	Small intestine (enzymatic)	Total soluble sugars	2.3 - 8.9
			Glucose	ND - 0.04
			Fructose	0.01 - 0.30
			Sucrose	1.1 - 3.0
			Maltose	0.05 - 0.33
Starch	Available carbohydrates	Small intestine (enzymatic)	Starch	34.7 - 65.0
			Amylose	26.1 - 34.0[a]
α- galactosides	Unavailable carbohydrates	Large intestine (Microbial)	Total α-galactosides	1.8 - 6.8
			Raffinose	0.16 - 1.49
			Stachyose	1.1 - 3.1
			Verbascose	ND - 1.35
			Ciceritol	0.24 - 1.99
Storage polysaccharides	Unavailable carbohydrates	Large intestine (Microbial)	Cellulose	4.1 - 5.33
			Hemicellulose	6.0 - 15.74

[a]- g/100 g starch; ND - not detected.

Values derived from: Reddy et al 1984; Vidal - Valverde and Frias 1992; Vidal-Valverde et al 1993a, b; Frias et al 1994, 1995, 1996; Jood et al 1998; Sotomayer et al 1999; Aranda et al 2001; Fasina et al 2001; Cai et al 2002; Porres et al 2002; El-Adway et al 2003; Wang and Daun 2004, 2006; Costa et al 2006; Urbano et al 2007, Grusak et al 2009; Tahir et al 2011, 2012; Wang et al 2014.

laxation, and/or blood cholesterol attenuation, and/or blood glucose attenuation" (DeVries 2003). In eight Canadian lentil cultivars the concentration of total dietary fiber varied from 131 to 147 g/kg while the water-soluble dietary fiber and water-insoluble dietary fiber ranged from 15.2 - 19.4 g/kg and 114 - 128 g/kg, respectively (Wang et al 2009a). Many of the beneficial health effects of pulses have been attributed to dietary fiber including, constipation relief, and the lowering of blood cholesterol and glucose levels in humans (DeVries 2003).

2.3.2 Protein composition

Lentils have high protein concentration when compared with cereals, and are rich in essential amino acids including lysine, but deficient in sulphur containing amino acids. Protein concentrations are in the range of 15.9 to 31.4 g/100g DM (Grusak 2009). Essential amino acids present in lentils are histidine (1.3 - 4), isoleucine (2.6 - 9.6), leucine (5.7 - 15.9), lysine (4.0 - 12.6), threonine (2.5 - 7.6), tryptophan (ND - 2.6), valine (3.3 - 11.6), phenylalanine (3.6 - 10.6), methionine (0.8 - 1.2) and cysteine (required by infants/children; "conditionally essential") (0.8 - 1.6), where all ranges are in g/16g N (Urbano et al 1995; Wang and Duan 2006; Urbano et al 2007; Grusak 2009; Boye et al 2010). Most essential amino acids are present in concentrations higher than those recommended by the FAO except tryptophan and sulphur containing amino acids methionine and cysteine. Therefore, the consumption of lentils with cereals contributes to a balanced diet, as cereals are deficient in lysine, but have adequate amounts of methionine and cysteine.

2.3.3 Vitamins and minerals

Vitamins can be classified into water soluble and fat soluble. Based on the maximum recommended dietary allowance (RDA) provided by the USDA (United States Department of Agriculture) lentils are good source of folate/vitamin B9 (Table 2.2). Lentils are also good source of thiamin (vitamin B1), pantothenic acid, and pyridoxine (vitamin B_6) contributing 28, 26 and 21% of adult RDA, respectively (Grusak 2009). Lentils are low in vitamin C while vitamin B_{12} and vitamin D are not found in plant tissues (Table 2.2). Minerals are a part of the ash fraction and can be classified as macronutrients (K, P, Ca, Mg, Na), micronutrients (Fe, Zn, Cu, Mn) and trace elements (Al, Cr, Ni, Pb, Co, Se, Mo) (Urbano et al 2007; Grusak 2009). Based on RDA and mineral composition, lentils are a good source of copper and phosphorous (Table 2.2).

Mineral availability in lentils is reduced due to the presence of antinutritional factors such as phytates and RFO. These antinutritional factors interfere with nutrient absorption in humans and monogastric animals (Martínez-Villeluanga et al 2008a).

2.4 Antinutritional factors

The antinutritional factors of pulses can be classified into compounds of either protein or non-protein origin. Compounds of protein origin include protease inhibitors and heamagglutinins. Antinutritional compounds of non-protein origin include raffinose family oligosaccharides, phytic acid, tannins and alkaloids (Gulewicz 2014). Compounds of protein origin are sensitive to heat and can be removed by cooking of lentil seeds. In Canadian lentil varieties, the concentration of trypsin inhibitors was reported to be in the range of 1.91 to 2.77 mg/g, but could be reduced by dehulling and cooking (Wang et al 2009a). However, cooking was recognized as an effective method to reduce these compounds in seeds (Wang et al 2009a). For Canadian lentil cultivars, phytic acid concentrations range from 6.2 to 8.8 g/kg DM and tannin concentrations ranged from 3.4 to 6.1 g/kg DM (Wang et al 2009a). RFO, which include raffinose, stachyose and verbascose were reported to be in the range of 1.95 - 2.22, 2.20 - 2.55 and 1.15 - 1.33 g/100 g, respectively, in selected Canadian lentil varieties (Tahir et al 2011).

2.4.1 Raffinose family oligosaccharides (RFO)

Alpha-galactosides or raffinose family oligosaccharides (RFO) are one of the factors contributing to reduced lentil consumption in the Western world . Humans and monogastric animals lack the enzyme α-galactosidase, that is required to break the α-(1→6) linkages which are characteristic to RFO (Figure 2.3). Bacteria present in the lower bowel ferment RFO producing large amounts of carbon dioxide, hydrogen and methane gases resulting in flatulence in humans and animals (Granito et al 2005). Though RFO produce stomach discomfort and flatulence, they may have potential beneficial effects as prebiotics as they have been shown to promote the growth of bifidobacterium and lactobacillus species in the large intestine (Roberfroid 1998, 1999; Guillon 2002; Martínez-Villaluenga et al 2008b). RFO are essential for plant growth and also play an important role in abiotic stress tolerance. The key lies in the reduction of RFO to an optimal amount so as to reduce stomach discomfort while still providing

Table 2.2: Vitamin and mineral composition of lentil seeds

	Concentration (amount in 100g DM)	Maximum adult RDA
Water Soluble Vitamins		
Biotin	ND - 132[a]	
Folate	40 - 535[a]	400[a]
Pantothenic Acid	0.4 - 2.4[b]	
Vitamin B_1 (Thiamin)	0.13 - 0.90[b]	1.2[b]
Vitamin B_{12}	not found in plants	2.4[a]
Vitamin B_2 (Riboflavin)	0.11 - 0.46[b]	1.3[b]
Vitamin B_3 (Niacin)	0.6 - 3.6[b]	16[b]
Vitamin B_6 (Pyridoxine)	0.16 - 0.60[b]	1.7[b]
Vitamin C	0.0 - 7.7[b]	90[b]
Fat Soluble Vitamins		
Choline	109[b]	
Vitamin A	2.2 - 3.4[a]	900[a]
Vitamin D	Not found in plants	
Vitamin E (α-tocopherol)	0.34 - 1.60[b]	
Vitamin E (γ-tocopherol)	0.31- 0.64[b]	15[b]
Vitamin K	5.6[a]	120[a]
Minerals		
Calcium (Ca)	42 - 165[b]	1200[b]
Copper (Cu)	0.4 - 9.9[b]	0.9[b]
Iron (Fe)	3.1 - 13.3[b]	18[b]
Magnesium (Mg)	13 - 167[b]	420[b]
Manganese (Mn)	0.6 - 1.0[b]	2.3[b]
Phosphorous (P)	240 - 1287[b]	700[b]
Potassium (K)	38 - 1360[b]	4700[b]
Zinc (Zn)	2.3 - 10.2[b]	11[b]
Selenium	0.009 - 0.012[b]	0.055[b]
Sodium	0.4 - 79[b]	
Boron	0.6 - 1.0[b]	
Chromium	0.03[b]	
Nickel	0.12 - 0.35[b]	
Cobalt	0.04[b]	
Molybdenum (Mo)	0.08 - 0.22[b]	

Reproduced from Grusak et al 2009 with modifications. Concentrations are represented as [a] µg and [b] as mg.

RDA - Recommended dietary allowance; ND - not detected.

adequate amounts required for seed germination and plant growth. Research in this thesis mainly focuses on the genes participating in RFO biosynthesis during lentil seed development so as to improve our understanding of the biosynthetic pathway. These results can be used to develop judicious strategies to reduce RFO concentration in lentil seeds.

2.4.2 Structure of RFO

Alpha-galactosides are second in abundance to sucrose in plant cells. RFO, a group of α-galactosides derivatives of sucrose, are ubiquitous in plant seeds (Blöchl et al 2007), accumulate during seed maturation (Peterbauer et al 2001a) and degrade rapidly during germination events.

RFO are made up of raffinose, stachyose, verbascose and ajugose. Ciceritol, a low abundant RFO was first purified and found in chickpea and hence the name ciceritol (Quemener and Brillouet 1983). The RFO nomenclature is as follows:

raffinose [α-D-galactopyranosyl-(1→6)-α-D-glucopyranosyl-(1→2)-β-D-fructofuranoside],

stachyose [α-D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→6)-α-D-glucopyranosyl-(1→2)-β-D- fructofuranoside],

verbascose [α-D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→6)-α-D-glucopyranosyl-(1→2)-β-D- fructofuranoside],

ajugose [α-D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→6)α-D-glucopyranosyl-(1→2)-β-D- fructofuranoside]

ciceritol [α- D-galactopyranosyl-(1→6)-α-D-galactopyranosyl-(1→2)-1D-4-O-methyl-chiro- inositol]

Improvement in carbohydrate analysis methodology with more sensitive techniques which include high performance anion exchange chromatography with pulse amperometric detection (HPAEC-PAD) has improved RFO analysis in lentils over previously used chromatography techniques (Gangola et al 2014). This analytical technique has also led to the identification of new compounds in lentils like manninotriose (dos Santos et al 2013). Analysis of carbohydrates from more than 100 plants using HPAEC led to the identification of manninotriose from *Lamium purpureum* and structure was later confirmed by NMR spectroscopy. Manninotriose is stachyose devoid of the fructose residue on sucrose, similar to that observed for melibiose, a raffinose compound devoid of fructose on the sucrose moiety

(Figure 2.3; dos Santos et al 2013). Manninotriose and melibiose were proposed to be synthesized by invertase or β-fructosidase hydrolysis of raffinose and stachyose, respectively (dos Santos et al 2013).

2.5 RFO biosynthetic pathway

2.5.1 RFO biosynthesis in plant cells

Raffinose family oligosaccharides are formed by α-(1→6) galactoside linkages between the linear chain galactosyl residues to the glucose moiety of sucrose (Avigad and Dey 1997) (Figure 2.3). The three galactosyl donors involved in RFO biosynthesis are, UDP-D-galactose, galactinol and RFO themselves. The biosynthesis of RFO is initiated by galactinol synthase (EC 2.4.1.123) which catalyses the transfer of a galactosyl residue from UDP-D-galactose to *myo*-inositol to synthesize galactinol (Figure 2.4). Raffinose synthase (EC 2.4.1.82) catalyzes the synthesis of raffinose by the transfer of a galactosyl residue from galactinol to sucrose. Stachyose synthase (EC 2.4.1.67) catalyzes the synthesis of stachyose by the transfer of a galactosyl residue from galactinol to raffinose. The enzyme verbascose synthase catalyzes the synthesis of verbascose by addition of a galactosyl residue from galactinol to stachyose. The main enzymes involved in the RFO pathway, galactinol synthase, raffinose synthase and stachyose synthase have been isolated from some plants (Tables 2.3, 2.4 and 2.5) and the gene sequences coding for these enzymes have been deposited in NCBI or patented (Allan and Hitz 2000; Oosumi et al 1998). RFO are also synthesized via a galactinol independent biosynthetic pathway. In *Ajuga reptans* the enzyme galactan:galactan galactosyl transferase (GGT) that catalyzes chain elongation by galactosyl transfer between two RFO molecules, has been reported (Bachmann et al 1994).

In *Ajuga reptans* two different RFO pools were reported: (i) storage pool-RFO synthesized in mesophyll cells, and (ii) transport pool-RFO synthesized in intermediary cells involved in phloem transport (Bachmann et al 1994). Further compartmentalization studies by purification of vacuoles from mesophyll cells indicated that GGT, stachyose and higher RFO (verbascose) were vacuolar; and GS, STS, *myo*-inositol, galactinol, sucrose and fructose were extravacuolar. Raffinose was reported to be distributed in both the vacuole and the cytoplasm (Bachmann and Keller 1995). Stachyose synthesized in the cytoplasm was proposed to be transferred through a stachyose transporter in the tonoplast (Bachmann and Keller 1995).

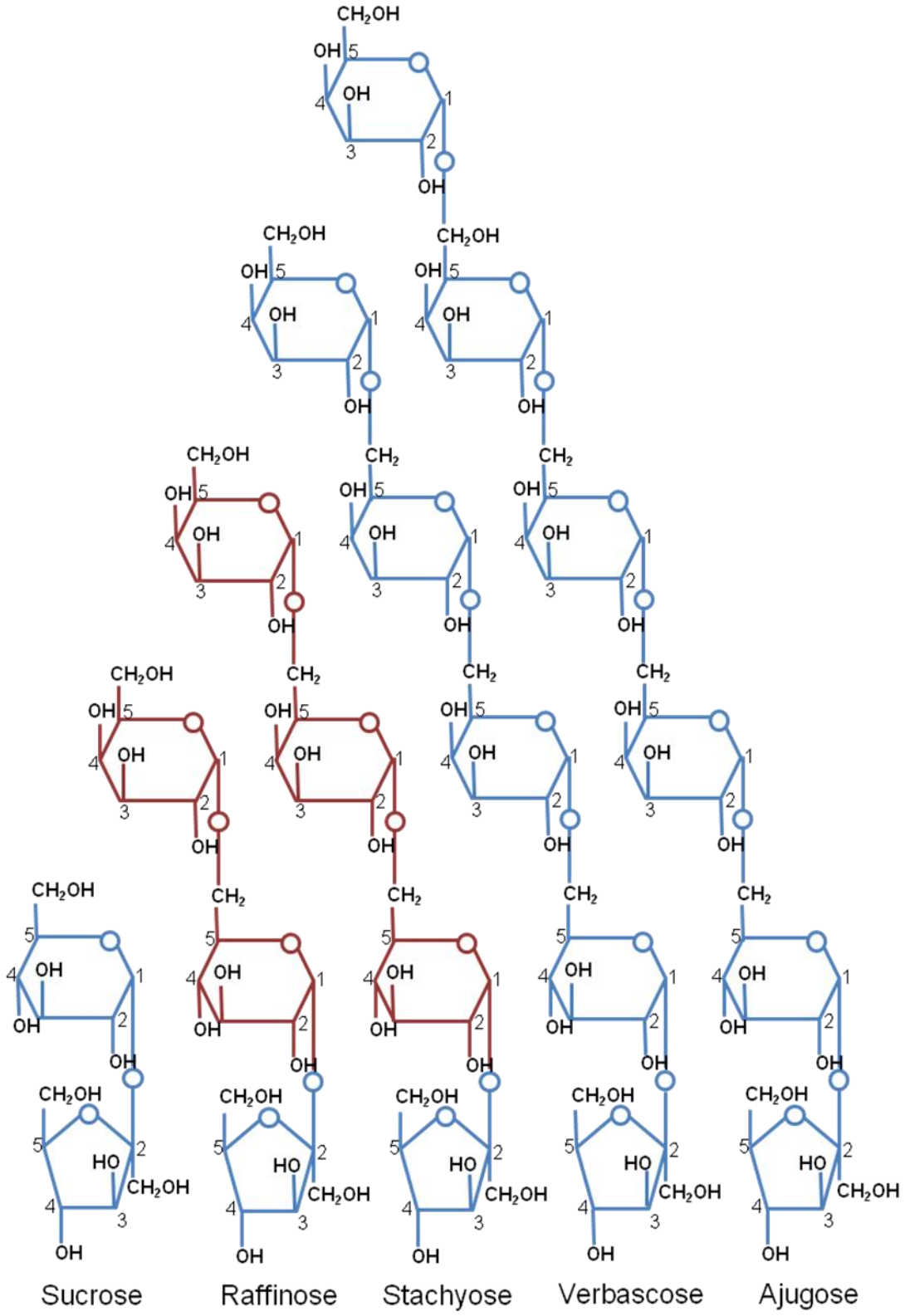


Figure 2.3: Chemical structure of sucrose, raffinose, stachyose, verbascose and ajugose. The brown colored regions of raffinose and stachyose indicate melibiose and manninotriose, respectively.

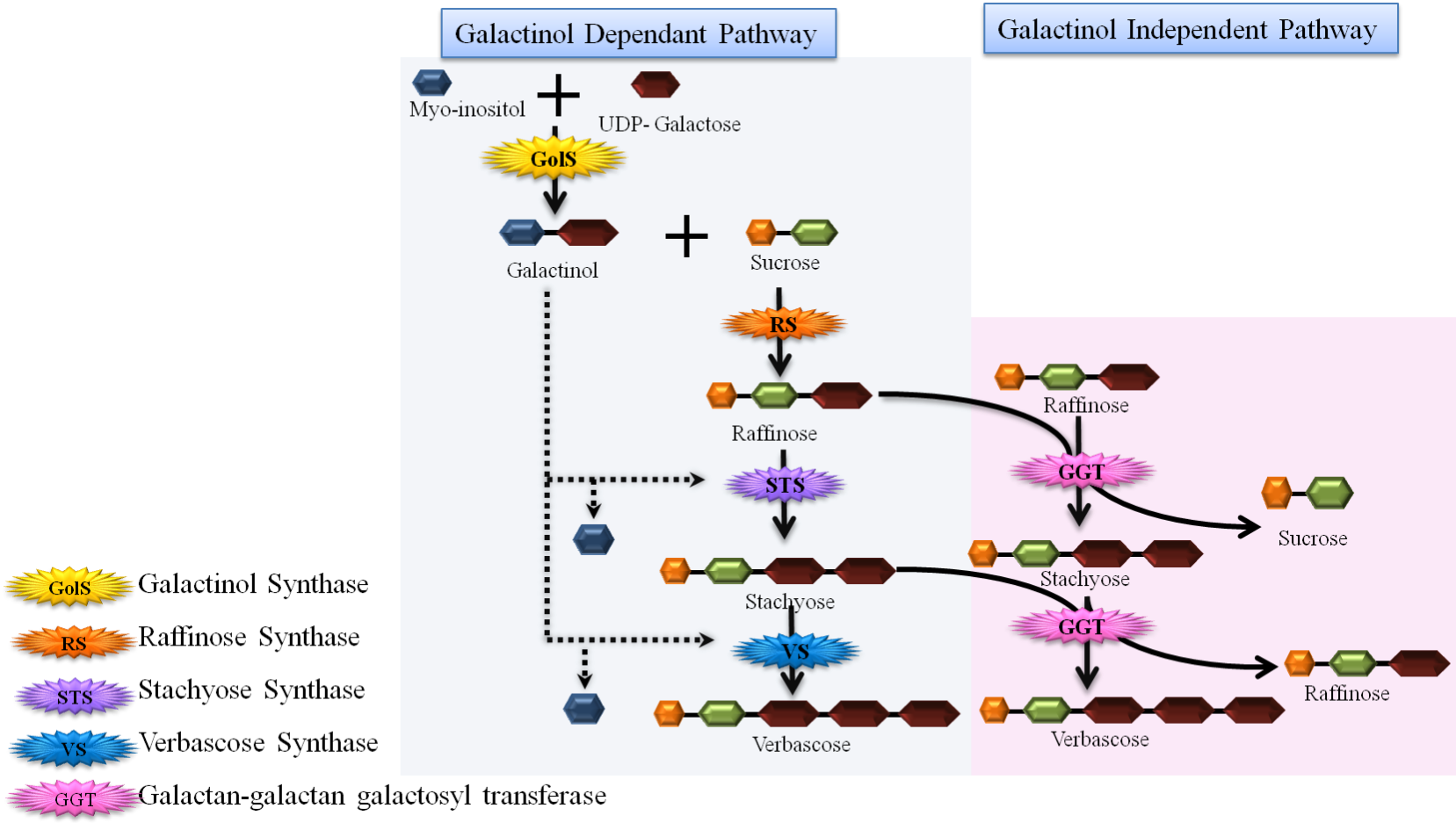


Figure 2.4: Schematic representation of RFO biosynthetic pathway in plants

Two allelic variants of *GolS* were isolated in *Ajuga reptans* (Sprenger and Keller 2000). Gene expression, RFO accumulation and GS activity suggested functional differences among the two isoforms. *ArGolS1* was predominantly present in the storage RFO pool in mesophyll cells and *ArGolS2* was predominant in transport RFO pools in intermediary cells (Sprenger and Keller 2000).

2.5.2 *myo*-inositol phosphate synthase

Myo-inositol is the main galactosyl donor for the synthesis of galactinol, a precursor for RFO synthesis. *Myo*-inositol phosphate synthase (MIPS; EC 5.5.1.4) catalyzes the conversion of glucose-6-phosphate to 1-D-*myo*-inositol-3-phosphate. Conversion of 1-D-*myo*-inositol-3-phosphate to *myo*-inositol is catalyzed by the enzyme *myo*-inositol-phosphate monophosphatase (IMP; EC 3.1.3.25) (Obendorf et al 2009). A low raffinose, stachyose and phytic acid (LRSP) phenotype was associated with a mutant *Mips*) in soybean (Hitz et al 2002). In soybean, four *Mips* genes (*Mips*1, *Mips*2, *Mips*3, *Mips*4) were identified, among which *Mips1* was highly expressed in immature seed and *Mips4* showed high expression in leaves (Hegeman et al 2001; Chapell et al 2006; Nunes et al 2006; Chiera and Grabau 2007). A mutation in the *Mips*1 gene that changed amino acid residue 396 from lysine to asparagine caused a 90% reduction in the seed specific MIPS activity (Hitz et al 2002) resulting in the LRSP phenotype. However, these lines showed reduced field emergence and were sensitive to imbibitional chilling (Meis et al 2003; Obendorf et al 2008).

2.5.3 Galactinol synthase (GS, *GolS*)

In *Cucurbitaeceae* leaves GS is a monomeric polypeptide of 38-43 kDa with 318-348 amino acid residues (Table 2.3). GS enzyme activity has a pH optima between 5.6 and 7.5 and a Mn^{2+} requirement. The K_m values for UDP-D-galactose and *myo*-inositol range from 0.16 – 0.53 mM and 4.0 – 6.5 mM, respectively (Keller and Pharr 1996; Peterbauer et al 2001b). Most of the studies on GS suggest that RFO accumulation is controlled by the levels of the initial substrates, *myo*-inositol and sucrose, rather than by galactinol synthase activity (Karner et al 2004).

The presence of more than a single isoform of *GolS* showing differential expression

Table 2.3: Galactinol synthase genes isolated from different plant species

Plant	Gene	Tissue	Amino acid	Accession No.	Role of Galactinol synthase	Reference
Arabidopsis thaliana	AtGols1		344	O22893	Isolated from database based on *Oryza sativa* sequence Present in mature seeds, induced during water and salt stress	Taji et al 2002
	AtGols2	Leaf	335	Q9FXB2	Induced by salt and drought stress and not by cold stress	
	AtGols3		334	O80518	Present in mature seeds, induced expression during cold stress and not by salt or drought stress	
Ajuga reptans	GolS-1	Leaf	333	CAB51533.1	Primarily expressed in the mesophyll, storage	Sprenger and Keller 2000
	GolS-2		292	CAB51534.1	Expressed in the phloem-associated intermediary cell-transport	
Boea hygrometrica	BhGolS1	Leaf	334	ACI62176.1	Expression induced by dehydration and induced by abscisic acid	Wang et al 2009b
Brassica napa	Galactinol Synthase		341	AAD26116.1	Expressed in leaf and developing seeds	Bock et al 2009
Coffea arabica	CaGolS1	Leaf	338	ADM92588.1	Expressed in normal conditions and most responsive during stress	Dos Santos et al 2011
	CaGolS2	Leaf, endosperm	334	ADM92589.1	Detected only under severe water deficit and salt stress	
	CaGolS3	Leaf, flowers, roots	344	ADM92590.1	Expressed under moderate and severe drought	
Cucumis melo	GolS1	Leaves	331	AAL78687.1	Expressed in mature leaves and developing seeds	Volk et al 2003
	GolS2		332	AAL78686.1	Expressed in mature leaves	

Table 2.3 Continued: Galactinol synthase genes in different plant species

Plant	Gene	Tissue	Amino acid	Accession No.	Role of Galactinol Synthase	Reference
Cucurbita pepo		Leaf	41	Q6U2M6	Partial sequence	Liu et al 1995
Fagopyrum esculentum	FeGolS-1	Developing seeds	333	AAM96870.1	Catalyse the synthesis of fagopyritol A1	Ueda et al 2005
	FeGolS-2		354	AAM96868.1	Catalyze synthesis of fagopyritol B1	
	FeGolS-3		255	AAM96869.1	Partial sequence	
Lycopersicon esculentum	LeGolS-1	Seeds	318	AF447452	Expressed in mature seeds, Induced by cold and dehydration in leaves	Downie et al 2003
Medicago sativa	MsGolS	Crown	325	Q84MZ5	Induction in roots during cold stress	Cunningham et al 2003
Pisum sativum	PsGolS1	Seeds	334	CAB51130.1	Expression observed in developing seeds	Peterbauer et al 2001b
Populus (*Populus alba x grandidentata*)	Pa3gGolSI	Leaf	337	AEN74905.1	Expression regulated by temperature. May be involved in seasonal regulation of carbohydrates	Unda et al 2012
	Pa3gGolSII	Leaf	337	AEN74906.1	Expressed in all tissues. May play a role in RFO storage	
Xerophyta viscosa	XvGolS	Leaf	337	A0MNW8	High expression during water stress	Peters et al 2007
Zea mays	ZmGolS1	Seeds	344	AF497507	Expression not detected in tissues	Zhao et al 2004
	ZmGolS2		348	AF497508	Expressed when seed germination was interrupted by desiccation	
	ZmGolS3		345	AF497509	Expressed during seed development	

during abiotic stresses has been reported in several plant species. Three genes coding for GS were characterized in *Arabidopsis thaliana* as, *AtGolS1*, *AtGolS2* and *AtGolS3*. Recombinant *AtGolS1*, *AtGolS2* and *AtGolS3* proteins expressed in *E.coli* all showed GS activity. Differential expressions of GS genes were obtained during abiotic stresses, where, *AtGolS1* and *AtGolS2* were induced during drought and salinity stress, but not by cold stress. The third isoform, *AtGolS3* was induced by cold stress but not by drought and salinity stress (Taji et al 2002).

In *Zea mays* three *GolS* isoforms were isolated, but they showed differential gene expression. *ZmGolS1* expression was not observed during seed development. Transcript accumulation of *ZmGolS2* was observed towards later stages of seed development and rapid decrease of transcripts was observed at imbibition during seed germination. *ZmGolS3* transcripts were only detected during seed germination when interrupted by desiccation (Zhao et al 2004). In *Coffea arabica* three allelic variants, *CaGolS1*, *CaGolS2* and *CaGolS3* that coded for polypeptides with 388, 334, 344 amino acids, also showed differential transcript accumulation under drought, salinity and heat stress conditions. *CaGolS1* showed high expression during stressed and non-stressed conditions, *CaGolS2* was expressed only during severe water deficit and *CaGolS3* was expressed during all experimental stresses but at reduced level than *CaGolS1* (dos Santos et al 2011). In tomato (*Lycopersicon esculentum*) *LeGolS1* transcripts were detected 35 days after anthesis until seed maturity was achieved (60 days after anthesis) (Downie et al 2003).

Two GS isoforms in *Ajuga reptans*, *ArGolS1* and *ArGolS2* have been characterized. *ArGolS1* is source-leaf specific and *ArGolS2* participated in RFO transport. Gene expression studies showed that *ArGolS1* transcripts were found in mesophyll and *ArGolS2* in intermediary cells explaining its role in phloem transport (Sprenger and Keller 2000).

2.5.4 Raffinose synthase (RS)

The purification of raffinose synthase (RS) was first reported by Lehle and Tanner (1973) from *Vicia faba*. This purified RS had a molecular mass of 90 kDa and exhibited a pH optimum between 6.5 and 7.0. In a subsequent study in pea (*Pisum sativum*) RS was partially purified, which showed a pH optimum of 7.0, and Km values of 7.3 mM and 22.9 mM for galactinol and sucrose, respectively (Peterbauer et al 2002a). Expression of a RS cDNA clone in *Spodoptera frugiperda* Sf21 insect cells, produced recombinant RS, and the kinetic properties of the

recombinant RS were found to be similar to those of the purified RS (Peterbauer et al 2002a). A RS cDNA clone isolated from rice (*Oryza sativa*; Table 2.4) was expressed in *E. coli* to produce recombinant RS. The rice recombinant RS also showed maximum activity at pH 7.0 at 45 ^{o}C (Li et al 2007).

In Arabidopsis, five putative RS genes (*AtRS1-5*) or seed imbibition proteins (SIP) have been described (Nishizawa et al 2008). Among the five *AtRS* genes described, *AtRS5* showed high sequence similarity to the RS characterized in *Pisum sativum.* Heterologous expression of recombinant *AtRS5* showed RS activity (Egbert et al 2013). Further, reduced levels of raffinose were observed in *AtRS5* mutant seeds in Arabidopsis. No *AtRS5* expression or activity was detected in leaves in mutant plants under unstressed or stressed conditions (Egbert et al 2013). *RS2*/*ATSIP2* showed sequence similarity to α-galactosidase genes. Recombinant protein of *ATSIP2* expressed in Sf9 insect system showed raffinose specific α-galactosidase activity (Peters et al 2010). RS was found to be the most unstable enzyme in the RFO biosynthetic pathway (Castillo et al 1990; Peterbauer et al 2002a).

2.5.6 Stachyose synthase (STS)

Stachyose synthase (STS) has been purified from adzuki bean, kidney bean, lentil and pea (Tanner and Kandler 1968; Peterbauer and Richter 1998; Hoch et al 1999; Peterbauer et al 2002b). STS purified from mature lentil seeds had a specific activity of 9.09 pkat/mg protein, a molecular mass of 88.6 kDa and an isoelectric point of 4.8 (Hoch et al 1999). The amino acid sequence of STS (853-868 amino acids; Table 2.5) was first obtained from *Vigna angularis* (Peterbauer et al 1999). The molecular weight of STS was 85 to 95 kDa, with an enzyme activity pH optimum of 6.5 - 7.0 (Richter at al 2000). STS shows a broad range of substrate specificity which includes inositols and inositol O-methyl ethers. STS from adzuki bean shows no conversion of pinitols, whereas lentil STS catalyzed the synthesis of galactopinitol A and ciceritol, in addition to stachyose synthesis (Hoch et al 1999; Peterbauer et al 2001a). STS purified from adzuki bean and lentil showed no synthesis of verbascose (Peterbauer and Richter 1998; Hoch et al 1999).

In *A. reptans,* RFO synthesis also occurred through a non-galactinol independent enzyme galactan:galactan galactosyl transferase (GGT) present in leaf vacuoles (Bachmann and Keller

Table 2.4: Raffinose synthase genes isolated from plant species

Plant	Gene	Tissue	Amino Acid	Accession number	Role of Raffinose synthase	Reference
Arabidopsis thaliana	AtRS	Leaf	783	At5g40390; NP_198855.1	Expressed in leaf	Zuther et al 2004, Egbert et al 2013
Cucumis sativus	CsRS	Stem	784	Q9ZT6	Expressed in roots, fruits, and stems. Induced by low temperature and abscisic acid (ABA)	Sui et al 2012
Glycine max	GmRS2		781	ACD13461.1	Expressed in all tissues	Dierking and Bilyeu 2008
	GmRS3		758	ACD13462.1		
	GmRS1			E25448		
Oryza sativa	RS	Rice seedlings	783	Q5VQG4-1	Recombinant enzyme expressed in *E. coli*	Li et al 2007
Pisum sativum	PsRS	Developing seeds	798	Q8VWN6-1	Expressed in Spodoptera frugiperda Sf21 insect cells	Peterbauer et al 2002a
Vicia faba	RS	Seeds			Enzyme purified	Lehle and Tanner 1973

Table 2.5: Stachyose synthase genes isolated from different plant species

Plant	Gene	Tissue	Amino Acid	Accession number	Role of Galactinol synthase	Reference
Alonsoa meridionalis	AmSTS1	Leaves	868	Q8L5R0	Expressed in intermediary cell but not in companion cell	Voitsekhovskaja et al 2009
Arabidopsis thaliana	AtSTS		876	NP_192106.3		GenBank
Glycine max		Seed	437	D2E9R7-1	Partial sequence	GenBank
Pisum sativum	STS1	Seed	853	CAC38094.1	Expressed in developing seeds	Peterbauer et al 2002, 2003
	STS1-2	Seed		CAD55555.1	Expressed in developing seeds. Catalyzes synthesis of verbascose	
Vigna angularis	VaSTS1	Seed	857	CAB64363.1	Expressed in developing seeds	Peterbauer et al 1999

1995; Haab and Keller 2002). GGT sequence showed high similarity to α-galactosidases and a non-sequence specific vacuolar sorting determinant at the C-terminal of the sequence (Haab and Keller 2002; Tapernoux-Luthi et al 2007). The presence of GGT activity (neutral pH) was reported in pea seeds with a high verbascose concentration and undetectable activities in a low verbascose pea line (Peterbauer et al 2002b; Peterbauer et al 2003, Obendorf and Gorecki 2012).

2.6 Lentil transcriptome analysis

Kaur et al (2011) sequenced lentil transcriptome and an EST (Expressed Sequence Tags)-derived dataset was reported. The EST dataset was comprised of 84,074 unigenes and around 25,000 annotated genes. Sharpe et al 2013, reported transcriptome sequencing of nine *L. culinaris* genotypes and two *L. ervoides* genotypes with CDC Redberry as the reference genotype. CDC Redberry 454 sequencing yielded 1.03×10^6 reads consisting of 50,146 contigs. Filtering of these contigs for duplication, overlap and size yielded around 27,921 contigs. Single nucleotide polymorphisms identified using Golden gate array platform were used to develop a SNP genetic linkage map in lentil (Sharpe et al 2013). Parallel to this research, Verma et al 2013 also reported sequencing of lentil transcriptome, where 119,855,798 short reads were generated by Illumina paired-end sequencing. A total of 42,196 non-redundant transcripts were obtained and 20,009 showed similarity to existing sequences in the database. Using these sequences 5,673 SSR markers were developed which could be used for crop improvement (Verma et al 2013). A search for RFO biosynthetic genes based on sequence similarity in the EST databases only provided with partial sequences for RFO biosynthetic genes (Table 2.6). Transcriptome shotgun assembly showed presence of partial gene sequences, whereas other lentil EST libraries showed no hits for RFO biosynthetic genes. Hence, there is a need for a cDNA library to isolate full length sequences of RFO biosynthetic genes.

2.7 RNA Isolation Methods

Isolation of RNA in adequate quantities and purity is essential for gene expression and transcriptomic studies. Developing lentil seeds accumulate high amounts of polysaccharides, phenolic compounds and proteins during seed development and these compounds interfere with RNA isolation. Polysaccharides co-precipitate with RNA thus reducing RNA purity. Traditional methods for RNA extraction using acidic phenol guanidine methods, successfully

Table 2.6: Partial sequences of RFO biosynthetic genes in the EST database (based on sequence similarity search)

Gene	Contig Number	Accession Number	Contig Length (bp)	Database	Database Type
Galactinol Synthase	Contig 02236	JI848511.1	1324	Kaur et al 2011	Transcriptome shotgun assembly
	Contig 12545	JI858791.1	524	Kaur et al 2011	Transcriptome shotgun assembly
	Contig 15115	JI861352.1	468	Kaur et al 2011	Transcriptome shotgun assembly
Raffinose Synthase	-				
Stachyose Synthase	contig06676	JI852941.1	1373	Kaur et al 2011	Transcriptome shotgun assembly
	contig01829	JI848108.1	452	Kaur et al 2011	Transcriptome shotgun assembly
	contig12445	JI858691.1	451	Kaur et al 2011	Transcriptome shotgun assembly
	contig08687	JI854942.1	463	Kaur et al 2011	Transcriptome shotgun assembly

used to isolate RNA from developing wheat seeds, did not yield good quality RNA from developing lentil seeds as the RNA co-precipitated with starch and other phenolic compounds (Kannan et al 2014). Isolation of RNA with buffers containing sodium tetraborate and sarkosyl followed by LiCl precipitation was used to isolate RNA from lentil leaves, roots, stem, seedling, developing and mature seeds (Dash 2013). RNA yield obtained using this method for leaf, root and stem tissues was: 568 ± 41, 201 ± 23 and 186 ± 18 µg/g of tissue, respectively. However, very low yields were reported for RNA isolated from etiolated seedling tissue (58 ± 11 µg/g), developing (4.7 ±1.9 µg/g) and mature (0 µg/g) seeds (Dash 2013). Recently, commercial kits (Qiagen Inc., Venlo, Limburg, Netherlands) have been used to isolate RNA from lentil leaves and seeds (Kaur et al 2011, Sharpe et al 2013, Verma et al 2013). RNA obtained using these methods were used for cDNA library construction. Trizol reagent based methods have been used to isolate RNA from chickpea, bean and cereal grains that contain high amounts of starch (Garg et al 2010; Blair et al 2011; Ganeshan et al 2012; Wang et al 2012).

2.8 RNA Quantification Techniques

RNA quantification using UV spectroscopy is the most commonly used method to determine small quantities of RNA. Purity of RNA could be estimated based on the $A_{260/280}$ ratio; where a value of 2.0 is predicted to be pure RNA (Wilfinger et al 1996). This quantification technique is reliable and simple, however it has some limitations. Ionic strength of water and pH of the solution influence RNA quantification. Aqueous alkaline solutions (pH > 8.0), produced consistent results and also efficiently determined protein contamination (Wilfinger et al 1996). RNA quantification with 1-3 mM Na_2HPO_4 buffer and pH ranging from 8.0-8.5 was suggested. This method was preferred because preparation of Na_2HPO_4 buffer was easier than Tris-based buffers (Wilfinger et al 1996).

To determine the integrity of RNA, it is separated by agarose gel electrophoresis, stained with ethidium bromide and the RNA bands are visualized under UV light. RNA integrity was determined based on 18S and 28S band intensities. The ratio of 28S: 18S bands of 2.0 or higher implies that the RNA purity iss adequate for use in transcriptomic studies (Schroeder et al 2006).

The newer methods for RNA quantification include utilization of a micro-spectrophotometer ND-1000 (Thermo Fisher Scientific Inc., Waltham, MA, USA), Quant-iT RiboGreen (Invitrogen Inc., CA, Carlsbad, USA) and an Agilent RNA 6000 kit (Agilent

Technologies Inc, Santa Clara, CA, USA) in conjunction with an Agilent 2100 Bioanalyzer. The nanodrop ND-1000 required only 1µL of RNA for estimation. The quantification detection range for ND-1000 was 500 to 2.5 ng μL^{-1}. The accuracy of the determination by ND-1000 was higher when the RNA samples were denatured at 70 ^{o}C in Tris-EDTA buffer (pH 8.0) before quantification (Aranda et al 2009). The Agilent 2100 Bioanalyzer was efficient in assessing sample purity but the determination of RNA concentration was reported to be more accurate using the ND-1000. Quant-iT RiboGreen is a fluorescence based quantification method with detection ranges of 1 to 0.001 ng μL^{-1} (Aranda et al 2009). This technique detects and quantifies RNA at low concentrations (pg μL^{-1}) with high reproducibility and less variation among sample replicates (Aranda et al 2009). The Agilent 2100 Bioanalyzer separates RNA based on their molecular weight in microfabricated chips. An electropherogram is generated after RNA detection and based on these results the software generates a RNA integrity number (RIN). RNA integrity values above 8.0 indicate RNA of high purity and values below 8.0 indicate RNA degradation (Schroeder et al 2006).

2.9 Physiological role of RFO in plants

2.9.1 Role of RFO during seed germination

RFO are rapidly hydrolyzed by α-galactosidases (E.C.3.2.1.22) during early seed germination events. Other polysaccharides, oils and proteins take several hours to days to metabolize and therefore do not meet the carbon source demand during seed germination. Alpha-galactosidases hydrolyze RFO during seed germination by cleaving the terminal α-galactosyl moiety (Blöchl et al 2007). Inhibition of α-galactosidase activity by the addition of a specific inhibitor, 1-deoxygalactonojirimycin delayed seed germination by 70% in pea seeds. It was concluded that RFO are essential for seed germination in pea (Blöchl et al 2007). However, inhibition of galactosidases with 1-deoxygalactonojirimycin in normal and low RFO soybean showed no significant difference in germination levels suggesting that RFO might not play a major role during germination events (Dierking and Bilyeu 2009).

2.9.2 Role of RFO in seed desiccation

Raffinose family oligosaccharides confer desiccation tolerance by two mechanisms, water replacement and glass state formation in the cell. Water loss increases solution viscosity

and proceeds to a stage where its properties are similar to a plastic solid, and this is known as the glass state (Koster 1991). Glass state formation provides stability to the cell by filling up space so as to prevent cellular collapse in biological structures. It also inhibits diffusion, preventing any deleterious reactions in the cell (Leopold et al 1994). The longevity of seeds is correlated with total RFO content and with the ratio of RFO to sucrose. Seeds with sucrose to oligosaccharide ratio < 1.0 had storability half-viability of greater than10 years, whereas, ratios > 1.0 showed storability half-viability of < 10 years (Horbowicz and Obendorf 1994; Lin and Huang 1994).

Accumulation of sucrose is crucial during the early stages of seed drying and accumulation of raffinose prevents sucrose crystallization (Sun et al 1994; Brenac et al 1997). A sucrose to raffinose ratio less than 20 to 1 was associated with desiccation tolerance in maize seeds (Brenac et al 1997). In *Fagus sylvatica L.* (beech), seeds stored for 7 and 12 years showed a decrease in germination capacity, which was associated with an increase in sucrose / (raffinose + stachyose) ratio and increased α-galactosidase activity. Stachyose, the predominant RFO in seeds correlated positively with germination capacity in beech seeds (Pukachka et al 2009).

2.9.3 Role of RFO in phloem transport

Plants translocate carbon fixed in leaves to sink tissues which are utilized for growth and development of the plant. Two loading mechanisms known as Type 1 and Type 2 have been proposed for photoassimilate translocation (Turgeon 1996). Type 2 plants transport sucrose apoplastically, where sucrose transferred from mesophyll cells into apoplasm and is subsequently transferred to sieve element-companion cell complex (SE - CCC) by co-transport with protons (McCaskill and Turgeon 2007). These type 2 plants which include plants such as sugar beet and tobacco are characterized by few plasmodesmata connecting the SE - CCC.

Type 1 plants transport sugars symplastically and are characterized by the presence of numerous plasmodesmata connected to specialized companion cells called intermediary cells. Sucrose diffuses into intermediary cells and is converted into raffinose and stachyose in the intermediary cells. Galactinol synthesized in the intermediary cells is utilized to synthesize raffinose in the intermediary cells (Turgeon 1996). The synthesized raffinose and stachyose molecules are larger in size and thus cannot diffuse back through the intermediary cells and this mechanism is referred to as 'polymer trap' mechanism. This 'polymer trap' mechanism creates a concentration gradient which motivates a phloem flux (Haritos et al 2000). Downregulation of

two GolS genes (*VpGAS1* and *VpGAS2*) in *Verbascum phoenicum* L. (expressed in the intermediary cells in wild type plants) affected photoassimilate translocation with increased accumulation of carbohydrate in the lamina and also showed impaired growth, cholorosis of the leaves and curling of leaf margins (McCaskill and Turgeon 2007).

2.9.4 Role of RFO in abiotic stress

Arabidopsis plants, in response to cold stress, accumulated high concentrations of raffinose in high cold tolerant lines (Col0), whereas only three-fold increase in fructose was observed in low cold tolerant lines (Col24) (Klotke et al 2004). Reciprocal crossing of Col0 and Col24 resulted in heterosis for increased cold tolerance in the F_1 generation. A strong correlation ($r = 0.97$ and 0.96) was obtained between raffinose content and leaf-freezing tolerance (Rohde et al 2004). Further, to understand the role of raffinose in cold acclimation in Arabidopsis, two mutants, (i) overexpressing GolS gene and (ii) RS knockout mutant were synthesized and the cold tolerance of the plants was studied. Overexpression of *GolS*, showed higher concentrations of raffinose in leaf tissues, but no increase in cold tolerance was observed in cold acclimated and non acclimated conditions. No significant difference in cold tolerance was observed between RS mutant and the wild type, suggesting that raffinose is not essential for cold tolerance in Arabidopsis (Zuther et al 2004).

Using a recombinant inbred line population (RIL) derived from crossing cv. Champagne a frost tolerant pea cv Terese a frost sensitive, genetic mapping identified two QTLs for raffinose concentration on linkage groups 5 and 6 that co-localized with QTL for frost damage (Dumont et al 2009).

A decrease in RFO concentration during drought tolerance was reported in *Coleus* plants suggesting that they might not play a role in drought tolerance (Pattanagul and Madore 1999). Galactinol and raffinose scavenge hydroxyl radicals to prevent plants from oxidative damage in *Arabidopsis thaliana* (Nishizawa et al 2008). In summary, there are some suggestions that RFO play a role in abiotic stress tolerance, however it is unclear if the RFO are the cause for abiotic stress tolerance or an effect of abiotic stress conditions.

2.10 Nutritional impacts of RFO in humans

2.10.1 Antinutritional effects of RFO

Raffinose family oligosaccharides, which are considered as antinutritional factors, contribute to limit lentil consumption by humans. RFO have been described as causing flatus in humans and monogastric animals. RFO undergo anaerobic fermentation in the hind gut producing carbon dioxide, hydrogen and methane (Reddy et al 1984). Flatulence can result in stomach discomfort, abdominal rumblings, cramps, pain, and diarrhea. Further, RFO in animal diets have also been associated with a reduction in net dietary energy. Adult roosters fed with diets containing 5.3% RFO showed a 20% reduction in net metabolizable energy compared to a diet containing 1% RFO (Coon et al 1990). Diets with high RFO content caused osmotic imbalance (before fermentation by microbial flora) resulting in reduced absorption in the small intestine and protein utilization (Wiggins 1984; Van Barneveld 1999).

2.10.2 Beneficial effects of RFO

Despite the antinutritional effects of RFO, they have also been suggested to be beneficial to humans. RFO are considered to be prebiotics, as they stimulate the growth of microbial flora in the colon contributing to beneficial effects in hosts (Martínez-Villaluenga 2008b). In rats, diets rich in RFO showed an increase in bifidobacterial growth and increased immune response (Gulewicz et al 2002). In humans, consumption of soybean α-galactosides increased bifidobacterial and eubacterial growth in the large intestine (Hayakawa et al 1990; Wada et al 1991). Prebiotics also affect lipid metabolism, reducing serum cholesterol by precipitation and excretion of bile acids thus allowing the liver to use serum cholesterol for bile acid synthesis (Pedersen et al 1997; Martínez-Villaluenga et al 2008a). Other health benefits of prebiotics include, modulation of immune response (reported in animals), reduction in blood pressure, and reduction of toxic metabolites by alleviating the detoxifying load of the liver (Takasoye et al 1990; Masai et al 1987; Schley and Field 2002).

2.11 Strategies to reduce RFO

2.11.1 Processing methods for RFO reduction

2.11.1.1 Addition of α-galactosidase

Raffinose family oligosaccharides or alpha galactosides are degraded by α-galactosidases (EC 3.2.1.22). Alpha-galactosidases exist as tetramers with molecular mass ranging from 29 to 209 kDa with a broad range of substrate specificity. In *Stachys sieboldii* tubers, the presence of

free galactose (inhibitor of α- galactosidases) facilitated the co-existence of RFOs and galactosidases (Keller and Matile 1985). Inhibition of α-galactosidases showed delayed germination in pea seeds (Blöchl et al 2007). α-galactosidases from *Cladosporium cladosporides*, *Aspergillus oryzae* and *Aspergillus niger* showed 100% reduction in raffinose oligosaccharides in chickpea flours (Mansour and Khalil 1997). Machaiah et al (1999) demonstrated degradation of RFO's with γ-irradiation at levels sufficient for insect disinfestation without damage to seed viability in mung bean.

2.11.1.2 Germination

Germination of seeds also alters soluble carbohydrate and RFO content in seeds. Germination for three days reduced RFO concentration by 18 - 40% in lentil seeds (Frias et al 1996). Germination of pigeon pea (*Cajanus cajan*) for four days at 20 ^{o}C in dark reduced α-galactosides up to 83% (Torres et al 2007). Addition of commercial α-galactosidases reduced the RFO content in lentil and peas (Frias et al 2003). Presoaking and cooking of chickpea seeds also decreased RFO concentration (Frias et al 2000). Soaking of seeds reduced RFO concentration by 16 - 27%, however, the soaking efficiency was also influenced by soaking time, temperature, seed to water ratio and by addition of sodium bicarbonate (Abdel-Gawad 1993; Vijayakumari et al 1996; Ibrahim 2002). Six days of germination in light for bean and lentil while germination in darkness for pea increased the nutritive value of the flours (Vidal-Valverde et al 2002).

2.11.1.3 Fermentation

Significant reduction of RFO was achieved through the fermentation of flour dough. Fermentation improved texture and aroma and also improved biological value of flour of *Phaseolus vulgaris* (Granito et al 2003). A reduction of 27% and 37% in stachyose content was reported when pinto bean and field pea flours were treated with *Lactobacillus fermentum* or *Lactobacillus plantaram* (Duszkiewicz-Reinhard et al 1994). Significant reduction in RFO was also obtained by natural fermentation of soybean and corn (Chompreeda and Fields 1984). Frias et al (1996) reported that a decrease in RFO through natural fermentation depended on the amount of lentil flour and fermentation temperature. In *Phaseolus vulgaris* seeds, natural fermentation showed higher reduction in RFO compared to controlled fermentation. However,

natural fermentation for 96 h reduced stachyose by 95%, whereas only 11% reduction was obtained in controlled fermentation (Granito et al 2003).

2.11.2 Molecular strategies to reduce RFO

Wide genetic variation is essential for molecular breeding to reduce the concentration of RFO in lentil seeds. In lentil, a wide variation in RFO has been reported in domesticated and wild lentil genotypes (Tahir et al 2011a, 2011b, 2012). Similar natural variation has also been reported in pea where variation in RFO concentration was determined in 78 pea lines (Jones et al 1999). During pea seed development, differences in gene expression and enzyme activity of RFO biosynthetic enzymes was found, suggesting that RFO biosynthetic enzymes affect seed RFO concentration (Peterbauer et al 2001b).

When no significant natural variation in a trait is available, it can be achieved through induced mutagenesis; either chemically or by UV irradiation. Chemical mutagenesis was used in soybean and a low RFO line was identified and was subsequently used to develop a mapping population so as to identify candidate genes coding for low RFO (Hitz et al 2002; Dierking and Bilyeu 2008). The low raffinose and stachyose phenotype was associated to a raffinose synthase gene by QTL analysis (Skoneczka et al 2009). DNA markers were developed based on mutation in RS gene to screen soybean lines with low RFO (Skoneczka et al 2009). However, with induced mutagenesis background mutations can occur and as a result several back crosses are needed to eliminate the background mutations.

The presence of ciceritols has been reported in chickpea and lentil. Frias et al (1999) reported a negative correlation between verbascose and ciceritol concentrations in lentil seeds. Based on these results they suggested the presence of a metabolic link between RFO and ciceritol synthesis. Alpha-galactosidases hydrolyze ciceritols slower than RFO, suggesting lower flatus production by ciceritols compared to RFO (Quemener and Brillouer 1983; Frias et al 1999). STS purified from lentil was also reported to catalyze the synthesis of galactopinitol A, galactopinitol B and ciceritol (Hoch et al 1999). Manipulation of the RFO biosynthetic pathway for carbon partitioning in seeds with low RFO and high ciceritol might be another strategy to reduce seed RFO. The role(s) of ciceritol towards abiotic stress in plants and human nutrition is yet to be established (Hedley 2001).

Molecular approaches targeting over-expression or down-regulation of individual enzymes in the RFO biosynthetic pathway would be another strategy to reduce RFO in seeds. A strategy involving genetic transformation for reduction of RFO post seed harvest was suggested by Griga et al (2001). Genetic transformation of legumes with α-galactosidases from thermophilic bacteria (*Thermotoga* species) into legumes has also been suggested. These α-galactosidases have been reported to have temperature optima of 100 ^{o}C with decreased activity at lower temperatures (McCutchen et al 1996). The α-galactosidase gene would be heat activated post harvest in seeds resulting in RFO degradation (Griga et al 2001; Wang et al 2003). Expression of an α-galactosidase gene from *Coffea arabica* in *Pisum sativum* showed a reduction in RFO content up to 40% (Polowick et al 2009). These seeds with reduced RFO showed high germination rates (96%) suggesting that RFO might not be the only compounds contributing to seed germination. In *Brassica napus*, down-regulation of GS by an antisense approach also showed RFO reduction in seeds (Bock et al 2009).

The beneficial and antinutritional effects of RFO depend upon the amount of RFO in a pulse seed crop and to some extent the human sensitivity to RFO. Therefore, it is essential to develop pulse crops with optimal concentrations of RFO, so that the beneficial effects of RFO are retained without causing the adverse affects on human health. To achieve the optimal RFO concentration in pulses it is imperative to have a good understanding of RFO biosynthesis at both the molecular and biochemical level, so that precise targets can be identified to alter the concentration and composition of RFO in lentils.

3. A CTAB BASED PROTOCOL TO ISOLATE HIGH QUALITY RNA IN ADEQUATE QUANTITIES FOR GENE EXPRESSION ANALYSES IN DEVELOPING SEEDS OF LENTILS (*Lens culinaris* Medik.)

Study 1*

In this study, a CTAB-LiCl based protocol was optimized for efficient RNA isolation from lentil seeds at various stages of seed development.

***Kannan U**, Ganeshan S, Chibbar RN (2014) A hexadecyltrimethylammonium bromide (CTAB) based protocol to isolate high quality RNA in adequate quantities for gene expression analyses in developing seeds of lentils (*Lens culinaris* Medik.). Gene Expression to Genetical Genomics 7: 7-16.

3.1 ABSTRACT

RNA isolation from polysaccharide- and phenolics- rich plant tissues such as developing lentil *Lens culinaris* Medik.) seeds is challenging. RNA of high purity is needed in adequate quantities or transcriptome analysis to study seed quality traits. To date, a suitable method to isolate RNA of high -purity and -quantity from lentil seeds has not been reported. The objective of this study was to develop a simple and reproducible method to isolate high-quantity and -purity RNA from developing lentil seeds for gene expression analysis. Methods based on Trizol™ reagents and phenol:guanidine gave low yields of RNA. A method based on CTAB followed by a lithium chloride precipitation yielded RNA in high quantity (210 - 260 μg from 200 mg of seeds) and purity (*A* 260/280 ratio of about 2.2). Isolated RNA was used to study the expression of the *granule-bound starch synthase I* (*GbssI*) during lentil seed development by RNA gel blot and quantitative real-time polymerase chain reaction. The expression pattern of the *GbssI* was similar to that reported for pea *GbssI* gene.

3.2 Introduction

Lentil (*Lens culinaris* Medik.) is a cool season legume crop producing seeds with high nutritional value, widely consumed in the Mediterranean and developing countries. Lentil seeds have high concentration of protein rich in essential amino acids like lysine and threonine which are deficient in cereals. Lentil seeds are also a good source of carbohydrates, minerals, and micronutrients. Carbohydrates are the major storage components in lentil seeds and contribute to their beneficial effects on human health (Chibbar et al 2010). In recent years, there is an increase in global demand for lentils with desirable seed quality. Thus, seed quality improvement has become a major focus of research to reduce some of the antinutritional constituents such as raffinose family oligosaccharides (RFO) and tannins. Although RFO cause stomach discomfort (Jood et al 1985), tannins bind mineral ions thereby reducing their bio-availability to humans (Davis 1981).

High-throughput transcriptomic studies involving deep sequencing have contributed valuable genomic resources for lentil (Kaur et al 2011). RNA extracted from mixed tissues which included young and mature leaf tissues, stem, flowers, and immature and mature pods was used to construct a cDNA library for transcriptome sequencing, which identified over 84,000

unigenes and enabled validation of sets of SSR primer pairs for polymorphisms among cultivated genotypes as well as among non-domesticated genotypes (Kaur et al 2011). This transcriptomic resource set (Kaur et al 2011), while valuable, is challenging to mine for seed-specific transcripts, especially low-abundant transcripts.

For large-scale transcriptome profiling for lentil seed quality traits, it is imperative to establish high-purity and quantity RNA isolation procedures from various seed developmental stages. To date, high-purity and -quantity RNA isolation methods from lentil seeds have not been reported. Although this requirement may seem trivial due to the availability of numerous RNA isolation protocols from a number of different species, several of the published protocols could not be used to extract RNA of high purity in adequate quantities from developing seeds of lentil. Most RNA isolation methods reported in the literature employ commercial kits, which yield pure RNA with low quantities adequate for applications such as quantitative real-time polymerase chain reaction (QPCR) and high-throughput transcriptome analyses such as microarrays. For cereal and leguminous seeds rich in polysaccharides, commercial kits with minor modifications have been successful to isolate RNA. However, a major limitation of commercial kits is insufficient amounts of RNA for applications such as RNA gel blot analyses. In addition, resource limitations for many laboratories may prevent purchase of these commercial kits for routine use.

A number of RNA isolation methods, both conventional hot acidic phenol:guanidine (Ganeshan et al 2010) or formulation thereof like Trizol™ reagent (Ganeshan et al 2012; Wang et al 2012) have been used for starch-rich seed tissues. The conventional methods are cost-effective and yield higher amounts of RNA than commercial kits, although down-stream applications can sometimes be inhibited due to carry-over of isolation reagents such as phenol, guanidine, and chloroform. The objective of this study was to develop a simple, reproducible, high-yielding and high-purity RNA isolation protocol from developing seeds of lentil. Published methods from our laboratory for developing wheat grains based on hot phenol:guanidine (Ganeshan et al 2010) and Trizol (Ganeshan et al 2012) followed by a column cleanup were not applicable to developing lentil seeds. A CTAB method, generally used for DNA isolation (Rogers and Bendich 1985; Doyle and Doyle 1990) followed by a lithium chloride RNA precipitation step was developed to isolate RNA of high-purity in adequate quantities from developing lentil seeds. The optimized protocol also included polyvinylpyrrolidone (PVP),

which effectively removed polyphenolic compounds during nucleic acid isolation (Rogers and Bendich 1985; Doyle and Doyle 1990). CTAB-based RNA isolation methods have been used for a number of plant tissues, including polysaccharide-rich loquat (Zhang et al 2013), lipid-rich seeds (Lan et al 2013) and polyphenols-and polysaccharide-rich mangrove plants (Jorge and Omar 2011). However, with any new study, methods need to be re-evaluated prior to selecting the most efficient method.

3.3 Materials and Methods

3.3.1 Plant material and RNA integrity determination.

Lens culinaris Medik. cv CDC Redberry (Vandenberg et al 2006) seeds were germinated and grown in growth chambers with 18/6 h photoperiod [250 μmol m^{-2} sec^{-1} Photosynthetic Photon Flux Density (PPFD]and 21/15 °C day/night temperatures. Plants were fertilized every two weeks with N:P:K; 20:20:20 (Plant Products Co. Ltd., Brampton, Ontario, Canada). At anthesis, flowers were tagged and the developing seeds were collected at two-day intervals from 8 to 32 days after flowering (DAF), frozen in liquid N2 and stored at -80 °C for total RNA isolation. For seedling tissues, seeds were germinated for three days in sterile petri dishes containing half-strength Hoagland's solution (Hoagland 1918). Germinated seeds were grown hydroponically in sterile Magenta boxes containing glass beads (1 - 5 mm diameter) and half-strength Hoagland's solution for two weeks. Leaves, stems, and roots were collected separately and stored at - 80 °C for RNA isolation.

Total RNA concentration was determined using a spectrophotometer at absorbance *A*260 nm and RNA purity at ratio of absorbance *A*260 nm/280 nm. RNA integrity was determined on a Bioanalyzer 2100 (Agilent Technologies Inc., Santa Clara, CA, USA) or by electrophoresis in HAE buffer (0.2 M Na HEPES, 50 mM sodium acetate, 10 mM EDTA, 0.7% (v/v) formaldehyde) and formaldehyde agarose gels (0.2 M Na HEPES, 50 mM sodium acetate, 10 mM EDTA, 0.7% formaldehyde (v/v), 1% agarose (w/v), 0.1 μg/mL ethidium bromide) and visualized under UV light.

3.3.2 Trizol reagent RNA isolation method

Total RNA was isolated from lentil leaves, stems, and roots using the Trizol reagent (Life Technologies Inc., Burlington, Ontario, Canada) and column cleanup by the PureLink RNA kit

(Life Technologies Inc., Burlington, Ontario, Canada) as optimized for total RNA isolation from wheat leaves (Ganeshan et al 2008). Total RNA isolation from lentil seeds was then tested using this method. An isopropanol precipitation step prior to column cleanup was also performed with another set of samples. Briefly, lentil seeds (200 mg) were ground in liquid N_2 in a -20 °C pre-cooled RNase-free mortar. The ground seeds were transferred to an RNase-free container at room temperature containing 2 mL Trizol reagent. The slurry was mixed well and transferred to a 15 mL tube. The tube was vortexed for 10 sec, incubated at room temperature for 5 min, and centrifuged at 15,000 x g for 10 min at 2 °C. The supernatant was transferred to a new RNase-free tube, 1 mL of chloroform was added and the tube was vortexed for 15 sec. The slurry was centrifuged at 15,000 x g for 10 min at 2 °C. The colorless aqueous phase was transferred to a new RNase-free tube and re-extracted with 1 mL of chloroform. The supernatant was transferred to a new RNase-free tube and 1 mL of isopropanol was added, mixed, incubated at room temperature for 10 min, and centrifuged at 1000 x g at 2 °C for 5 min. The aqueous phase was discarded. The RNA pellet was washed with 1 mL of 75% (v/v) ethyl alcohol, briefly air-dried and resuspended in 50 μL of DEPC-treated water.

3.3.3 Phenol:guanidine RNA isolation method

A second isolation method tested was the hot phenol:guanidine method optimized for total RNA isolation from wheat seeds (Ganeshan et al 2010). Briefly, lentil seeds (200 mg) were ground as described earlier. The ground seeds were transferred to an RNase-free container containing 2 mL of hot 65 °C phenol:guanidine buffer (50% (v/v) acidic phenol, pH 4.3; 30% (w/v) guanidine–HCl; 5% (w/v) sodium dodecyl sulfate). The slurry was mixed well, transferred to a 15 mL tube and centrifuged at 15,000 x g for 10 min to pellet down cell debris and starchy components. The supernatant was transferred to a new RNase-free tube and equal volume of chloroform was added to the tube and vortexed for 15 sec. The mixture was centrifuged at 15,000 x g for 10 min at room temperature. The aqueous phase was transferred to a new RNase-free tube and re-extracted with equal volumes of chloroform until the inter-phase was clear. The supernatant was transferred to a new RNase-free tube and 1 mL isopropanol was added and incubated at -80 °C for 10 min and centrifuged at 1000 x g for 5 min. The supernatant was discarded. The RNA pellet was washed in 1 mL of 75% (v/v) ethyl alcohol, briefly air-dried, and resuspended in 50 μL of DEPC-treated water.

3.3.4 Lysis™ buffer RNA isolation method

This method employed the Lysis™ buffer available with the PureLink RNA Mini Kit as per the manufacturer's protocol. Briefly, ground lentil seeds (200 mg) were transferred to an RNase-free container at room temperature containing 4 mL Lysis buffer and 40 µL of β-mercaptoethanol. The slurry was mixed well and transferred to a 15 mL tube and incubated at room temperature for 2 – 3 min. The mixture was centrifuged at 15,000 x g for 10 min. The supernatant was transferred to a 15 mL tube and 0.5 volume of 100% ethyl alcohol was added and mixed well by vortexing. The sample was added to a spin column and all subsequent procedures were as per the manufacturer's instructions. The RNA was eluted from the column with a final volume of 50 µL RNase-free water.

A modification of the procedure described above included mixing the ground seeds with the Lysis buffer, followed by addition of phenol:guandine and chloroform extraction and subsequent precipitation of RNA with isopropanol as described earlier.

3.3.5 CTAB -LiCl RNA isolation method

Another method tested involved the CTAB buffer. Ground lentil seeds (200 mg) were transferred to an RNase-free container containing 2 mL of CTAB buffer (4% (w/v) CTAB; 1% (w/v) PVP; 20 mM EDTA and 1.4 M NaCl) maintained at 65 °C and containing 20 µL β-mercaptoethanol. The slurry was mixed well, transferred to a 15 mL tube and incubated at 65 °C for 15 min with intermittent shaking. The mixture was centrifuged at 15,000 x g for 10 min at room temperature. The supernatant was transferred to a 15 mL tube, equal volume of chloroform was added and the tube was vortexed for a few sec. The mixture was centrifuged at 15,000 x g for 15 min at room temperature. The chloroform extraction was repeated until the inter-phase was clear. The supernatant was transferred to a new RNase-free tube and LiCl was added to final concentration of 4 M. The mixture was incubated at -20 °C for 30 min and centrifuged at 15,000 x g for 15 min at 2 °C. The supernatant was discarded and RNA pellet was washed with 1 mL of 75% (v/v) ethyl alcohol, briefly air-dried and resuspended in 50 µL of DEPC-treated water.

3.3.6 Validation of isolated RNA for gene expression analyses

To test the CTAB-isolated RNA for gene expression analyses, RNA gel blot was performed according to standard procedures (Sambrook and Rusell 2001). Total RNA (25 µg) was loaded on a 1% (w/v) agarose gel containing 0.7% formaldehyde and electrophoresed in HAE buffer for 1.5 – 2 h at 100 V. Ribosomal RNA present on the gel was visualized with ethidium bromide to check for equal loading of RNA samples. Transfer of total RNA from the gel to a Hybond–N+ membrane (GE Healthcare, Baie d'Urfe, Quebec, Canada) was done using a positive pressure blot, PosiBlot™ apparatus (Agilent Technologies Inc., Santa Clara, CA, USA) for 1 h at 75 mm Hg pressure. RNA was cross-linked to the membrane in a UV cross-linker (UVP Inc., Upland, CA). Transferred RNA on the membrane was hybridized in Church hybridization buffer (0.5 M Na2HPO4, 7% SDS (w/v), and 1 mM EDTA) at 65 °C with a 175 bp probe corresponding to the *granule-bound starch synthase I* (*GbssI*) labeled with [α32-P] dCTP. The membrane was washed two times in 1 × SSPE (150 mM NaCl; 10mM NaH2PO4. H2O; 1 mM EDTA; 0.1% SDS (w/v)) for 10 min each, and then in 0.1 × SSPE, 0.1% SDS (w/v) until background activity was negligible. Membrane was exposed to Biomax MR film (Kodak) at -80 °C for 24 h to several days depending on observed signal intensities.

For QPCR analyses, 5 mg of DNase-treated total RNA was reverse transcribed using 200 U Superscript III (Life Technologies Inc., Burlington, Ontario, Canada). The 25 µL reaction volume of PCR consisted of 5 µL of a 1/15 dilution of the cDNA, GBSS_1F and GBSS_1R primers for the *GbssI* at 300 nM concentrations and 1 x Maxima™ SYBR Green I QPCR Master Mix (Fisher Scientific Company, Ottawa, Ontario, Canada). Real-time PCR amplification was performed in a Mx3000P real-time PCR machine (Agilent Technologies Inc., Santa Clara, CA, USA) under the following conditions: initial denaturation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 sec, 60 °C for 1 min, and 72 °C for 1 min. The ARF_1F and ARF_1R primers corresponding to the lentil *adenosyl ribosylation factor* (*ARF*) gene was amplified for use as reference gene. Threshold values (Ct) were used to calculate relative expression by the 2-ΔΔCt method (Livak and Schmittgen 2001) using the 32 DAF sample as a calibrator. A dissociation step was also performed at the conclusion of the amplification to ensure that a single specific product was amplified for each gene. Real-time PCR experiments were repeated twice with two technical replicates for each sample. Analysis of variance (ANOVA) was performed using the Minitab software (Version 16) (Minitab, Inc., State College, PA, USA).

3.4 Results and Discussion

3.4.1 General considerations for total RNA isolation from lentil seeds

Total RNA isolation from developing seeds of field crops has always been challenging due to the complexity of seed composition and seed sizes at various stages of development. In cereals, for example, starch constitutes about two-thirds of the grain dry weight and the interference of starch during RNA isolation leads to reduction of RNA yield and purity. Nonetheless, RNA isolation methods have been developed for successful use in transcriptome profiling studies of developing cereal grains (Laudencia-Chingcuanca et al 2007; McIntosh et al 2007; Ganeshan et al 2010). However, in pulse crops such as lentil, similar and further challenges are encountered. Besides starch, which constitutes 35 – 65% dry matter in lentil seeds, oligosaccharide constituents such as raffinose, stachyose, verbascose, and α-galactosides account for 6 – 18% seed dry weight as reviewed in Chibbar et al 2004, and phenolic compounds such as phenolic acids (0.25 – 0.47%) and tannins (0.02 – 1.0%) reviewed in Grusak 2009 also interfere with RNA isolation. Furthermore, lentil seeds at early stages of development are very small (Figure 3.1). Considering all these factors, we first tested the Trizol method followed by column cleanup, since this is the most widely used method. We have successfully used the Trizol/column cleanup method for RNA isolation from wheat leaves (Ganeshan et al 2008) and developing wheat grains (Ganeshan et al 2012) for QPCR analyses. When tested with 100 mg leaf, stem, and root tissues of lentil, yield of RNA was high using this method, about 50, 25, and 65 µg, respectively. The purity of RNA was also good (Figure 3.2A).

3.4.2 Recalcitrance of lentil seeds to Trizol method for RNA isolation

Since the Trizol RNA isolation method produced RNA of high-purity and -quantity from leaf, stem, and root tissues of lentil seedlings, we proceeded with this method for RNA isolation from 22 DAF lentil seeds. This stage was selected as it was in the mid-developmental stage and considered ideal for RNA optimization experiments. However, using the Trizol method, RNA yield was generally low, ranging from 2 to 2.7 µg from 200 mg of seed tissues (Table 3.1). Most often, RNA was also degraded. One factor considered for improving the Trizol isolation method was to increase the Trizol reagent to 4 mL to dilute and more effectively remove contaminants

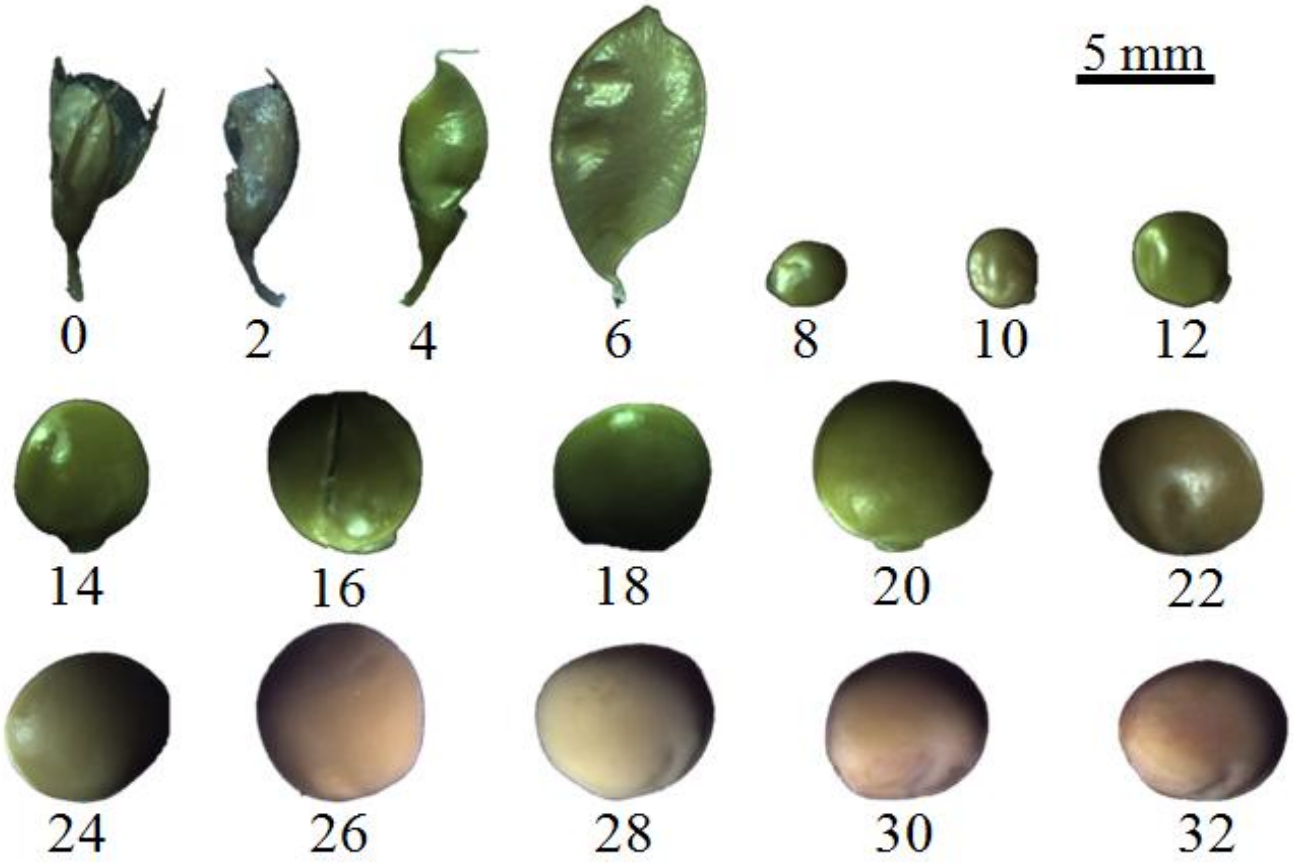


Figure 3.1: Sizes of lentil seeds at various stages of development from 0 – 32 days after flowering (DAF). Day 0 is equivalent to the stage at which pollination has just occurred. DAF 2 to 6 are shown with the developing seeds still within the pods.

during the chloroform extraction step, prior to column cleanup. While the *A*260/280 reading was 2.1, a value acceptable for RNA, the yield was still low at 2.1 μg (Table 3.1). Cloudiness of the supernatant upon addition of equal volume of 70% ethyl alcohol was also observed. Manufacturer's instructions suggest that cloudiness generally disappears upon vortexing, but it still persisted after mixing vigorously. It is suspected that the ethyl alcohol may have precipitated some of the RNA causing reduced RNA yield. A further modification included isopropanol precipitation after Trizol and chloroform extractions. RNA yield was in the range of 73 – 140 μg, with a low *A*260/280 reading of 1.6 (Table 3.1) from 400 mg of seeds, indicating protein contamination. Successful use of Trizol reagent for RNA isolation from developing cotyledons of legumes such as *Phaseolus vulgaris* (Blair et al 2011) has been reported. It was therefore conceivable that the complexity of the lentil seed composition may have reduced the efficacy of the Trizol reagent isolation method, and subsequent modifications of the original method were tested.

3.4.3 Unsuitability of in-house phenol:guanidine method

Since the Trizol method did not show promising results, a hot phenol:guanidine method previously optimized for RNA isolation from developing wheat grains (Ganeshan et al 2010) was tested. Generally, degraded RNA (data not shown) was observed. When RNA was not degraded, 200 mg lentil seed meal yielded only about 2 μg RNA with *A*260/280 ratio of 2.6 (Table 3.1). With the hot phenol:guanidine isolation procedure, the slurry with the ground seeds became gelatinous. This could be due to the starchy components of the seeds, as reported in a previous study (Li and Trick 2005). This would explain the low RNA yield because it could become trapped within this gelatinous mass and would also account for its subsequent degradation. To prevent interference of the starch during RNA isolation when the ground seeds were mixed with the phenol:guanidine and incubated at 65 °C, a pre-spin for 10 min at 15,000 x g was performed. The supernatant was transferred to a new tube and extracted with equal volume of chloroform followed by column cleanup. This modification improved the yield of RNA; however, its purity was still poor as reflected by the *A*260/280 ratio of 1.5 (Table 3.1). Another modification was lithium chloride precipitation of RNA, followed by resuspension in RNase-free water and column cleanup. The yield was still low (20 μg/200 mg seed meal) with *A*260/280 ratio of 1.7 (Table 3.1). The phenol:guanidine method was found to be unsuitable for RNA isolation from lentil seeds.

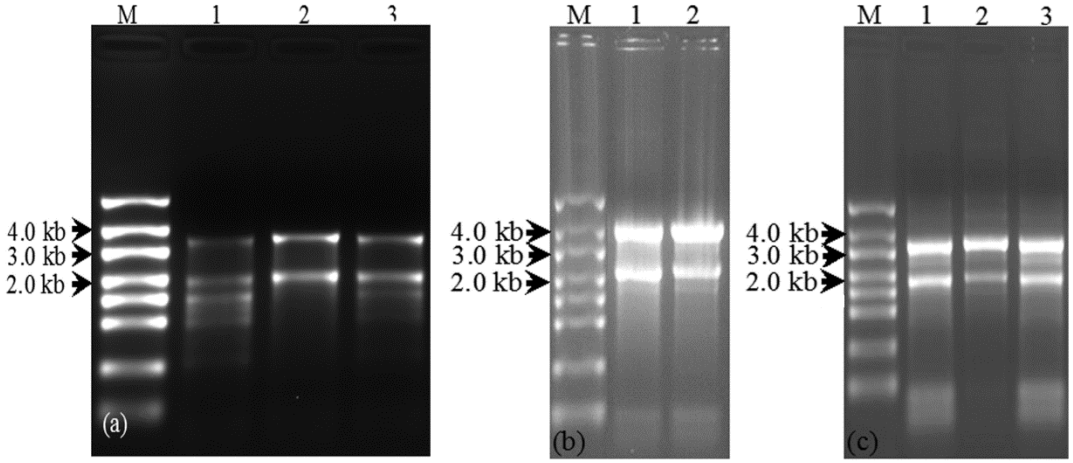


Figure 3.2: Gel electrophoresis of RNA samples resolved on formaldehyde agarose gels (a) isolated using the Trizol™ method from different tissues of lentil seedlings: Lane 1-Leaf ; Lane 2 - Root; Lane 3- Stem; (b) CTAB-extracted RNA: Lane 1 - before Trizol Clean-up; Lane 2 - after Trizol clean-up; (c) Modifications of the CTAB extraction: Lane 1 – CTAB-LiCl with column clean up; Lane 2 - Incubation of ground seeds in CTAB + βME buffer for 30 min at -20 ^{o}C; Lane 3 - Incubation of ground seeds in CTAB + βME buffer for one h at room temperature. M – RNA ladder.

Table 3.1: RNA isolation methods and modifications tested with lentil seeds.

Extraction Method	Seed weight (mg)	$A_{260/280}$ ratio	Total RNA yield (µg)
Trizol reagent			
Trizol (2 mL) + PureLink RNA mini Kit	200	1.7	2.7
Trizol (4 mL) + PureLink RNA mini Kit	200	2.1	2.1
Trizol + isopropanol precipitation	400	1.6	73 - 140
Phenol:guanidine			
Phenol-guanidine	200	2.6	2.0
Phenol-guanidine pre-spin	200	1.5	17.5
Phenol-guandine with LiCl precipitation	200	1.7	20
Lysis Buffer			
Lysis buffer+ β-ME* + column	200	2.1	19.5
Lysis buffer+βME + isopropanol precipitation	200	1.6	10
Lysis +βME +phenol-guanidine+ isopropanol precipitation	200	1.8	220
Lysis buffer +Trizol	200	1.8	16-21.7
CTAB			
CTAB +βME +4 M LiCl precipitation	200	2.2	140
CTAB+ βME + LiCl precipitation for 30 min at -20 ^{o}C	200	2.2	260
CTAB +βME + 1 h at room temperature	200	2.2	210
CTAB+ LiCl before clean up	200	2.2	247
CTAB after Trizol clean up	200	2.2	180

* βME = β- mercapto ethanol

3.4.4 Commercial lysis buffer method improves RNA isolated from lentil seeds

Since the two previous methods did not yield intact RNA in adequate quantities, Lysis buffer supplied with the PureLink RNA Mini Kit as per the manufacturer's instructions was tested for lentil seed tissue lysis and extract RNA. The resultant RNA showed *A*260/280 ratio of 2.1 and yields of 19.5 μg (Table 3.1). RNA purity check on the 2100 Bioanalyzer also indicated RNA of high purity (Figure 3.3A). Although this kit-based method was easier and faster than our previous methods tested, RNA yield was still inadequate for applications such as RNA gel blots. The increase in yield may be attributed to the Lysis buffer (which contains an unspecified concentration of guanidine isothiocyanate), to more effectively prevent RNase activity, thereby yielding RNA of high purity. It is therefore surprising that our in-house phenol:guanidine mix did not yield pure RNA with improved yield. In fact, acidic phenol in addition to preventing RNase activity allows partitioning of RNA into the aqueous phase to preferentially maximize RNA isolation (Chomczynski and Sacchi 1987). It is possible that the gelatinous mass observed in the hot phenol:guanidine isolation procedure used earlier was due to the phenol forming a complex with the carbohydrates and simultaneously trapping most of the RNA. This was confirmed when we combined the Lysis buffer with the phenol– guanidine mix, as discussed below.
The Lysis buffer with an isopropanol precipitation step, followed by resuspension in RNase-free water and column cleanup did not increase the yield or purity of RNA (10 μg) (Table 3.1, Figure 3.3B). The next procedure included mixing lentil seed meal with the Lysis buffer followed by addition of hot phenol:guanidine, chloroform extraction of RNA, and its precipitation with isopropanol. While the RNA yield was much higher at 220 μg (Table 3.1), RNA tended to co-precipitate with starch and other polysaccharides upon incubation on ice. Heating the samples dissolved the pellet, but it led to unpredictable amounts of RNA for gel loading, unless the samples were quantified prior to each use. This is not practical for routine RNA gel blot analyses. For RNA of some seed samples, mostly at the later stages of maturity, dissolution of the pelleted gelatinous mass was also difficult even after incubation at 65 °C for 30 min. Freeze-thaw cycles of the samples at -80 °C further exacerbated the dissolution of the pellet. A similar problem was encountered with the use of the Lysis buffer with Trizol, with RNA yield being much low at 16 – 22 μg (Table 3.1).

3.4.5 Combined CTAB–LiCl method consistently yields high-quantity and -pure RNA

It is evident that conventional methods initially tested for RNA isolation from lentil seeds were more challenging than anticipated. As mentioned earlier, the lentil seed contains relatively high concentrations of carbohydrates and phenolic compounds. One of the well-documented methods for nucleic acid isolation from plant tissues is the CTAB method, for protocols and references therein see Allen et al 2006; Macrae 2007. Originally developed for isolation of high molecular weight DNA from plant tissues (Murray and Thompson 1980), the CTAB buffer was further modified to include PVP to efficiently bind phenolic compounds to prevent their interference with DNA isolation (Rogers and Bendich 1985; Doyle and Doyle 1990). Similarly, PVP was included in the RNA isolation buffer to remove phenolic compounds (Claros and Canovas 1998)

The CTAB buffer for RNA isolation was the same buffer that has been previously successfully used for large-scale DNA isolation from cereal leaf tissues (Baga et al 2007) and was based on the modified method of Doyle and Doyle (1990). The lentil seed meal was mixed with 4% CTAB buffer supplemented with β-mercaptoethanol to isolate RNA which was then extracted with chloroform and finally precipitated with 4 M LiCl. The RNA yield from 200 mg of seed sample was 140 – 260 µg with *A*260/280 readings of about 2.2 (Table 1). RNA purity as determined on the 2100 Bioanalyzer was high (Figure 3.3C, 3.3D). LiCl selectively precipitates RNA, albeit slowly (Cathala et al 1983). However, DNA can also co-precipitate (Figure 3.2B) and this can be circumvented by a Trizol cleanup procedure of the RNA sample (Figure 3.2B). The RNA sample is mixed with an equal volume of Trizol, followed by chloroform extraction and reprecipitation with isopropanol. Although the yield of RNA decreased from 248 to 180 µg with this step (Table 3.1), the yield was adequate for RNA gel blot analyses. A 30-min incubation at -20 °C precipitated similar amount of RNA compared to the overnight precipitation initially performed (Table 3.1) and the RNA was of high purity (Figure 3.2C). Similarly, the duration that the ground samples could be left in CTAB buffer + β-mercaptoethanol at room temperature also showed that one h incubation did not affect RNA yield (Table 3.1) or purity (Figure 3.2C). This was done to ascertain that if multiple samples are being processed for RNA isolation that the initial few samples ground and mixed with CTAB buffer + β-mercaptoethanol did not start to degrade. The optimized protocol (Figure 3.4) was used for RNA isolation with high purity from developing lentil seeds (Figure 3.5).

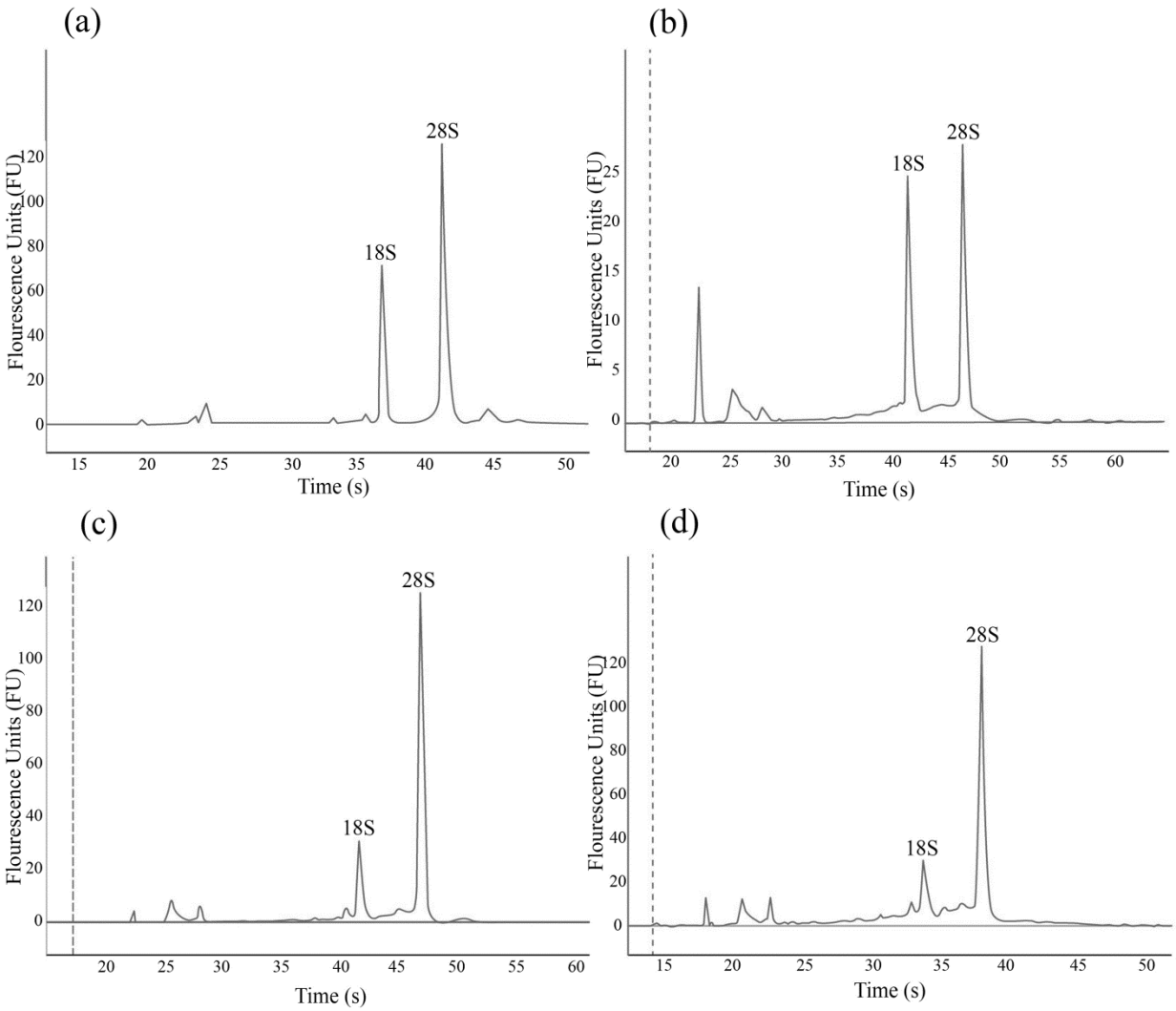


Figure 3.3: Electropherogram of microfluidic electrophoresis of RNA isolated from lentil seeds. **(a)** Lysis Buffer + β-mercaptoethanol method, followed by column clean-up using the PureLink RNA mini kit; **(b).** Lysis Buffer + β-mercaptoethanol method combined with phenol:guanidine and column-clean-up. **(c)** CTAB method, followed by LiCl precipitation and pellet re-suspended in RNase-free water; **(d)** CTAB method, followed by LiCl precipitation which was re-extracted after Trizol reagent treatment.

3.4.6 CTAB-based RNA isolation protocol is suitable for gene expression analyses

A combined CTAB-lithium chloride method described above was used to isolate RNA from developing lentil seeds of 12, 16, 20, 22, 24, and 28 DAF, and used in an RNA gel blot analysis to study the expression of *GbssI* during seed development (Figure 3.6A). Gel blot revealed differential expression of the *GbssI* gene during seed development, showing maximum transcript abundance at 24 DAF (Figure 3.6A). For QPCR analysis, the expression of *GbssI* was assessed using cDNA synthesized from RNA extracted from 12 to 32 DAF seeds. Peak expression of the *GbssI* was observed at 24 DAF (Figure 3.6B), similar to the observed transcript abundance at 24 DAF on the RNA gel blot (Figure 3.6A). The expression patterns corroborate accumulation patterns of *GbssI* in developing seeds of pea, where accumulation of *GbssI* was observed to increase during seed development and decreased toward seed maturity (Dry et al 1992).

3.5 Conclusions

Four methods were tested for RNA isolation from lentil seeds. The most commonly used Trizol reagent and hot phenol:guanidine methods did not yield satisfactory RNA quantity or purity. Similarly, the Lysis buffer available with the RNA PureLink Mini kit, while showing improvements, did not yield RNA of adequate quantity. Despite several modifications of the first three protocols, isolation of high-yield RNA with high purity from the lentil seeds could not be obtained. Subsequently, the CTAB-LiCl protocol was found to be the most efficient for RNA isolation from developing lentil seeds. The developed isolation protocol is cost-effective and rapid. Furthermore, the CTAB buffer is the same buffer used for DNA isolation and therefore serves as a dual purpose nucleic acids isolation buffer. The method presented in this study is believed to be more efficient compared to some of the methods mentioned earlier (Zhang et al 2013; Lan et al 2013; Jorge et al 2011) because of the high CTAB concentration used, inclusion of PVP, and increased LiCl molarity to 4 M for precipitation. Besides validating the RNA for use in gene expression analyses, we have successfully used it to generate a cDNA library from developing lentil seeds, which is being used to isolate and characterize full-length cDNA clones of genes influencing grain quality trait (manuscript in preparation). Contaminating DNA in the isolated RNA samples can be effectively removed by Trizol reagent treatment for less-sensitive

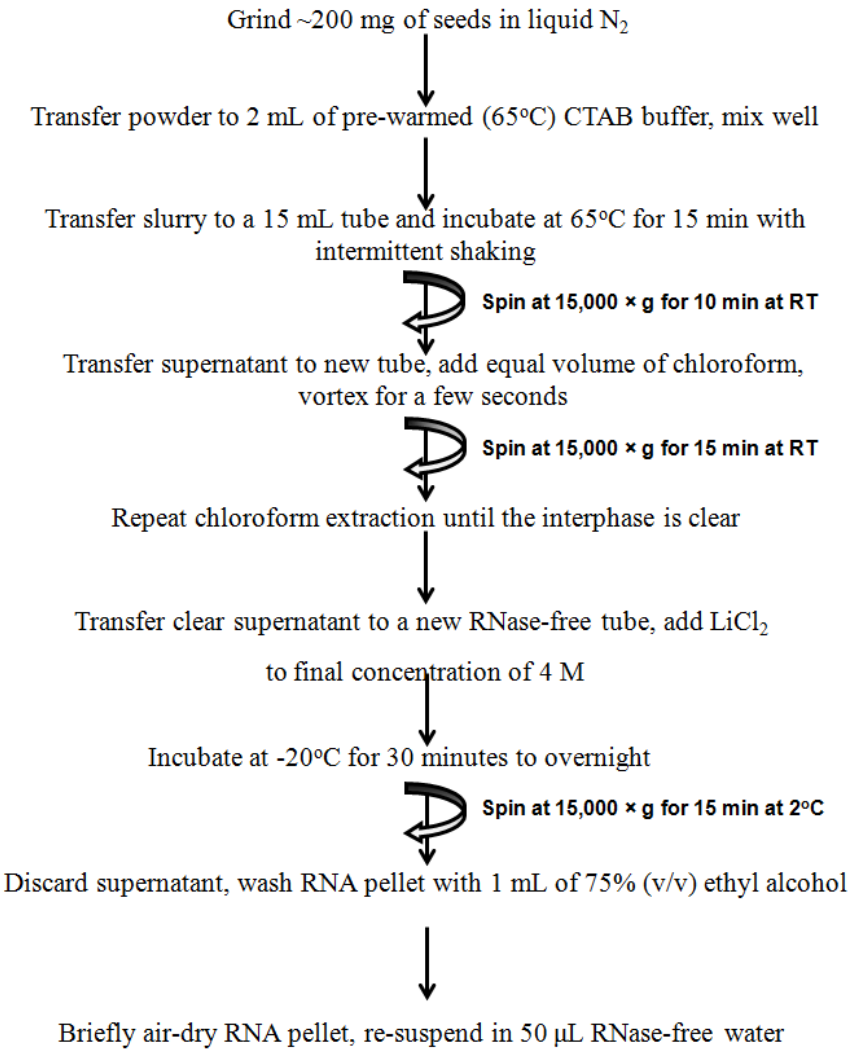


Figure 3.4: Flow-chart of optimized RNA isolation protocol from lentil seeds using the CTAB-LiCl method.

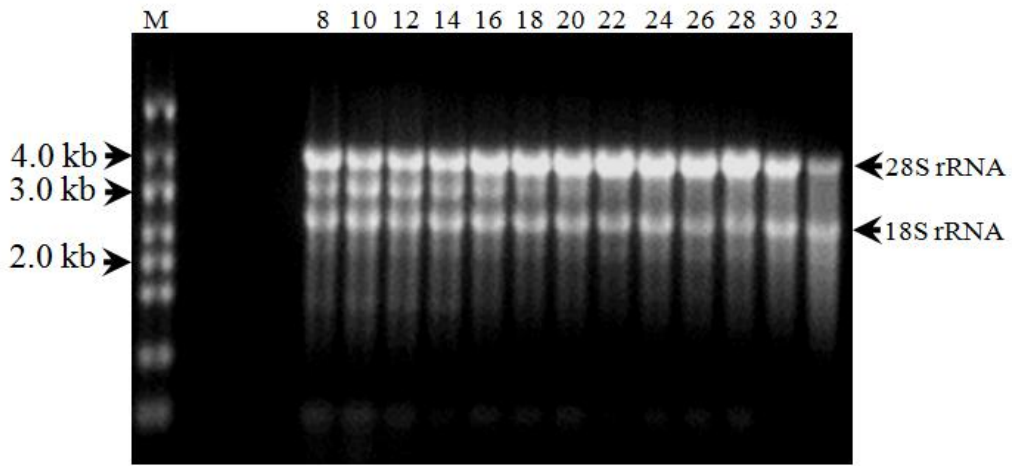


Figure 3.5: CTAB-extracted RNA from developing lentil seeds (8 – 32 DAF) resolved on formaldehyde agarose gel. M- RNA ladder.

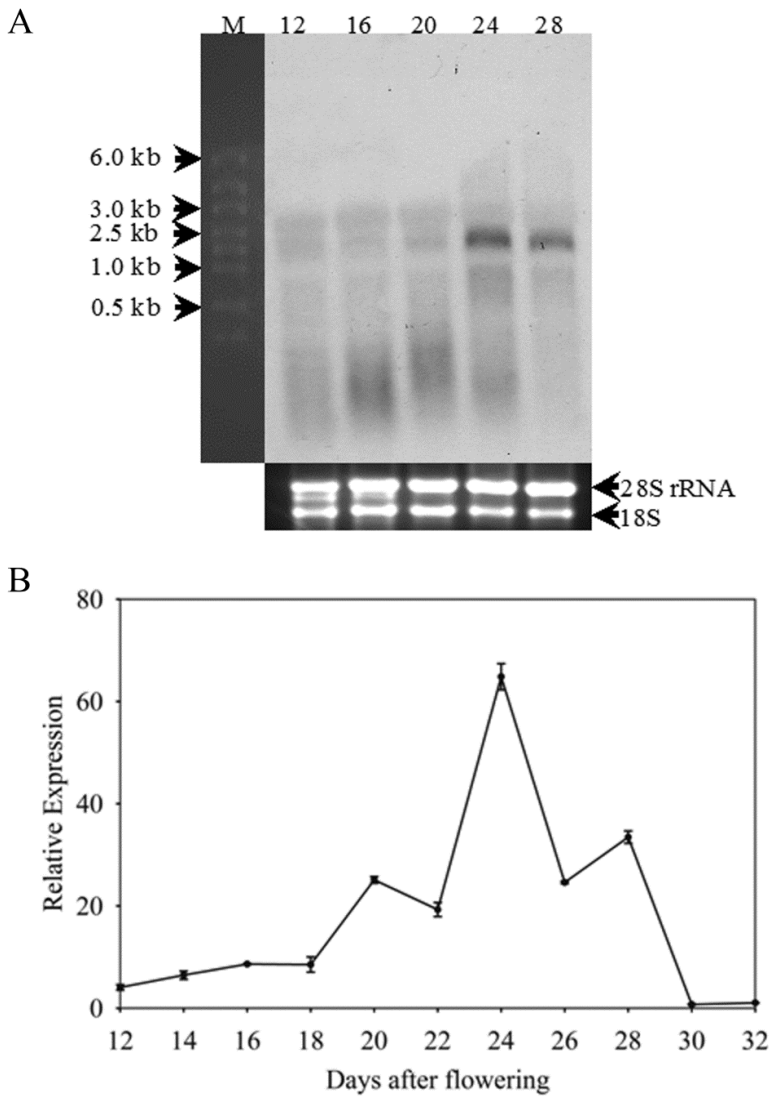


Figure 3.6: (a) RNA gel blot for *GbssI* expression from developing seeds of lentil from 12,16, 20, 24 and 28 days after flowering. (b) Relative expression of *GbssI* determined by QPCR from developing seeds of lentil from 10 to 32 days after flowering. Error bars indicate standard error of the mean of four determinations at each point.

applications such as RNA gel blot analyses and by DNase treatment for more sensitive applications such as QPCR analyses. We have successfully used this protocol for RNA isolation from developing wheat grains, with RNA yields ranging from 90 to 115 μg with *A*260/280 ratio of 2.2 from 100 mg of ground 6, 18 and 40 DPA kernels (Chibbar, personal communication). We foresee that the optimized RNA isolation protocol developed in this study will be valuable for RNA isolation from seeds of other pulse crops.

4. ISOLATION OF TWO GALACTINOL SYNTHASE ISOFORMS *LcGolS1* AND *LcGolS2* IN DEVELOPING LENTIL (*Lens culinaris* Medik. cv CDC Redberry) SEEDS.

Study 2

In this study, two genes coding for galactinol synthase (*LcGolS1*, *LcGolS2*) were isolated from a cDNA library constructed using RNA from developing lentil seeds (18-24 DAF). During seed development gene expression, enzyme activity and galactinol accumulation was studied.

***Kannan U,** Sharma R, Khedikar Y, Ganeshan S, Gangola M and Chibbar RN (2014). Isolation of two galactinol synthase isoforms *LcGolS1* and *LcGolS2* in developing lentil (*Lens culinaris* Medik.) seeds (to be submitted)

4.1 Abstract

Galactinol synthase (GS, EC 2.4.1.123) catalyzes the transfer of a galactosyl residue from UDP-galactose to *myo*-inositol to synthesize galactinol, a precursor for raffinose family oligosaccharide (RFO) biosynthesis. A cDNA library constructed with RNA isolated from developing lentil seeds, was screened to isolate cDNA clones for two isoforms of GS, *LcGolS1* (1336 bp, ORF-1002 bp, 334 amino acids) and *LcGolS2* (1324bp, ORF-975bp, 325 amino acids) with predicted molecular weights 38.7 kDa and 37.6 kDa. Gene expression analyses during seed development showed differential transcript accumulation with a maximum for *LcGolS1* and *LcGolS2* at 32 and 24 days after flowering (DAF), respectively. During seed development, GS showed maximum enzyme activity at 26 and 28 DAF. In developing lentil seeds, GS enzyme activity correlated with galactinol accumulation, which rapidly increased in concentration at 22 DAF with maximum accumulation at 26 DAF. During seed development, *myo*-inositol accumulated at early stages, but gradually decreased from 20 DAF to 32 DAF when galactinol concentration increased coinciding with increased GS enzymatic activity.

4.2 INTRODUCTION

Lentils which are traditionally consumed in Asia and Mediterranean countries, are also gaining more popularity in the western world. Lentils have a high protein content when compared to cereals and have essential amino acids including lysine which makes it a good substitute for animal protein. Consumption of pulses has been encouraged by health organizations as they reduce serum cholesterol levels, thereby reducing the occurrence of diabetes and coronary heart diseases (Leterme 2002; Chibbar et al 2010). However, the presence of anti-nutritional compounds like raffinose family oligosaccharides (RFO) contribute to reduced consumption of lentils in some parts of the world.

The three sucrosyl oligosaccharides raffinose, stachyose and verbascose are the three predominant RFO in lentils. Galactinol synthase (GS, EC 2.4.1.123) catalyzes the synthesis of galactinol by transferring a galactosyl residue from UDP- galactose to *myo*-inositol (Saravitz et al 1987; Keller and Pharr 1996; Peterbauer et al 2001a). Galactinol is the major galactosyl donor in the galactinol dependant pathway for RFO biosynthesis. A reversible reaction of raffinose synthesis is catalyzed by the enzyme raffinose synthase by a transfer of a galactosyl residue from galactinol to sucrose. Stachyose and verbascose are synthesized by the transfer of a galactosyl

residue from galactinol to raffinose and stachyose, respectively, catalyzed by enzymes stachyose synthase and verbascose synthase. The increase in GS activity prior to RFO accumulation in soybean seeds suggests that GS plays a regulatory role in carbon partitioning between sucrose and RFO (Saravitz et al 1987; Lowell and Kuo 1989). However, no significant correlation was observed between GS activity and RFO accumulation in developing pea seeds (Peterbauer et al 2001b). Similarly, in developing seeds of tomato, GS gene expression could not be correlated with raffinose accumulation (Downie et al 2003). In soybean, mutation in the *myo*-inositol phosphate synthase gene reduced RFO concentration in seeds, suggesting that *myo*-inositol is required for RFO accumulation (Hitz et al 2002). In pea seeds with contrasting RFO content, GS activity and *myo*-inositol concentration correlated with RFO accumulation (Karner et al 2004).

The presence of more than one isoform of GS has been reported in different plants. In Arabidopsis, three isoforms of GS were characterized in response to abiotic stress with *AtGolS1* and *AtGolS2* induced by drought and high salinity, and *AtGolS3* induced by cold stress (Taji et al 2002). Similar differential expression of GS genes in response to drought and salt were also reported in tomato, maize and coffee (Downie et al 2003; Zhao et al 2004; dos Santos et al 2011). These reports suggest different functional roles of GS in combating several abiotic stresses.

The concentration of RFO in mature lentil seeds has been previously reported (Tahir et al 2011), however, accumulation of RFO and its precursors galactinol and *myo*-inositol in developing lentil seeds have not been reported. To improve our understanding of RFO biosynthesis in developing lentil seeds, two cDNA clones for *GolS* were characterized and their expression during seed development was analyzed.

4.3 MATERIALS AND METHODS

4.3.1 Plant Material

Lens culinaris Medik. cv CDC Redberry (Vandenberg et al 2006) plants were grown under controlled conditions in a growth chamber with an 18 h photoperiod [250 mmol m^{-2} s^{-1} PPFD (photosynthetic photon flux density)] at 21 ^{o}C/15 ^{o}C day/night temperatures. Plants were fertilized every two weeks with NPK 20:20:20. Young leaves (two weeks old) were collected and stored at -80 ^{o}C for genomic DNA extraction. Completely opened flowers were tagged and developing seeds were collected at two day intervals from 10 days after flowering (DAF) to 38

DAF and stored at -80 ^{o}C. To avoid variation in sucrose concentration, sample collections were done at the same time (3:00 - 5:00 pm) of the day during the entire collection period.

For the reference gene normalization study, plants were germinated in sterile petri dishes for three days. The germinated seeds were transferred to Magenta boxes containing half strength Hoagland's solution (Hoagland 1918) and sterile glass beads (1 - 5mm). Plants were harvested for root, stem and leaf collection. Leaf samples were collected every 5 h continuously for 30 h for gene expression normalisation studies along circadian rhythm.

4.3.2 Isolation of gene sequences for *LcGolS*1 and *LcGolS*2

Genomic DNA was extracted from frozen leaf tissues (200 mg) using the CTAB method (Doyle and Doyle 1990). Partial sequences of *LcGolS1* and *LcGolS2* were obtained by amplification of genomic DNA with degenerate primers. Degenerate primers were designed at conserved regions by aligning nucleotide sequences of *Medicago sativa* (AY126615.1), *Pisum sativum* (AJ243815), *Ammopiptanthus mongolicus* (DQ519361.1) and *Glycine max* (AY126715.1). An oligonucleotide primer pair DG/GS_F1 (forward) and DG/GS_R1 (reverse) (Appendix 4.2) was used in a PCR reaction comprised of 20 mM Tris-HCl (pH 8.4), 50 mM KCl, $MgCl_2$ concentrations (1.5mM, 2.5 mM, 3.5 mM), primer concentrations (0.2 µM and 0.4 µM), 0.2mM each dNTP and 3 units Taq Polymerase (Life Technologies Inc., Burlington, Ontario, Canada). The amplification conditions were: initial denaturation at 95 ^{o}C for 4 min 15 sec, followed by 32 cycles of 95 ^{o}C for 45 sec, 20 sec annealing temperature (tested at 54 ^{o}C, 57 ^{o}C and 60 ^{o}C), 1 min 30 sec extension at 72 ^{o}C and final cycle at 72 ^{o}C at 10min. The amplified fragments of sizes 312 bp and 400 bp were analyzed by agarose (1% w/v) gel electrophoresis (Appendix 4.4). The amplified fragments were DNA sequenced to confirm sequence identity. The isolated partial sequences of *LcGolS1* and *LcGolS2* were searched for sequence similarity against *Lens culinaris* Medik. transcriptome assembly (Kaur et al 2011) to obtain GS sequences. Primers were designed for *LcGolS1* based on the sequence obtained from lentil transcriptome assembly (Kaur et al 2011), amplified and cloned using a CloneJet PCR cloning Kit (Thermo Fisher Scientific Inc., Pittsburgh, PA, USA). Plasmids were digested with restriction enzymes *BsrG1* and *Tth111I* to obtain a 848 bp fragment for use as a probe to screen the cDNA library.

4.3.3 RNA Isolation, construction and screening of cDNA library

A CTAB LiCl based RNA isolation method (Kannan et al 2014) was used for RNA isolation from developing seeds (5 g). The total RNA was purified using PureLink RNA Mini Kit (Life Technologies Inc., Burlington, Ontario, Canada) as per manufacturer's instructions. To construct a cDNA library, 5 µg of Poly(A)$^+$ RNA was purified from total RNA isolated from developing lentil seeds (18, 20, 22 and 24 DAF) using PolyATract® mRNA Isolation System I (Promega Corporation, USA). The cDNA library was constructed using a λ ZAP-cDNA® Gigapack III gold cloning kit (Agilent Technologies Inc., Santa Clara, CA) as per manufacturer's instructions. All the aliquots of amplified cDNA library were stored in DMSO (7% v/v) at - 80 °C. For library insert determination, a total of 50 plaques were randomly cored from the agar plate containing the amplified cDNA library (1 x 10^9 pfu/mL) for *in vivo* excision to obtained phagemids. The phagemids were digested with *EcoR* I and *Xho* I to determine insert lengths.

The cDNA library (1 x 10^9 pfu/mL) was screened using radiolabelled ([α^{32}-P] dCTP) partial GS gene 848 bp (Probe 1; Figure 4.1A) as per manufacturer's instructions (Agilent Technologies Inc., Santa Clara, CA). Plaque lifting and hybridizations were done as per standard protocols (Sambrook and Russell 2001). Primary screening resulted in 72 phagemids and 12 phagemids were selected for secondary screening based on amplification using *LcGolS1* gene specific primers GS1_1F and GS1_1R (Appendix 6.3). Secondary screening resulted in 16 positive phagemids and these phagemids were plaque purified and recovered by *in vivo* excision as per manufacturer's instructions (Agilent Technologies Inc., Santa Clara, CA). For *LcGolS2* full length sequences were obtained by GeneRacer® Kit with SuperScript® III RT (Life Technologies Inc., Burlington, Ontario, Canada). 5' UTR and 3'UTR sequences were amplified using primers reverse GS2_RACE/R and forward GS2_RACE/F (Appendix 6.3).

4.3.4 Phylogenetic Analysis

The sequences of GS genes from other plant species were obtained from GenBank, NCBI. Sequence alignments were performed using Clustal Omega (Sievers et al 2011). A phylogenetic tree was constructed using MEGA 6 (Tamura et al 2013). Both the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method and the Poisson correction method were used to calculate evolutionary relationship and distances. The phylogenetic tree was constructed using of *LcGolS1*, *LcGolS2* and *GolS* polypeptide sequences from several plant species (Table 4.1).

4.3.5 Protein Structure prediction for LcGolS1 and LcGolS2

Amino acid sequences of LcGolS1 and LcGolS2 were provided as inputs in I-Tasser (Roy et al 2010) software for three dimensional structure predictions of LcGolS1 and LcGolS2. Predicted models were provided as inputs in Consurf (Celniker et al 2013) to obtain conserved residues in predicted structure models for LcGolS1 and LcGolS2. Predicted structures and superposition of LcGolS1 and LcGolS2 were performed using the PyMOL Molecular Graphics System, Version 1.7.2 (Schrödinger, LLC, New York, USA).

4.3.6 Transcript accumulation

4.3.6.1 RNA gel blot

Total RNA (20 µg) was electrophoresed on a 1% agarose (w/v) gel containing 0.66 M formaldehyde in HAE buffer (0.2 M Na-HEPES, 50 mM NaAc, 10 mM EDTA, pH 7.0) for 3-4 h at 100V. Ethidium bromide was added to the sample and visualized under UV to estimate RNA concentration and its integrity. RNA was transferred from agarose gel to a Hybond-N^+ membrane (GE Healthcare, Baie d'Urfe, Quebec, Canada) using a PosiBlot^{TM} apparatus (Agilent Technologies Inc., Santa Clara, CA, USA) for 2 h at 75 mm Hg pressure. RNA was cross-linked to the membrane in a UV cross-linker (UVP Inc., Upland, CA, USA). The membrane was hybridized in Church buffer (0.5 M Na_2HPO4, SDS (7% w/v) and 1 mM EDTA) at 65 ^oC with the 848 bp probe towards the 3' end of the *LcGolS1* labelled with [α^{32}-P] dCTP. Membrane was rinsed twice with 1 x SSPE (150 mM NaCl; 10 mM NaH_2PO_4. H_2O; 1 mM EDTA; SDS [0.1% w/v]) for 10 min each at room temperature, and then washed twice in 0.1× SSPE, SDS (0.1%, w/v) at 65 ^oC, until minimal background activity was observed. The membrane was exposed to Biomax MR film (KODAK® BioMax® Maximum Resolution (MR) Autoradiography Film) at -80 ^oC for varied intervals based on the signal intensities observed.

4.3.6.2 Quantitative real time PCR analysis

A total of ten genes were selected based on the commonly used genes for *Cicer arietinum* (Garg et al 2010), *Pisum sativum* (Die et al 2010), *Glycine max* (Hu et al 2009),

Table 4.1: Amino acid sequences used for phylogenetic analysis of *GolS* genes.

Plant	Gene	Accession Number	Amino acid length
Arabidopsis thaliana	GolS1	NP_182240.1	344
	GolS2	NP_176053.1	335
	GolS3	NP_172406.1	334
	GolS4	NP_176250.1	334
Boea hygrometrica	GolS	ACI62176.1	334
Coffea arabica	GolS1	ADM92588.1	338
	GolS2	ADM92589.1	334
	GolS3	ADM92590.1	344
Fagopyrum esculentum	GolS1	AAM96870.1	333
Fagopyrum esculentum	GolS2	AAM96868.1	354
Fagopyrum esculentum	GolS3	AAM96869.1	255
Glycine max	GolS	AAM96867.1	328
Medicago truncatula	GolS	AES82223.1	339
Populus trichocarpa	GolS1	ACA04027.1	338
	GolS2	ACA04030.1	336
	GolS3	AAQ07250.1	337
Pisum sativum	GolS	CAB51130.1	334
Solanum lycopersicum	GolS1	AAO72744.1	318
Zea mays	GolS1	AAQ07248.1	344
	GolS2	AAQ07249.1	348
	GolS3	AAQ07250.1	345

(Paolacci et al 2009) and *Zea mays* (Manoli et al 2012). Since these genes were not annotated in lentil, genes previously annotated from *C. arietinum, P. sativum* and *G. max* were searched against the lentil EST database to identify reference gene sequences from lentil. Sequences that showed high similarity were selected and primers were designed with lengths varying from 20 to 24 oligonucleotides and product length of 50-200 bp and melting temperatures of 60 ^{o}C ± 1 ^{o}C using Primer 3 software (Untergrasser et al 2012). For reference gene validation, reference gene stability was calculated using GeNorm algorithm (Vandesompele et al 2002). Amplified partial gene sequences were cloned and DNA sequenced to confirm gene identity.

For qPCR analysis, total RNA (5 µg) was DNase treated and 1 µg was used for cDNA synthesis. Single stranded cDNA was synthesized using 200 U superscript III (Life Technologies Inc., Burlington, Ontario, Canada), 500 ng OligodT and 900 ng of random primer. PCR amplification was set up using 5 µL of 1:15 dilution of cDNA, 300 nM primers and 2X Maxima^{TM} SYBR green qPCR master mix (Thermo Fisher Scientific Inc., Pittsburgh, PA, USA). Two controls were used: a no template control; and a no reverse transcriptase control. The presence of a single product was confirmed from the melt curve obtained from the dissociation step after PCR amplification. Gene specific primers (Appendix 4.3) for *LcGolS1* GS1_RT_F (forward) and GS1_RT_R (reverse); and *LcGolS2* GS2_RT_F (forward) and GS2_RT_R (reverse) were used (Appendix 6.3). Reference genes glyceraldehyde-3-phosphate dehydrogenase (GAPD) GAPD_RT_F and GAPD_RT_R; heat shock protein (HSP) HSP_RT_F and HSP_RT_R; elongation factor 1-α (EF1-α) EF_RT_F and EF_RT_R were used for validation (Appendix 6.3). The geometric mean of Ct values of reference genes was used to calculate relative expression (Vandesomple et al 2002). Relative expression of *LcGolS1* and *LcGolS2* was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak et al 2001) employing the 12 DAF sample as the calibrator. Experiments were repeated twice with two technical replicates.

4.3.7 Determination of galactinol concentration and galactinol synthase enzyme assay

Galactinol, *myo*-inositol, glucose/galactose concentrations were determined using high performance anion exchange chromatography with pulse amperometric detection (HPAEC-PAD). Developing seed samples were freeze dried, ground and 100 mg of sample was used for extraction of soluble sugars. Extraction, purification of soluble sugars, sample preparation,

chromatographic conditions and determination of oligosaccharide concentrations were performed as described in Gangola et al 2014.

GS activity was determined based on previously published methods (Peterbauer et al 1998; 2001b, Hitz et al 2002). The enzyme activity procedure optimized for pH, protein concentration, temperature and substrate concentration in *Cicer arietinum* (unpublished data; manuscript in preparation) was used in this study. Approximately 200 mg of sample was ground in liquid nitrogen and 2 mL of extraction buffer [50mM HEPES (2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid)-NaOH, pH 7.0 and 5 mM of 2-β mercaptoethanol] was added. The extract was transferred to a centrifuge tube and centrifuged at 12000 x g for 10 min at 4 ^{o}C. The supernatant was purified using a Sephadex G25 columns pre-equilibrated with 5 mL extraction buffer. Protein concentration was determined by a dye binding assay (Bradford 1976) using a commercial kit (Bio-Rad protein assay, Bio-Rad Laboratories Inc., Mississauga, Ontario, Canada) with bovine serum albumin as the standard. Ten micrograms of crude protein extract was incubated at 25 ^{o}C for 10 min with substrates *myo*-inositol (20 mM), UDP-galactose (10 mM), dithiothreitol (10 mM) and $MnCl_2$ (5 mM). Two controls, (i) protein without substrate, and (ii) substrate without protein were included for each assay set. Enzyme reactions were terminated by adding 50 µL of 95 % ethanol and samples were boiled for one min. The tubes were centrifuged at 12000 x g for 10 min and supernatant was transferred to a new tube. Ten milligrams of Dowex 50WX8 100 mesh ion exchange resin (Sigma-Aldrich Inc., St. Louis, Missouri, USA) was added to the sample and shaken at 200 rpm for 30 min. The tubes were centrifuged at 12000 x g for 10 min and the supernatant was transferred to a new tube and dried using a Speedvac concentrator and universal vacuum system (Thermo Savant, Holbrook, NY, USA). The dry pellet was dissolved in 500 µL water followed by centrifugation at 12000 x g for 10 min. The RFO extract was diluted (125 µL supernatant and 375 µL water) and analyzed by HPAEC-PAD as described by Gangola et al (2014). Any galactinol present in the crude protein extract or substrate was detected in the control and subtracted for galactinol concentration determination.

4.4 RESULTS AND DISCUSSION

4.4.1 Lentil cDNA library

A cDNA library was constructed from developing seeds from stages 18 - 24 DAF of *Lens culinaris* cv CDC Redberry (Vandenberg et al 2006). The titer of the primary library was 8.5 x 10^5 pfu/mL (~1 x 10^6 pfu/mL) and that of the amplified library was 1.0 x 10^9 pfu/mL. Insert size was determined by digesting *in vivo* excised phagemids from 50 random plaques cored from agar plates containing the amplified cDNA library. Digestion patterns of the plasmids for insert size showed sizes of 500 to 3000 bp (data not shown).

4.4.2 Amplification of DNA sequence corresponding to *GolS1* used as probe for cDNA library screening

Two amplicons of 312 (*LcGolS1*) and 400 bp (*LcGolS2*) were obtained by amplification using degenerate primers (Appendix 4.4). A sequence identity search of these two amplicons with the GenBank database showed high sequence identity (97%) to previously annotated GS, suggesting that both the amplicons were partial GS genes. The amplified sequences showed 80% sequence identity with each other, suggesting that the partial gene sequences could be two GS genes. The 302 bp *LcGolS1* nucleotide sequence showed high sequence identity (93%) with *Medicago sativa* GS and 93% identity with *Medicago sativa* subsp. falcate GS mRNA. The 400 bp *LcGolS2* partial sequence showed 80% sequence identity with *P. sativum* mRNA *GolS*. However, the sequence also showed presence of an intron in the sequence based on sequence alignment with *Pisum sativum GolS1*. *LcGolS2* partial sequence amplified from cDNA revealed the presence of a 90 bp intron in the partial *LcGolS2* amplified from genomic DNA (Fragments amplified from genomic DNA and cDNA shown in Appendix 4.5). A sequence search for enzyme active sites with the sequences *LcGolS1* and *LcGolS2* using Pfam (Punta et al 2014), showed the presence of active sites for glycosyl transferase family 8 which includes several enzymes including 1-alpha-galactosyltransferase.

During the course of this study, a *Lens culinaris* transcriptome shotgun transcriptome assembly was released (Kaur et al 2011) and made available in GenBank database. A 1324 bp nucleotide sequence (contig02236) was found by sequence similarity search of partial *LcGolS1* against *L. culinaris* transcriptome shotgun sequence assembly available in the database. The sequence was 100% identical to the partial *LcGolS1* obtained using degenerate primers. The amplified partial sequence was used as a probe to screen the cDNA library constructed using mRNA from developing lentil seeds (18-24 DAF).

4.4.3 Isolation of *LcGolS1* from lentil cDNA library

A total of 1.8 x 10^5 phage particles were screened (18% of the cDNA library) and yielded 72 positive clones. Twelve positive clones confirmed for insert presence by PCR were used for secondary screening based on the results obtained through amplification of gene specific primers. Plasmids were recovered from the bacteriophages for 16 clones obtained after secondary screening. After *in-vivo* excision of the plasmids, three colonies were picked from each plate and DNA sequenced. One plasmid from each triplicate was digested with restriction enzymes *EcoRI* and *XhoI* to confirm the presence of the insert and was sent for DNA sequencing (Appendix 4.6).

A total of 48 plasmids were sent for DNA sequencing and 43 clones showed 92% sequence identity to the existing GS sequences in the GenBank database. The nucleotide sequence showed 93% similarity to *Pisum sativum GolS* sequence. Nucleotide sequencing results obtained for *LcGolS1* clones were categorized into three variants based on sequence lengths:

(i) variant 1- 5'UTR-118 bp, ORF- 1002 bp and 3'UTR-10 bp followed by polyA (22 clones);

(ii) variant 2- 5'UTR- 55bp, ORF- 1002 bp and 3'UTR- 203 bp followed by polyA (18 clones);

(iii) variant 3-5'UTR- 55bp, ORF- 1002 bp and 3'UTR-279bp followed by polyA (3 clones).

All three variants (43 clones) obtained for *LcGolS1* had identical open reading frames. However, differences in the lengths of cDNA clones were due to variations in 5' and 3' UTR lengths.

The isolated *LcGolS1* with ORF 1002 bp coded for 334 amino acids with the predicted molecular weight of 38.8 kDa with a pI of 5.79 (Figure 4.1A; Appendix 4.7). Full length clone of *LcGolS1* sequence showed 91% sequence identity to *Pisum sativum* GolS mRNA and 95% similarity to *P. sativum* GolS polypeptide. The DxD motif, the manganese ion binding site and the HxxxGxKPW substrate binding site were conserved in LcGolS1 (Figure 4.1B).

4.4.4 Isolation of *LcGolS2* from lentil cDNA library

The partial nucleotide sequence of *LcGolS2* was used to isolate a full length sequence for *LcGolS2* using 5' and 3' RACE (Rapid amplification of cDNA ends; Primer positions indicated in Appendix 4.8). A full length 1324 bp sequence of *LcGolS2* was obtained with an ORF of 975

bp coding for 325 amino acids, with a predicted molecular weight of 37.6 kDa and a pI 5.67 (Figure 4.1A; Appendix 4.8). Isolated *LcGolS2* showed 87% identity to *Medicago truncatula* GolS mRNA and 89% similarity to *Medicago truncatula* GolS polypeptide. Conserved domains DxD motif and the manganese ion binding site and substrate binding domains site HxxxGxKPW were also conserved in LcGolS2.

Sequence alignment of *LcGolS1* and *LcGolS2* showed 76% identity to nucleotide sequences and 73% similarity between LcGolS1 and LcGolS2 polypeptides. *LcGolS2* was devoid of eleven amino acids present in *LcGolS1* towards the 3' end of the sequence (Figure 4.1A, B). A similar deletion was observed in the *G. max* sequence (data not shown). The hydrophobic penta-peptide APSAA conserved in GS was also conserved in *LcGolS1*, however a change from serine to asparagine was observed in *LcGolS2* (APNAA) (Figure 4.1A, B). Though the hydrophobic penta-peptide has been conserved in GS its function is yet to be elucidated.

4.4.5 Alternative polyadenylation in 3' UTR of *GolS* transcripts might contribute to difference in UTR lengths

LcGolS1 clones (43 clones) isolated from the cDNA library showed three variants based on differences in the 5' and 3' UTR lengths of the clones (Appendix 4.9). 3'UTR of different lengths: 10 bp (variant 1), 203 bp (variant 2), and 279 bp (variant 3) were obtained (Figure 4.2A). Variant 3 showed a change of G to T in the 53rd nucleotide in 3'UTR (Figure 4.2A). Similar differences in UTR lengths were also obtained for *LcGolS2*, where the length of 3'UTR after the stop codon was 182 bp, 211 bp and 240 bp followed by poly (A) (Figure 4.2B). 5'UTR length of *LcGolS2* was 88 bp. No nucleotide changes were observed in the open reading frame and 3'UTR of *LcGolS2* sequences.

An alternative polyadenylation site could be the reason for the differences in UTR lengths observed in the two cDNA clones. Though this factor has been overlooked for several decades, increasing cDNA library mining has focused on the regulatory roles of 3' UTR. Recent studies in *Arabidopsis* and *Medicago* have reported differential polyadenylation sites in the 3'UTR region of transcripts (Wu et al 2011; Wu et al 2014). In Arabidopsis the majority of polyadenylation clusters [poly A sites located within 24 nucleotides in the same gene] were

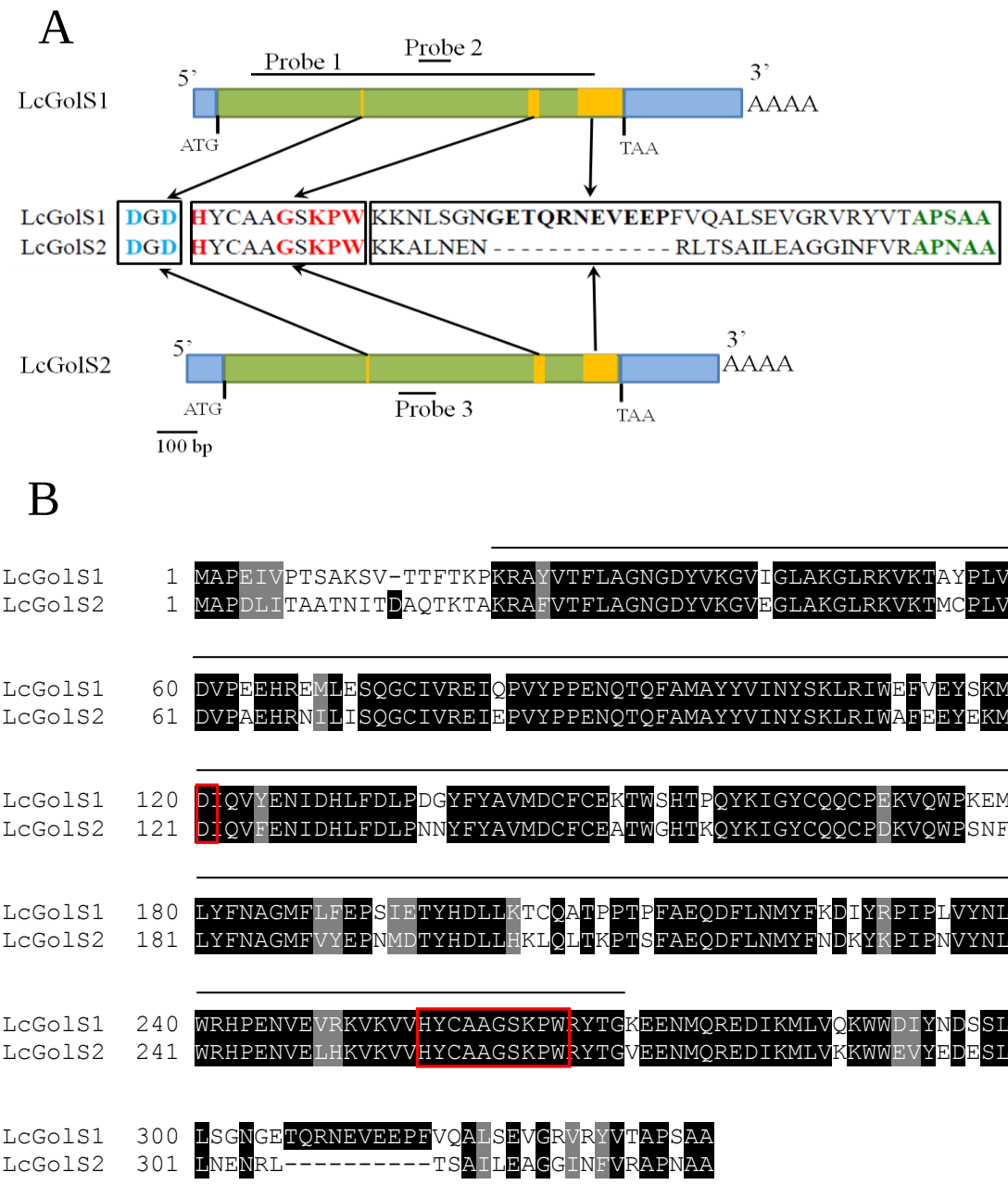


Figure 4.1: (A) Schematic representation of *LcGolS1* and *LcGolS2* cDNA clones. Open reading frame are represented in green, UTR regions in blue and important domains are represented in yellow in the pen reading frame of *LcGolS1* and *LcGolS2*. The amino acid residue DxD (indicated in blue), HxxxGxKPW (indicated in red), APSAA (indicated in green) constitute to manganese binding domain, substrate binding site and the hydrophobic penta peptide. Probe 1 indicates nucleotide sequence (848 bp in length) used as the probe for RNA gel blot analysis and cDNA library screening. Probe 2 and probe 3 indicate nucleotide sequences used for qRT-PCR analysis. **(B)** Alignment of deduced amino acid sequences of *LcGolS1* and *LcGolS2*. The amino acids highlighted black are highly conserved and residues highlighted grey indicate similar amino acids. The conserved DxD represent the manganese ion binding site, the HxxxGxKPW motif is the substrate binding site indicated by red boxes. The black line indicates the glycosyl transferase domain.

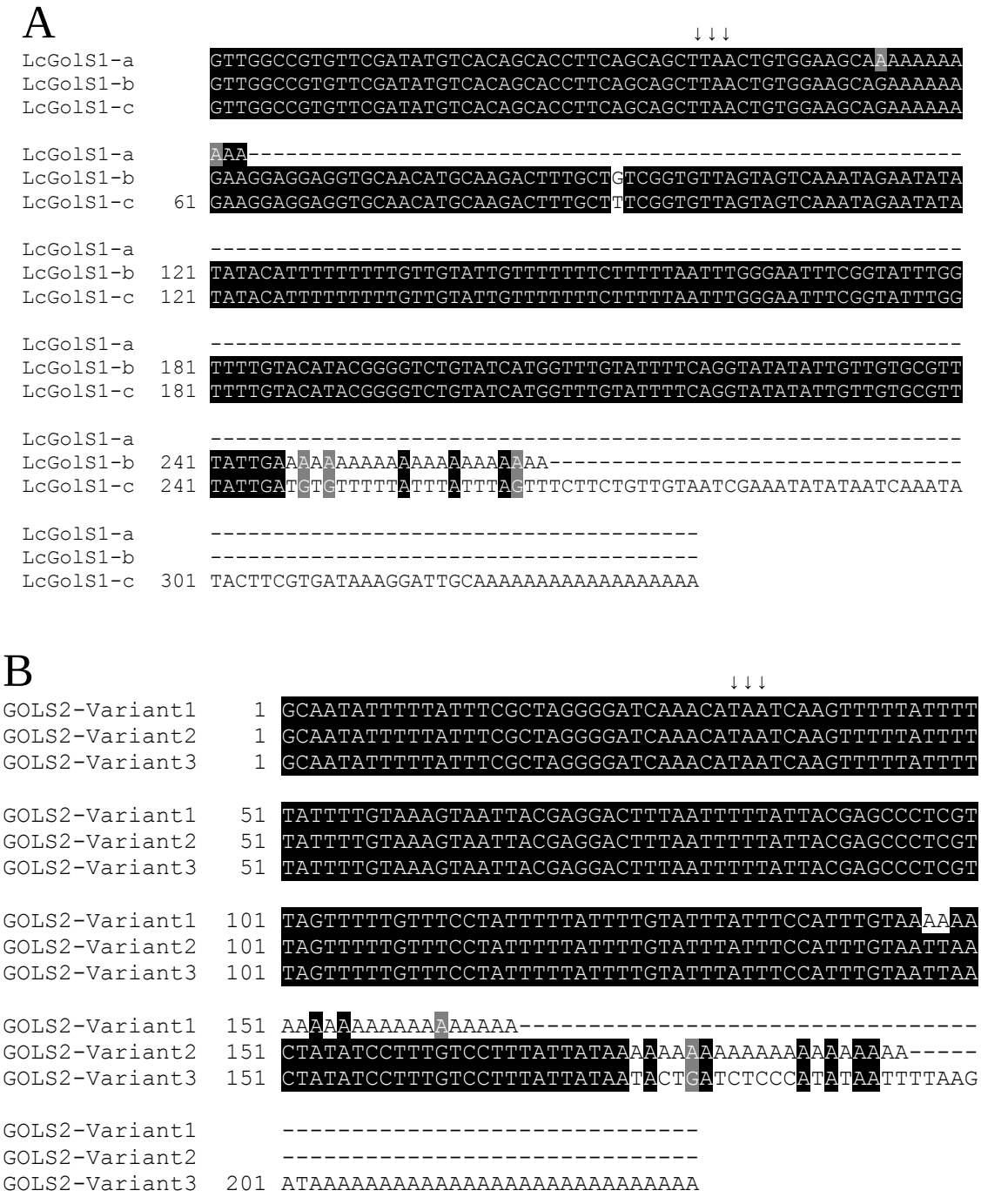


Figure 4.2: Sequence alignments of partial *LcGolS1* (A) and *LcGolS2* (B) nucleotide sequences with differences in 3' UTR lengths. Identical nucleotides are highlighted in black and similar nucleotides are highlighted in grey. Translation stop codon TAA is indicated by arrows.

reported in the 3'UTR regions. Also switching of poly A sites was also observed in genes based on the tissues in which they are expressed (Wu et al 2011). The seeds contained shorter 3' UTR compared to transcripts in the leaves (Wu et al 2011). Introns present in 3' UTR with classical polyadenylation signals could promote differential use of polyadenylation signals based on tissues where they are expressed (Wu et al 2011). Alternative polyadenylation sites present in the 3' UTR regions might contribute to regulation of mRNA by RNA- binding proteins or micro RNA (Wu et al 2014). Presence of shorter 3' UTR sequences might provide few chances of gene repression. Shorter transcripts reduce chances of transcript regulation whereas longer transcripts could be regulated by various post transcriptional mechanisms such as miRNA binding to 3' UTR sequences (Wu et al 2011).

4.4.6 Phylogenetic analysis of GolS sequences

Phylogenetic analysis of GolS sequences resulted in segregation of six clades A-F (Figure 4.3). *LcGolS1* in clade A aligned with *Pisum sativum, Medicago truncatula, Cicer arietinum, Arabidopsis thaliana* and *Fagopyrum esculentum GolS1* showing higher similarity to *Pisum sativum GolS1*. *LcGolS2* clustered in clade C with *Glycine max GolS, Boea hygrometrica GolS* and *Coffea arabica* isoform *GolS2*. *Coffea arabica GolS2* was expressed in both leaf and endosperm, however expression was detected only under severe drought and salt stress (dos Santos et al 2011). *Boea hygrometrica GolS* was also expressed during drought stress and gene expression was also induced by abscisic acid (Wang et al 2009b). With *LcGolS2* present in a group of *GolS* induced by abiotic stresses, suggests that *LcGolS2* might be stress inducible gene. However, experimental validation is necessary to understand the role of *LcGolS2* in abiotic stress.

In *A. thaliana,* three isoforms of *GolS* were identified. *AtGolS1* and *AtGolS2* were induced by drought and high salinity stress (Taji et al 2002) whereas, *AtGolS3* was induced during cold stress. In *Arabidopsis* transcription factors HSF (heat shock factors) and DREB1A/CBF3 (dehydration-response element binding factor 1A/cold responsive element binding factor 3) activated *AtGolS1* and *AtGolS3* expression (Panikulangara et al 2004; Ogawa et al 2007). *Boea hygrometrica* has four cis-elements in *BhGolS1* promoter which can be bound by WRKY gene. *BhGolS1 g*ene expression was induced during desiccation tolerance in leaves (Wang et al 2009b). Expression of GS genes in response to various stress responses and

elucidation of promoter sequences of *LcGolS2* might also provide more insights in the role of *GolS2*.

4.4.7 In-silico protein structure prediction of *LcGolS1* and *LcGolS2*

GS belongs to the GT8 (Glycosyl transferase family 8) class family of proteins. The GT8 class can be further divided into class I and class II proteins. Class I proteins are involved in cell wall synthesis (α-galactouronosyltransferase and galactouronosyl transferase like proteins), whereas, class II includes GS and plant glycogenin- like starch initiation proteins which are not related to cell wall synthesis (Yin et al 2010). All glycosyl transferase genes contain the GT-8 domain and motifs DxD, HxGxxKPW which are conserved across all genes in the GT-8 family (Figure 4.1B).The motifs conserved across clade II includes YNL, FAEQDFLNMYF, YFNAGMFLFEPS and KLRIWEFVEY (Yin et al 2010). Glycosyl transferases were classified based on the protein structure into GT-A and GT-B folds. GT-A fold proteins contain the presence of α/β/α which are very similar to a Rossmann fold type, whereas GT-B contain two Rossmann folds linked with catalytic sites present between the folds (Breton et al 2006). GT-C type fold is a recent addition to the existing classification and was reported with the crystal structure of sialyltransferase (CstII) (Chui et al 2004).

Search for conserved domain using Pfam (Punta et al 2014) for *LcGolS1* and *LcGolS2* showed the presence of GT-8 domain (PF01501) with conserved residues aligning from 24-267 for both the isoforms (Figure 4.1B). All conserved motifs were present in the GT-8 domain of *LcGolS1* and *LcGolS2*. No crystal structures are available for GS and three dimensional structures of GS could be predicted by 'threading method' based on fold recognition methods with structures of similar proteins in the PDB library. Three dimensional protein structure models were predicted based on threading alignments [I-Tasser Server (Roy et al 2010)] with rabbit muscle glycogenin (pdb 1LL0; Gibbons et al 2002). Normalized Z score, an indicator of the quality of alignment was 1.86 for LcGolS1 and 1.92 for LcGolS2, where values greater than 1 indicate good alignments. C- score for predicted protein structures for LcGolS1 and LcGolS2 were -0.86 and -0.69. C- score values range from -5 to 2 where higher scores indicate better structural models (Roy et al 2010). Predicted structures of LcGolS1 viewed through PyMol contained 7 β-sheets and 15 α- helices and LcGolS2 contained 3 β-sheets and 15 α- helices (Figure 4.5 A, B). The threading alignments with rabbit glycogenin align only in the Glycosyl

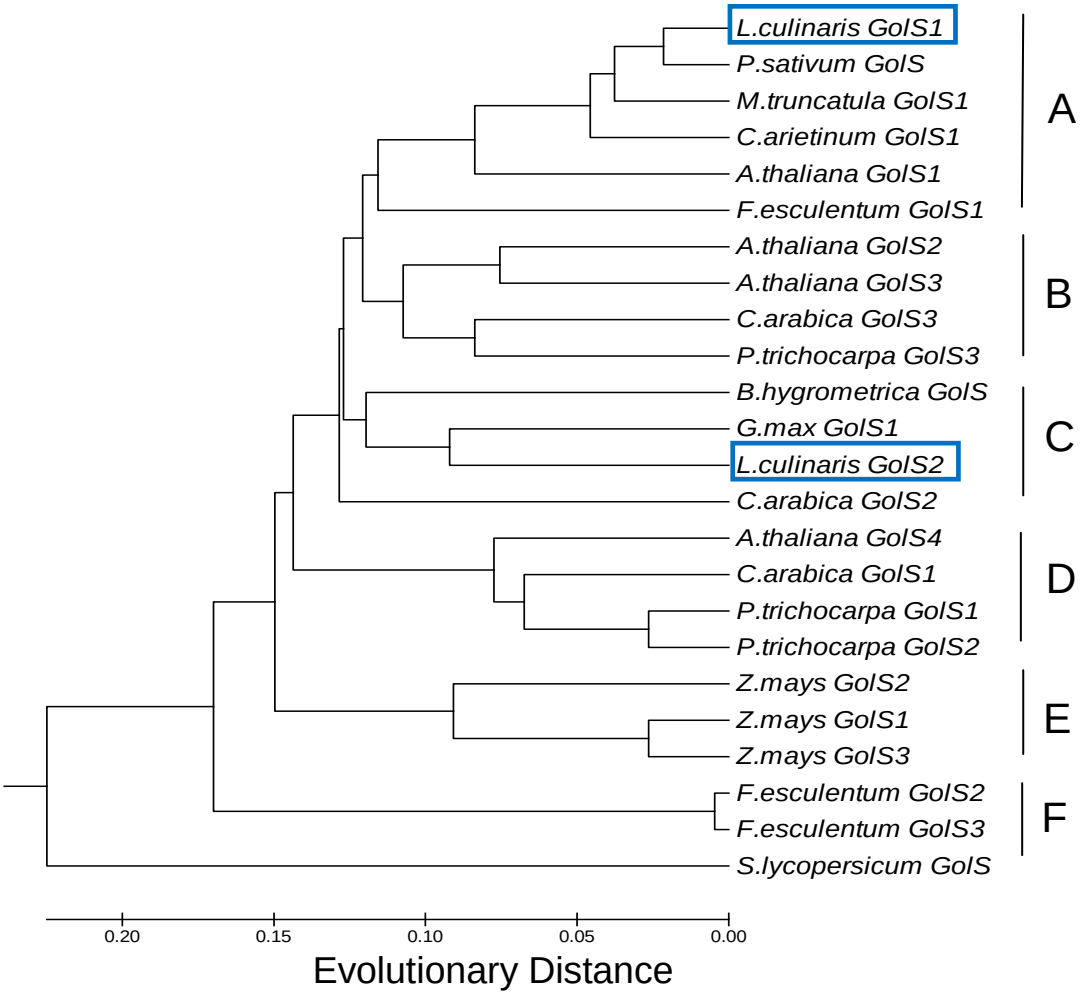


Figure 4.3: Phylogenetic analysis of galactinol synthase genes based on putative protein sequences of *LcGolS1* and *LcGolS2*. Branch lengths indicate evolutionary distances between the sequences.

A

```
GolS1       MAPEIVPTSAKSVTTFTKPKRAYVTFLAGNGDYVKGVIGLAKGLRKVKTAYPLVVAVLPDVPEEHREMLESQGCIVREIQPVYPPENQTQFAM
SecStr      CCCCCCCCCCCCCCCCCCCCCSSSSSSSSCHHHHHHHHHHHHHHHHHCCCCCSSSSSCCCCCHHHHHHHHHCCCCSSSSSCCCCHHHCCCCC
Score       999876876666667898871899999508727999999999999985899888999779987888999998089417971024733321233 4

GolS1       AYYVINYSKLRIWEFVEYSKMIYLDGDIQVYENIDHLFDLPDGYFYAVMDCFCEKTWSHTPQYKIGYCQQCPEKVQWPKEMGQPPSLYFNAGM
SecStr      CCHHHHHHHHHHCHHCCCCSSSSSCCCCCCCCCHHHHHCCCCCCSSSSCHHHCCCCCCCCHHHHHHHHHHCHHHHCCCCCCCCCCCCCCCCSS
Score       304667599875410262359993698410278899851864314530122202200145044445654274653253002567644265437

GolS1       FLFEPSIETYHDLLKTCQATPPTPFAEQDFLNMYFKDIYRPIPLVYNLVLAMLWRHPENVEVRKVKVVHYCAAGSKPWRYTGKEENMQREDIK
SecStr      SSSSCCHHHHHHHHHHHHHCCCCCCCCHHHHHHHCCCSSSSCCHHHHHCCCCCCCCCHHHHHHCCCSSSSSCCCCCCCCCCCCCCCCCCCCHHH
Score       999587999999999998428988531378887637847978705520111000461556524747998458999899888876674410189

GolS1       MLVQKWWDIYNDSSLDYKKNLSGNGETQRNEVEEPFVQALSEVGRVRYVTAPSAA
SecStr      HHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHCCCHHHHCCCCCCCCCCCCCCCC
Score       999999999987577776411231144554422612232143578655689998 9
```

B

```
GolS2       MAPDLITAATNITDAQTKTAKRAFVTFLAGNGDYVKGVEGLAKGLRKVKTMCPLVVAVLPDVPAEHRNILISQGCIVREIEPVYPPENQTQFA
SecStr      CCCCCCCCCCCCCCCCCCCCCCCSSSSSSSSCCHHHHHHHHHHHHHHHHHCCCCCSSSSSCCCCCHHHHHHHHHCCCSSSSSSSCCCCHHCCCCC
Score       999853555665668888888679999842887899999999999985899988999769888899878884598899973104620013333

GolS2       MAYYVINYSKLRIWAFEEYEKMIYLDGDIQVFENIDHLFDLPNNYFYAVMDCFCEATWGHTKQYKIGYCQQCPDKVQWPSNFGPKPPLYFNAG
SecStr      CCCHHHHHHHHHHHHHCCCCSSSSSCCCCCCCCCHHHHHCCCCCCSSSSSHHCCCCCCCCCCHHHHHHHHHCCHHHCCCCCCCCCCCCSSCCS
Score       530577879987522205445999569731206779986287511244320002431113302113453317144313222478885323032

GolS2       MFVYEPNMDTYHDLLHKLQLTKPTSFAEQDFLNMYFNDKYKPIPNVYNLVLAMLWRHPENVELHKVKVVHYCAAGSKPWRYTGVEENMQREDI
SecStr      SSSSSCCHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHCCCSSSCCHHHHHCCCCCCCCCHHHHHHCCCSSSSSCCCCCCCCCCCCCCCCCCHHHH
Score       899966789999999999703998874656898770791577887873133100126145663664699832899999858888777440158

GolS2       KMLVKKWWEVYEDESLDYKKALNENRLTSAILEAGGINFVRAPNAA
SecStr      HHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHCCCCCCCCCCCCC
Score       999999999975117776414421333332100014668899989
```

Figure 4.4: Predicted secondary structures of (A) LcGolS1 (B) LcGolS2. β-Sheets are indicated as S in blue, α- helices indicated as H in red and coils represented as C in black. Confidence scores are indicated below the predicted secondary structure with values from 0-9 with 9 indicating high probability of secondary structures. Highlighted region (Yellow) indicating highly variable regions in the 3-D structure observed during superposition of LcGolS1 and LcGolS2 predicted models (Figure 4.5: Indicated as gray and red in the three dimensional Figure)

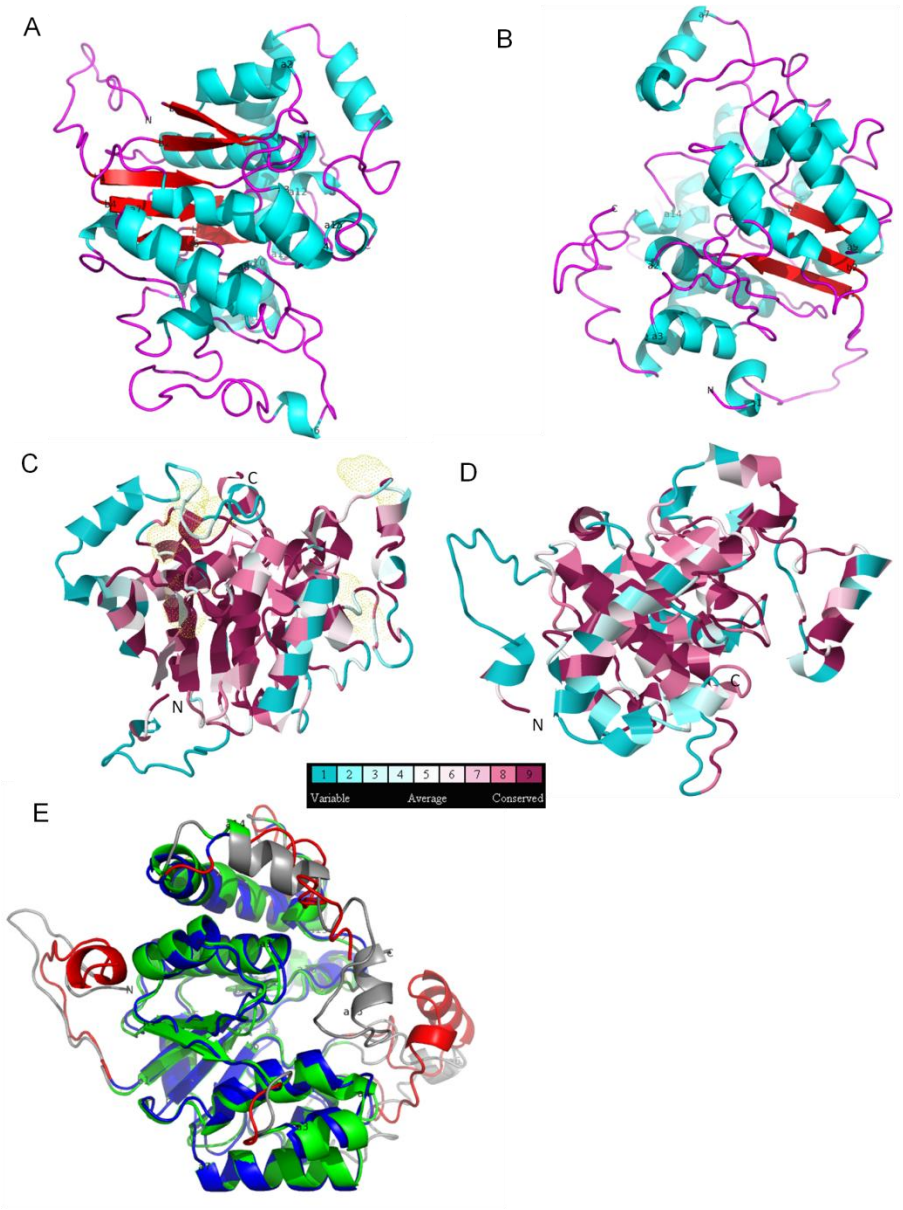


Figure 4.5: Predicted three dimensional structures (A) LcGolS1 and (B) LcGolS2 with α-helices represented in blue, β-sheets in red and coils in magenta. (C, D) Predicted model structures are represented with blue showing highly variable region and magenta indicating highly conserved regions in the structure for LcGolS1 (C) and LcGolS2 (D) obtained through Consurf. E represents superposition of LcGolS1 and LcGolS2 structures with LcGolS1 indicated in green and GolS2 indicated in blue. Variable regions of LcGolS1 are indicated as gray and LcGolS2 indicated as red

transferase domain. The N and C terminal are highly variable and this is evident with alignment and low confidence scores obtained in secondary structure prediction (Figure 4.4). No differences in structure were obtained in the substrate binding sites and manganese binding site in the protein. Superposition of LcGolS1 and LcGolS2 predicted structures align at conserved regions and significant differences were only obtained at the variable region (Figure 4.5). Results obtained were similar to predicted ZmGolS3 structure reported earlier (Sengupta et al 2012).

4.4.8 Transcript accumulation of *LcGolS1* and *LcGolS2*

Validation of stable reference genes for qRT-PCR analysis

Since the MIQE (Minimum information for publication of quantitative real-time PCR experiments) guidelines for real time PCR experiments (Bustin et al 2009) was established, several reports have been published where more emphasis has been placed on reference gene validation (Hu et al 2009; Paolacci et al 2009; Die et al 2010; Garg et al 2010; Manoli et al 2012). A single gene cannot be considered as the most stable reference gene, hence validation of reference genes is an important criterion for gene normalization. Literature search revealed no validation studies for stable reference genes during lentil seed development.

The most stable reference genes in lentil were found to be EF1α (elongation factor 1-alpha), GAPDH (glyceraldehyde-3-phosphate dehydrogenase) and HSP (heat shock protein) during seed development (Appendix 4.10). The most stable genes along the circadian rhythm were adipose ribosylation factor (ARF) and beta-tubulin (BTUB) and alpha-tubulin (ATUB) (Appendix 4.11). The stable reference genes along the circadian rhythm were different than those during seed development emphasizing the importance of reference gene validation in various tissues. Stability of reference genes across different tissues showed GAPD, HSP and EF1-alpha as the stable genes, which was similar to genes obtained during seed development for lentil (Appendix 4.12). The results obtained were similar to stable reference genes reported in *Cicer arietinum*, where, EF1- alpha and HSP90 were reported as the most stable genes in different tissues and developmental stages for qRT-PCR analysis (Garg et al 2010).

Transcript accumulation of LcGolS1

Expression of *LcGolS1* in developing lentil seeds was studied using RNA gel blot analysis. Probe 1(Figure 4.1A) produced a strong hybridization signal to a 1.5 kb long transcript

only (Figure 4.6B). No hybridization signal was obtained before 26 DAF and the 1.5 kb *LcGolS1* transcript increased steadily from 26 DAF with maximum transcript accumulation at 32 DAF. Beyond this stage, the amount of 1.5 kb *LcGolS1* transcript declined sharply (34 and 28 DAF). Nucleotide sequences of *LcGolS1* and *LcGolS2* were highly similar, hence primer specificity of *LcGolS1* and *LcGolS2* was assessed prior to quantitative real time PCR analyses (Figure 4.7A). A trend similar to RNA gel blot analysis was observed for *LcGolS1* gene expression by qRT-PCR analysis, where, transcript abundance increased from 22 DAF to later stages of seed development with maximum transcript accumulation at 32 DAF (Figure 4.7 B). GS gene expression pattern in lentil was similar to *P. sativum*, where *PsGolS* (Peterbauer et al 2001b) showed transcript accumulation towards later stages of seed development and presence of transcripts were also obtained in mature seeds. This observation concurs with results obtained in phylogenetic analysis, where high similarity was observed between *LcGolS1* and *PsGolS* (Figure 4.7). Similar, transcript accumulation pattern was also observed in developing seeds in *L. esculentum* (Downie et al 2003) and presence of transcript was also observed at seed maturity.

Transcript accumulation of LcGolS2 during seed development

Gene expression analysis for *LcGolS2* using qRT-PCR with gene specific primers showed a trend different to *LcGolS1* transcript accumulation (Figure 4.7C). Gene expression of *LcGolS2* increased from 20 DAF with maximum accumulation at 24 DAF. Beyond this stage, a sharp decline in transcript accumulation of *LcGolS2* was observed until later stages of seed development (Figure 4.7 C). Presence of transcript was also detected at seed maturity (32 DAF).

4.4.9 Accumulation of Galactinol and its precursors during seed development

Myo-inositol was detected at earlier stages (embryogenesis and early seed fill) with a gradual decrease from 22 DAF to 30 DAF, thereafter reaching a plateau towards 30 - 32 DAF (late maturation phase) (Figure 4.8 A). Combined glucose / galactose concentrations have been presented since this method was unable to differentiate glucose from galactose (Figure 4.8 B). Concentration of glucose/galactose was detected at early stages of seed development and its concentration decreased from 16 DAF and reached a plateau at 20 DAF. Galactinol

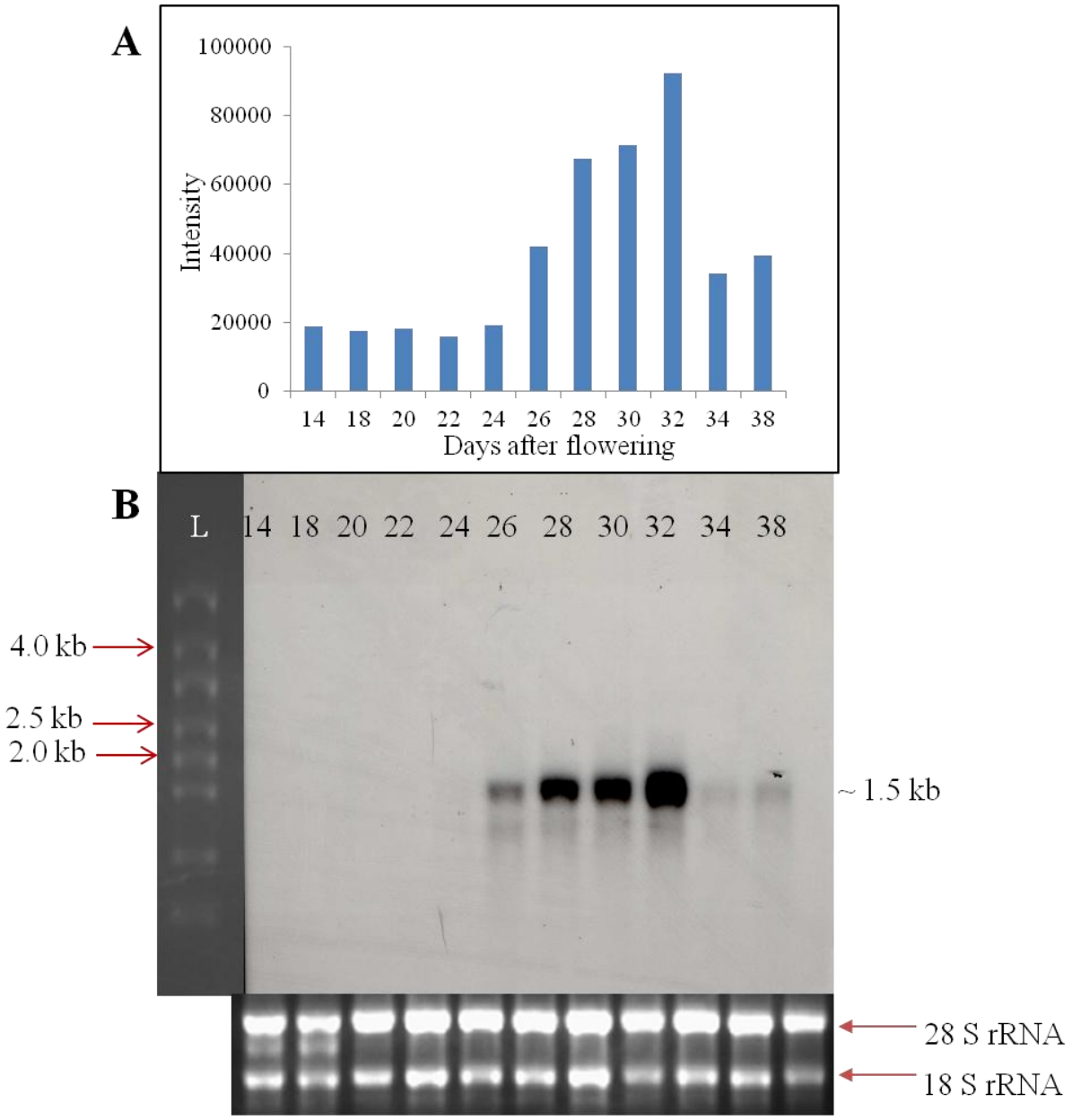


Figure 4.6: Transcript accumulation of *LcGolS1* (A) Graph plotted based on intensity of transcript obtained in the RNA blot with background as the reference (B) *LcGolS1* transcript accumulation and 28S and 18S rRNA band intensity observed in the gel before blotting.

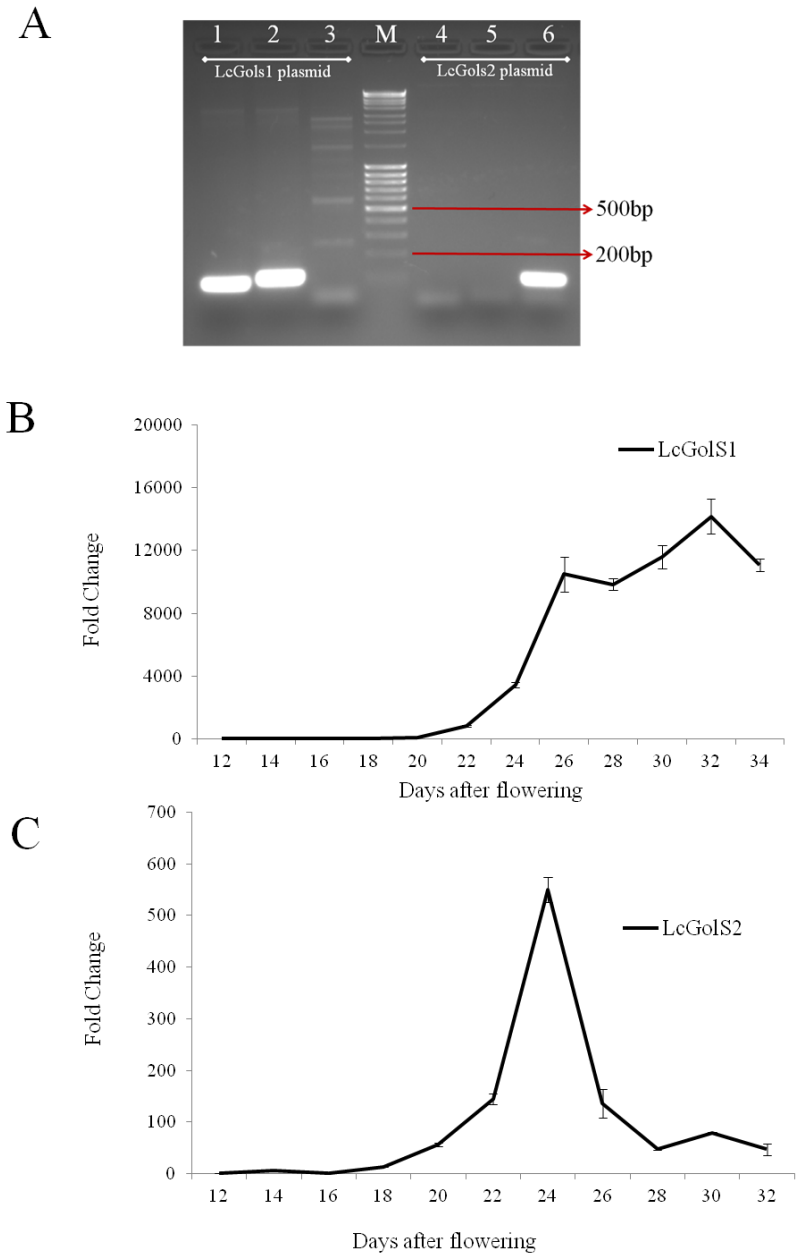


Figure 4.7: Transcript accumulation of *LcGolS1* and *LcGolS2* by qPCR analyses (A) Confirmation of primer specificities for LcGolS1 and LcGolS2 for qPCR analyses. Lane 1, 2 and 3 show amplification using plasmids with *LcGolS1* insert and lanes 4,5 and 6 show amplification using plasmids with *LcGolS2* insert. *LcGolS1* primers were used to amplify plasmids indicated in lanes 1, 2, 4, 5 and *LcGolS2* specific primers were used to amplify plasmids indicated in lanes 3 and 6 (B) *LcGolS1* expression during seed development (C) *LcGolS2* expression obtained during seed development. Error bars indicate standard error of the mean of four determinations at each time point.

concentration was detected at 22 DAF and concentration of galactinol increased with maximum accumulation obtained at 26 DAF (Figure 4.8C) decreasing gradually towards later stages of seed development.

4.4.10 Enzyme activity of GS and correlation with GolS transcript accumulation and galactinol accumulation

A trend similar to galactinol accumulation was observed in GS enzyme activity (Figure 4.8D). No activity was detected before 18 DAF. A rapid increase in GS activity was observed from 20 DAF with maximum activity at 26 DAF 1082.5 ± 127.4 pkat/mg protein to 28 DAF- 1019.2 ± 46.3 pkat/mg protein. GS activity decreased towards later stages of seed development however some activity was still detected at 32 and 34 DAF. *GolS* gene expression and GS enzyme activity was also observed 32 and 34 DAF where physiological maturity of seed had already been reached. Transcript accumulation of *LcGolS1* showed an abundant accumulation towards 32 DAF. Similar observation of transcript and activity abundance was also reported in *Pisum sativum* where transcript accumulation was several-fold higher than required amount for protein synthesis. (Peterbauer et al 2001b). This suggests a post transcriptional regulatory mechanism at 32 or 34 DAF, where reduced GS enzyme activity was observed. In *Medicago truncatula* genes induced during late maturation and post abscission phases of seed development, a 5- fold increase in genes related to transcriptional and translational regulation was observed at later stages of seed development (Verdier et al 2013). Further, delay in transcript accumulation and protein synthesis was reported to be due to post transcriptional regulation of genes. Protein accumulation of GS across different stages of seed development will facilitate understanding of any post translational regulatory mechanisms at these stages. Recombinant *GolS* isolated from soybean catalyzed the synthesis of fagopyritols but not galactopinitols (Obendorf et al 2004). Expression of recombinant enzymes of *LcGolS1* and *LcGolS2* would provide insights of substrate specificity of these enzymes.

Transcript accumulation patterns of *LcGolS2* correlates with GS enzyme activity and hence could be suggested that *LcGolS2* might contribute to GS activity in seeds. However, based on sequence similarity and phylogenetic analysis, where close similarity was observed with *P. sativum GolS* (Peterbauer et al 2001b), suggest that *LcGolS1* might contribute more towards GS activity in seeds. Zhao et al (2004), isolated two GolS genes *ZmGolS2* and *ZmGolS3*

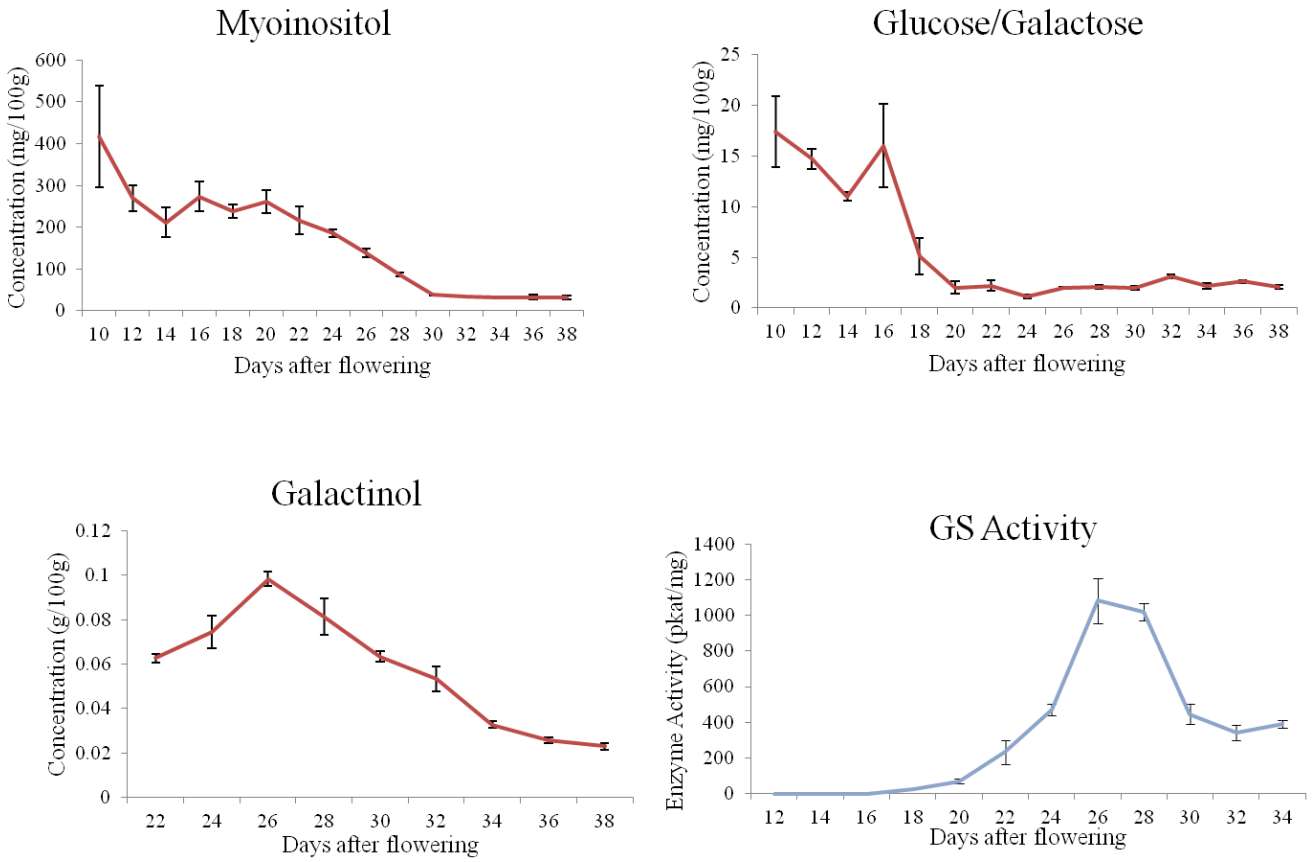


Figure 4.8: Accumulation of *myo*-inositol (A), glucose/galactose (B), galactinol (C) and galactinol synthase activity (D) during seed development. Error bars indicate standard error of the mean of six determinations at each time point.

from developing seeds and during seed germination. Transcript accumulation of *ZmGolS3* was observed to be present throughout seed development and also in mature seeds. *ZmGolS3* expression was observed to be dominant during seed development. *ZmGolS2* was primarily associated with post-imbibition dehydration stress (Zhao et al 2004). Gene expression of *LcGolS2* was not studied during germination or stressed conditions.

In conclusion, two genes encoding GS, *LcGolS1* and *LcGolS2* were isolated and gene expression was studied during seed development in lentil. In developing lentil seeds, a switch in expression from *LcGolS1* to *LcGolS2* was observed at 24 DAF suggesting *LcGolS1* induction of expression prior to seed desiccation. Accumulation of galactinol correlated with GS enzyme activity during seed development where maximum accumulation and maximum GS activity were obtained at 26 DAF. The contribution of individual *LcGolS1* and *LcGolS2* towards galactinol accumulation is yet to be elucidated.

5. CHARACTERIZATION OF RAFFINOSE SYNTHASE IN *Lens culinaris* Medik. DURING SEED DEVELOPMENT

Study 3

In this study, a gene coding for raffinose synthase (*LcRS*) was isolated from a lentil cDNA library constructed using RNA from developing lentil seeds (18-24 DAF). Further, *LcRS* gene expression, raffinose synthase enzyme activity and raffinose accumulation were also studied during seed development.

Kannan U, Gangola M, Ganeshan S, Chibbar RN (2014) Characterization of raffinose synthase in *Lens culinaris* Medik. during seed development (to be submitted).

5.1 Abstract:

Raffinose synthase (RS; EC 2.4.1.82) catalyzes the transfer of a galactosyl residue from galactinol to sucrose to synthesize raffinose. To understand the role of RS in RFO accumulation in lentil seeds, raffinose accumulation, *Lens culinaris RS* transcript accumulation and RS enzyme activity were analysed during seed development from 12-38 days after flowering (DAF). Concentration of raffinose showed a rapid increase from 20 days after flowering with maximum concentration at 28 DAF. Maximum accumulation of sucrose was observed at 16 DAF showing rapid decrease towards maturity. Screening a cDNA library, constructed from developing seeds (18 - 24 DAF), using degenerate oligonucleotide primers, identified a 2.7 kb cDNA clone corresponding to RS. The RS gene coded for 792 amino acids with a predicted molecular weight of 88.2 kDa and pI 6.09. Nucleotide sequence alignment and phylogenetic analysis of *LcRS* to other RS genes showed high similarity to *Pisum sativum RS*. Transcript accumulation obtained by RNA gel blot and qPCR analyses results show maximum *LcRS* transcript accumulation during 24-26 DAF. Maximum RS enzyme activity was also observed at 24 DAF. RS gene expression correlated with enzyme activity suggesting that the isolated gene contributes raffinose accumulation in developing lentil seeds.

5.2 INTRODUCTION

The 'classic' raffinose family oligosaccharides (RFO) are made up of mono-, di-, tri-, and tetra- galactosides of sucrose that include raffinose, stachyose, verbascose and ajugose, respectively. Raffinose, the smallest RFO member, is ubiquitous in plants, but stachyose the di-galactoside is the most abundant RFO in pulse seeds including lentil (Tahir et al 2011). Raffinose synthase (RS; E.C. 2.4.1.82) catalyzes the transfer of a galactose moiety from galactinol to sucrose synthesizing raffinose and *myo*-inositol (Lehle et al 1970; Peterbauer et al 2002a). Among the RFO biosynthetic enzymes, RS is the least studied enzyme, which is attributed to its instability *in vitro* (Castillo et al 1990; Peterbauer et al 1998, 2002). RS was first purified from seeds of *Vicia faba* (Lehle and Tanner 1973). *Vicia faba* RS polypeptide is 90 kDa and exhibits a pH optimum of 7.0, Km 7.3 mM and 22.9 mM for galactinol and sucrose respectively (Lehle and Tanner 1973). A cDNA for pea RS was expressed in *Spodoptera frugiperda* Sf21 insect cells to produce recombinant protein. Kinetic properties of the recombinant protein were similar to the purified RS from pea (Peterbauer et al 2002a). *Oryza*

sativa RS was expressed in *E. coli* and recombinant protein showed a pH optima 7.0 at 45 ^{o}C (Li et al 2007).

Biochemical characterization for RS has been done only in one pulse crop species, *Pisum sativum*, and in three other plants, *Cucumis sativus*, *Oryza sativa* and *Arabidopsis thaliana* (Peterbauer et al 2002a, Sui et al 2012, Li et al 2007, Egbert et al 2013). However, gene sequences coding for RS have been isolated from several plant species including *Pisum sativum* (AJ426475), *Cucumis sativus* (AF073744), *Oryza sativa* (BAD68247), *Glycine max* (EU651888.1, EU651889.1), *Arabidopsis thaliana* (At5g40390), *Vitis vinifera* (AM430487) and *Zea mays* (Q575Z6, Q575Z7, Q575Z8) (Dierking and Bilyeu 2008; Egbert et al 2013). Presence of more than six RS genes has been reported in *Arabidopsis thaliana* (*AtRS1-6*). *AtRS2* previously annotated as putative RS gene or seed imbibition protein was reported to be an alkaline α-galactosidase (Peters et al 2010). *AtRS5* was shown to be responsible for raffinose biosynthesis in leaves. Of the six RS genes, *AtRS5* showed high sequence similarity to RS previously characterized in *Pisum sativum* (Egbert et al 2013). In *Glycine max* presence of four RS genes was reported. Among all RS genes isolated in *Glycine max*, high sequence similarity was observed between *GmRS2* and *PsRS* (72 % identity). A variation in *GmRS2* was associated with low raffinose, low stachyose phenotype in line PI 500208 through candidate gene selection in soybean. The reduction in RFO was conferred due to a 3 bp deletion in RS2 gene (Skoneczka et al 2009; Dierking and Bilyeu 2008).

The main objective of this study was to characterize RS genes and enzyme activity during seed development in *Lens culinaris* cv CDC Redberry. Transcript accumulation, RS activity and accumulation of raffinose during seed development would provide us more insights about the role of this enzyme in the RFO biosynthetic pathway.

5.3 MATERIALS AND METHODS

5.3.1 Plant material

Lens culinaris Medik. cv CDC Redberry plants were grown in growth chambers with 18 h photoperiod (250 mmol m^{-2} s^{-1}, PPFD [photosynthetic photon flux density]), 21 ^{o}C / 15 ^{o}C day / night temperatures. Plants were fertilized every two weeks with NPK (Nitrogen: Phosphorous: Potassium) 20:20:20. Young leaves were collected from plants and stored at -80 ^{o}C for genomic DNA extraction. Flowers were tagged when they were fully opened and developing seeds were

collected at two day intervals from 10 days after flowering (DAF) to 38 DAF and stored at -80 ^{o}C. Sample collection was done at the same time (3:00 - 5:00 pm) during the day to avoid effect of circadian rhythm in photoassimilate accumulation.

5.3.2 DNA and RNA isolation

Genomic DNA was extracted from young leaf tissues (200 mg) of lentil using CTAB method (Doyle and Dolye 1990). RNA was isolated from developing lentil seeds using CTAB-LiCl based isolation procedure (Kannan et al 2014).

5.3.3 Isolation of partial *LcRS* sequence

Degenerate primers were designed for raffinose synthase (RS) gene by aligning *RS* sequence of closely related species. *RS* gene sequences of *Pisum sativum* (AJ426475), *Glycine max* (EU651889) and *Vicia faba* (E244231) were obtained from NCBI. The sequences were aligned using Clustal Omega (Sievers et al 2011). Oligonucleotide primers were designed in the conserved exonic region of the sequence excluding the exon/intron splice sites. Partial sequences were PCR amplified from lentil DNA using degenerate primers forward DG/RS_F and reverse DG / RS_R (Appendix 4.2). The PCR reaction comprised of 20 mM Tris-HCl (pH 8.4), 50 mM KCl, $MgCl_2$ concentrations (1.5 mM, 2.5 mM, 3.5 mM), primer concentrations (0.2 µM and 0.4 µM), 0.2 mM each dNTP and 3 units Taq Polymerase (Life Technologies Inc., Burlington, Ontario, Canada). The amplification conditions were: initial denaturation at 95 ^{o}C 4 min 15 sec, followed by 32 cycles of 95 ^{o}C for 45 sec, 20 sec annealing temperature (tested at 54 ^{o}C, 57 ^{o}C and 60 ^{o}C), 1 min 30 sec extension at 72 ^{o}C and final cycle at 72 ^{o}C at 10 min (Appendix 5.1). The amplified fragments were analyzed by agarose (1 % w / v) gel electrophoresis (Figure 5.1A). The amplified fragment (352 bp) was DNA sequenced to confirm its identity.

5.3.4 Screening of cDNA library

A cDNA library (titre $1x10^9$ pfu / mL) constructed using mRNA from developing lentil seeds (18, 20, 22 and 24 days after flowering, DAF) was screened to obtain *LcRS* sequence. The 1957 bp partial RS gene was digested with restriction enzymes *XhoI* and *NheI* to obtain 745 bp fragment which was used as a probe to screen cDNA library (Probe 1; Figure 5.1C). The digested fragment (745 bp) was radiolabelled with alpha ^{32}P- dCTP prepared using random

primer labelling kit (Life Technologies Inc., Burlington, Ontario, Canada). Plaque lifting and hybridization was performed using standard protocols (Sambrook and Russell 2001). Positive plaques were selected based on appearance of spots in replicate autoradiographs. To confirm the presence of insert in clones obtained from primary screening, gene specific primers forward F/R 17 and reverse R / R 1356 (Appendix 6.3) were used to amplify *RS* fragment. Primary clones which showed the presence of 1.3 kb fragment were selected for secondary screening. Positive plaques were cored from agar plate and phagemids were excised using ExAssist® SOLR system (Agilent Technologies Inc., Santa Clara, CA).

Rapid amplification of cDNA ends (RACE) strategy was used to isolate 5'UTR of *RS* by GeneRacer® Kit with SuperScript® III RT (Life Technologies Inc., Burlington, Ontario, Canada) using manufacturer's instructions. Oligonucleotide primers forward RS_RACE / F1 and RS_RACE / F2 (Appendix 6.3) were used for amplification.

5.3.5 RNA gel blot analysis

Total RNA (40 µg) was loaded on a 1% agarose gel containing 0.066 M formaldehyde and electrophoresed in HAE buffer (0.2 M Na-HEPES, 50 mM NaAc, 10 mM EDTA, pH 7.0) for 3 - 4 h at 100V. Ethidium bromide was added along with sample and ribosomal RNA was visualized under UV light to estimate concentration of RNA and its integrity. The gel was rinsed with (i) DEPC treated water (ii) 50 mM NaOH, 150 mM NaCl (iii) 100 mM Tris-HCl pH 7.5, 150 mM NaCl for 30 min each to remove traces of formaldehyde from gel. Total RNA was transferred to Hybond-N^+ membrane (GE Healthcare, Baie d'Urfe, Quebec, Canada) using a positive pressure blot, PosiBlot^{TM} apparatus (Agilent Technologies Inc., Santa Clara, CA, USA) for 2 h at 75 mm Hg pressure. Transferred RNA was cross-linked to the membrane in a UV cross-linker (UVP Inc., Upland, CA, USA). The membrane was hybridized in Church hybridization buffer (0.5 M Na_2HPO4, 7 % (w / v) SDS and 1 mM EDTA) at 65 ^oC with a 1068 bp probe (Probe 2; Figure 5.1C) towards the 3' end of *RS* gene labelled with [α^{32}-P] dCTP. Membrane were sequentially washed with (i) 1X SSPE (150 mM NaCl; 1 mM EDTA; 0.1 % (w/v) SDS (ii) 0.1× SSPE, 0.1 % (w / v) SDS twice ten min each until low background activity was observed. Membrane was exposed to Biomax MR film (KODAK® BioMax® Maximum Resolution (MR) Autoradiography Film) at -80 °C for 24 h to several days depending on signal intensity.

5.3.6 Quantitative real time PCR analysis

For Quantitative PCR analyses, DNase- treated total RNA (1 µg) was used for cDNA synthesis using 200 U Superscript III reverse transcriptase (Life Technologies Inc., Burlington, Ontario, Canada), 500 ng of Oligo dT, 900 ng of random primer. Mx3000P real-time PCR machine (Agilent Technologies Inc., Santa Clara, CA, USA) was used for PCR amplification with reaction mixture consisting 5 µL of 1 / 15 dilution of cDNA, 300 nM primers and 2x Maxima^{TM} SYBR Green I qPCR Master Mix (Thermo Fisher Scientific Inc., Pittsburgh, PA, USA). Two controls, one without addition of template and one with cDNA synthesized without addition of reverse transcriptase was used to check presence of genomic DNA contamination. Thermocycler conditions were: initial denaturation at 95 ^{o}C for 10 min, followed by 40 cycles of 95 ^{o}C for 15 sec, 60 ^{o}C for 1 min and 72 ^{o}C for 1 min. The dissociation step was performed to confirm single product amplification. Primers used for *RS* gene RS_RT_F (forward) and RS_RT_R (reverse)) reference genes GAPD (GAPD_RT_F, GAPD_RT_R), HSP (HSP_RT_F, HSP_RT_R) and EF1-α (EF_RT_F, EF_RT_R) were used to validate RS expression (Appendix 6.3). Geometric mean of reference gene Ct values was used to calculate relative expression (Vandesomple et al 2002). Relative expression of *RS* was calculated by $2^{-\Delta\Delta Ct}$ method (Livak et al 2001) using 12 DAF sample as the calibrator. The experiment was repeated twice with two technical replicates.

5.3.7 Phylogenetic Analysis

Sequences of *RS* genes in GenBank, NCBI were aligned and analysed by Clustal Omega (Sievers et al 2011) to construct a phylogenetic tree using MEGA 6 (Tamura et al 2013). UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method. Poisson correction was used to calculate evolutionary distances.

5.3.8 Determination of sucrose and raffinose concentration in seeds

Developing seed samples were freeze dried and ground using a pestle and mortar. Raffinose concentration was determined in dried samples (100 mg) using a previously described method for RFO extraction and its constituent determination (Gangola et al 2014). In a 15 mL conical tube, 2 mL ethanol was added to 100 mg sample, and the slurry was vigorously mixed on

a vortex mixer. The slurry was incubated at 60 ^{o}C for 45 min with intermittent mixing every 15 min. The slurry was centrifuged at 12100 x g for 10 min. The supernatant was collected and pellet was re-extracted as described above. The supernatants were pooled and purified by passing through pre-equilibrated C18 Solid Phase Columns (Honeywell Burdick and Jackson, Muskegon, MI, USA) attached to a vacuum manifold. The columns were first washed with 5 mL of methanol (99% v/v), followed by 5 mL of nanopure water. Purified extract (1.6 mL) was dried using Speedvac concentrator and universal vacuum system (Thermo Savant, Holbrook, NY, USA). The dried samples were dissolved in 500 µL water and were centrifuged at 12000 x g for 10 min. The samples were diluted (125 µL supernatant and 375 µL water) and analyzed using Dionex^{TM} ICS 5000 as described (Gangola et al 2014). RFO constituents were separated using a CarboPac PA 100 and CarboPac PA 100 guard column (Dionex Canada Ltd., Oakville, ON, Canada) with flow rate of 1 mL / min at 30 ^{o}C. Pure raffinose (Sigma-Aldrich Inc., St. Louis, Missouri, USA) was used to identify the peaks in chromatogram by comparing the retention times and to develop a standard curve. Peak area was used to calculate raffinose concentration in lentil seed meal.

5.3.9 Raffinose synthase enzyme activity assay

Raffinose synthase (RS) enzymatic activity was measured in crude lentil seed cell extracts, by determining the raffinose concentration using a HPAEC-PAD method essentially based on previously published methods (Peterbauer et al 1998; 2001, Hitz et al 2002). The enzyme activity method optimized for pH, protein concentration, temperature and substrate concentration in *Cicer arietinum* was followed (Gangola 2014). Lentil seeds (200 mg) were ground in a pestle and mortar with liquid nitrogen and 2 mL of extraction buffer containing 50 mM HEPES (2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid)-NaOH, pH 7.0 and 5 mM of 2-β mercaptoethanol. The extract was centrifuged at 12,000 x g for 10 min at 4 ^{o}C and supernatant was filtered through a Sephadex G25 column pre-equilibrated with 5 mL of extraction buffer. The protein concentration of the filtrate was determined using a dye binding assay (Bradford 1976) using a commercial kit (Bio-Rad protein assay, Bio-Rad Laboratories Inc., Mississauga, Ontario, Canada). Enzyme assay was performed using total cellular proteins (30µg); galactinol (10 mM), sucrose (40 mM) and dithiothreitol (10 mM) in a final reaction volume of 50 µL. The reaction pH was 7.0 and assay mixture was incubated for 60 min at 25 ^{o}C.

Two controls i) protein and no substrate ii) substrate and no protein was included for each assay. The reaction was terminated by adding 50 µL ethanol (95 % v / v) and incubating the tubes in a boiling water bath for 1 min. The tubes were centrifuged at 12,000 x g for 10 min and supernatant was transferred to a new tube containing 10 mg of Dowex 50WX8 100 ion exchange resin (Sigma-Aldrich Inc., St. Louis, Missouri, USA) and shaken at 200 rpm for 30 min. The tubes were centrifuged for 5 min at 10,000 x g for 10 min. The supernatant was transferred to a new tube and dried using Speedvac concentrator and universal vacuum system (Thermo Savant, Holbrook, NY, USA). Pellet was dissolved in 500 µL water and mixed well. The tubes were centrifuged at 12,000 x g for 10 min. The samples were diluted (125 µL supernatant and 375 µL water) and analyzed using Dionex^{TM} ICS 5000 as described (Gangola et al 2014).

5.4 RESULTS AND DISCUSSION

5.4.1 Amplification of partial raffinose synthase gene for probe synthesis

Genomic DNA extracted from *Lens culinaris* cv CDC Redberry in a PCR reaction with degenerate RS oligonucleotide primers amplified a 352 bp DNA fragment, which was cloned and DNA sequenced (Figure 5.1A). Nucleotide sequence of amplified *L. culinaris RS* showed 87 % and 95 % identity to *P. sativum* RS nucleotide and amino acid sequence, respectively.

Recently released lentil annotated unigenes and contigs (Kaur et al 2011) did not contain any *RS* sequences. However, five singletons showing sequence identity to *P. sativum RS* were identified and aligned with it as a reference gene (Figure 5.1B). Four singletons ls_639304 (325bp), ls_62278 (367bp), ls_560993 (413bp) and ls_47762 (188bp) aligned at 510 - 835 bp, 1353 - 1706 bp, 1825 - 2238 bp and 2294 - 2481 bp respectively to the 2671 bp *P. sativum RS* (AJ426475.2) nucleotide sequence. Primers designed at the 3' end of ls_63904 and at the 5' end of ls_417762 were used to amplify a partial gene sequence (1957 bp) from *Lens culinaris* cDNA (Primer positions shown in Figure 5.1 B). Partial *RS* obtained using degenerate primers (352 bp) showed complete sequence similarity to the 1957 bp fragment (Amplified regions indicated in Figure 5.2).

5.4.2 Isolation and characterization of raffinose synthase cDNA clones

Primary screening of the cDNA library (1.8 x 10^5 phage particles accounting for 18 % of the cDNA library) for *RS* yielded 50 positive clones. Amplification using gene specific primers

with the 50 primary clones as the template confirmed the presence of insert in several (32) clones (Appendix 5.2). Secondary screening of 12 selected clones resulted in 16 positive clones selected for *in-vivo* excision to recover plasmids. A total of 48 samples, 16 plasmids in triplicates were sent for DNA sequencing. Sequence similarity search of the 48 samples showed 32 samples to be highly similar to existing RS genes in the GenBank database. Sixteen clones of the total 48 samples showed no identity to *PsRS* sequence. DNA sequencing revealed different fragment sizes of isolated *RS* genes: 1336 bp (3 clones), 1671 bp (5 clones), 1872 bp (3 clones), 2422 bp (9 clones) and 2526 bp (12 clones). All the sequences were identical and had complete 3' UTR sequences, but differences in sequence length were due to regions in the 5' region of the sequence in all the cDNA clones.

Aligning the longest cDNA clone (2527 bp) with *P. sativum* revealed that 249 bp at the 5' end were absent from the complete sequence of *RS*. Full length sequence of RS could have been obtained through further screening of the cDNA library, however, due to the small length of sequence to be isolated (38 amino acids towards to the 5' end of the sequence), 5' UTR of the cDNA was isolated by RACE (rapid amplification of cDNA ends) strategy. This strategy provided us with the full length sequence of 2776 bp (Figure 5.2).

The 2776 bp cDNA corresponded to full length transcript with an open reading frame of 2376 bp (nucleotide 136 to 2511 bp in Figure 5.2) flanked by 5' UTR of 135 bp in length and 3' UTR of 262 nucleotide in length (Figure 5.2). A polyadenylation signal 'AATAA' (2314 - 2318 bp in Figure 5.2) was present in the 3' region of the sequence upstream of the poly (A) tail. Initiation of translation was at first 'ATG' encoding for the first methionine in the sequence. The open reading frame coded for a polypeptide of 792 amino acid in length with predicted molecular weight of 88.2 kDa and pI 6.09. *Pisum sativum* RS sequence was similar to *LcRS* sequence with 5'UTR of 147 bp in length, open reading frame of 2394 bp and 3' UTR of 127 bp length. *P. sativum RS* encoded a polypeptide of 798 amino acid residues with molecular weight of 88.7 kDa (Peterbauer et al 2002a). The *LcRS* was 89% identical to *PsRS* at nucleotide level and 93% similar at amino acid level. Prediction of active sites with Pfam search tool (http://pfam.sanger.ac.uk/; Punta et al 2014) showed the presence of active sites for RS or seed imbibition protein that belong to the Clan 0058 proteins which represents Tim barrel glycosyl hydrolase (GH) superfamily. Among the GH family proteins, GH36 protein family is further

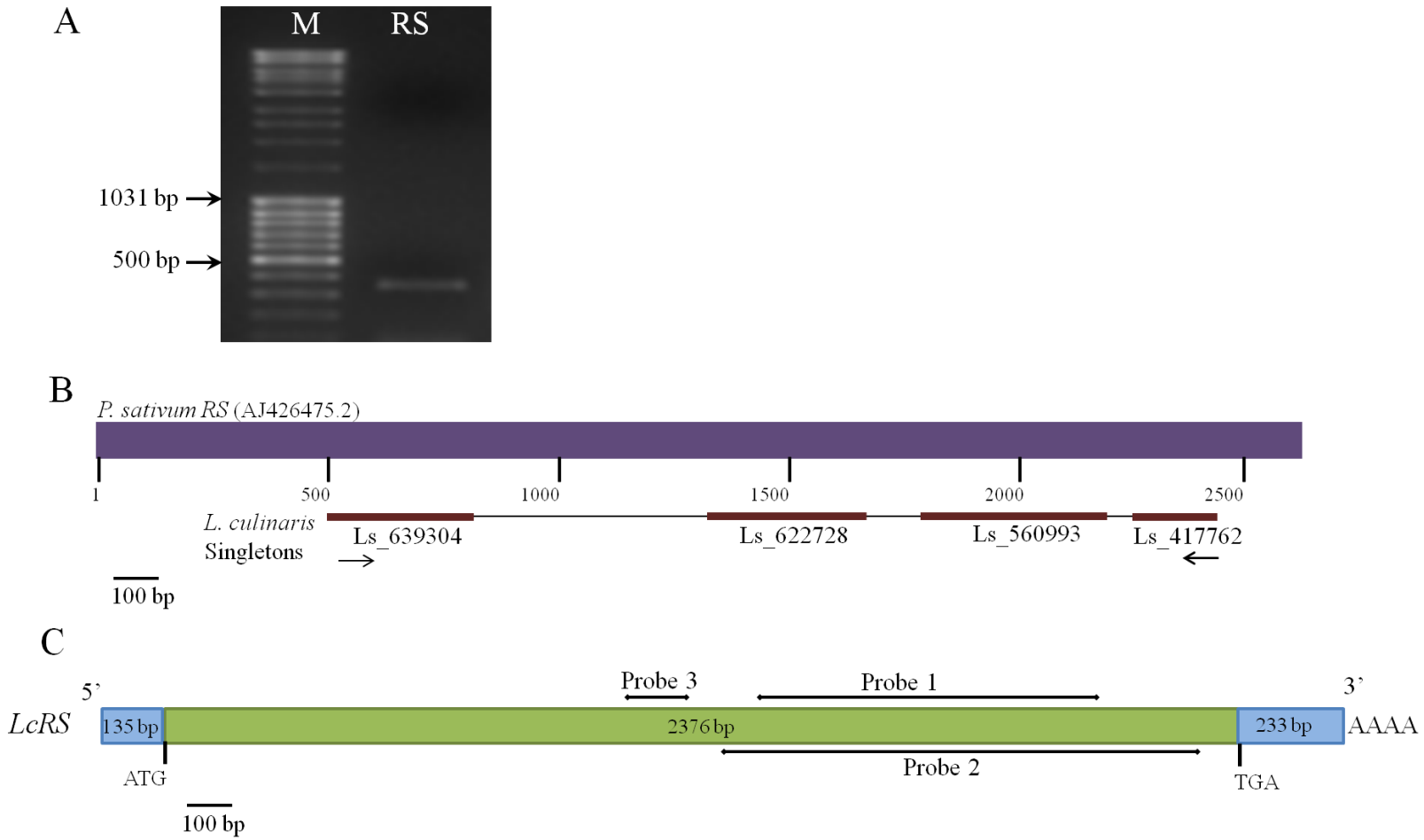


Figure 5.1: Isolation of partial raffinose synthase sequence (A) Partial *LcRS* obtained using degenerate primers. M indicates DNA ladder (B) Alignment of *L. culinaris* singletons with *P. sativum RS*. Arrows indicate primer positions used to amplify of partial *LcRS* of length 1975 bp (C) Schematic representation of full length *LcRS* sequence. Probe 1 indicates position of probe (745 bp in length) used for screening cDNA library; Probe 2 indicates position of probe (1048 bp) used for RNA gel blot analysis; Probe 3 indicates position of gene sequence used for qRT-PCR analysis (Primers were designed towards the 5' and 3' of Probe 3 to amplify 130 bp DNA fragment).

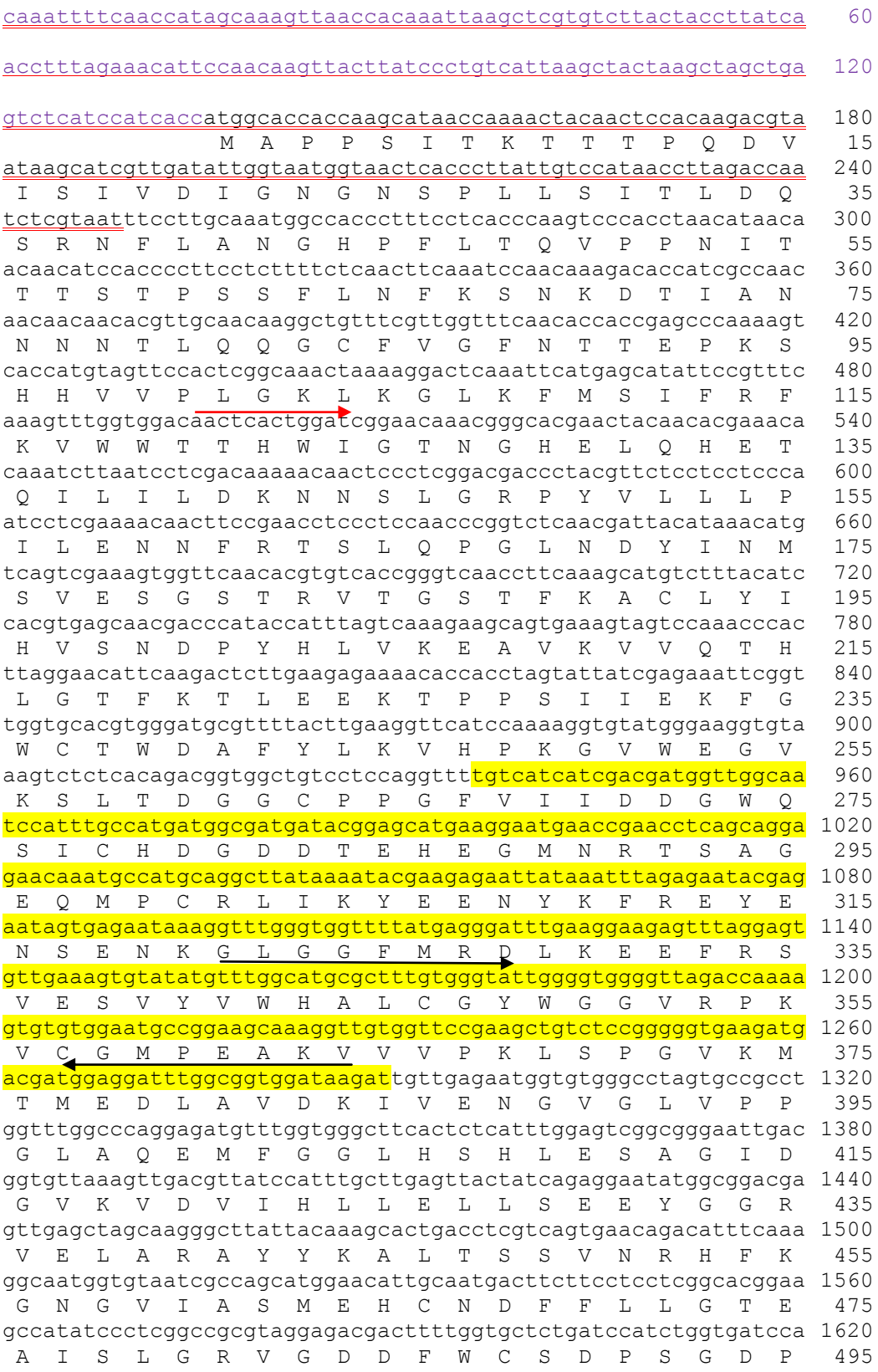


Figure 5.2: *LcRS* nucleotide sequence

```
aatggtacatattggctccaaggttgtcacatggtacattgtgcctacaacagtttatgg 1680
 N  G  T  Y  W  L  Q  G  C  H  M  V  H  C  A  Y  N  S  L  W   515
atgggaaatttcattcatccagattgggacatgtttcagtccactcacccttgtgctgaa 1740
 M  G  N  F  I  H  P  D  W  D  M  F  Q  S  T  H  P  C  A  E   535
tttcatgccgcctcaagagccatctctggtggaccaatttatgttagtgattgtgttggt 1800
 F  H  A  A  S  R  A  I  S  G  G  P  I  Y  V  S  D  C  V  G   555
aatcacaatttcaagttgctcaagtctcttgttttgccagatggttctatcttgcgttgt 1860
 N  H  N  F  K  L  L  K  S  L  V  L  P  D  G  S  I  L  R  C   575
caacattacgcactcccaacacgagattgcttgtttgaagacccttgcataatggcaaa 1920
 Q  H  Y  A  L  P  T  R  D  C  L  F  E  D  P  L  H  N  G  K   595
acaatgctcaaaatttggaatctcaacaaatatgcaggtgttttgggtttattcaattgc 1980
 T  M  L  K  I  W  N  L  N  K  Y  A  G  V  L  G  L  F  N  C   615
caaggtggcgggtggtgtcccgagacacggcgaaacaagagtgcctctgagttttcacac 2040
 Q  G  G  G  W  C  P  E  T  R  R  N  K  S  A  S  E  F  S  H   635
accgtgacatgttatgcaagtcctgaagatattgaatggtgcaatgggaaaatcccaatg 2100
 T  V  T  C  Y  A  S  P  E  D  I  E  W  C  N  G  K  I  P  M   655
ggcatcaaaggtgtggatgtatttgctgtgtatttcttcaaggagaagaaactgaggctc 2160
 G  I  K  G  V  D  V  F  A  V  Y  F  F  K  E  K  K  L  R  L   675
atgaagtgttctgatagattggaagtttcgctcgagccatttagttttgagctaatgaca 2220
 M  K  C  S  D  R  L  E  V  S  L  E  P  F  S  F  E  L  M  T   695
gtgtctccggtgaaagtcttttcgaataggttgatacagtttgctccgattgggttagtg 2280
 V  S  P  V  K  V  F  S  N  R  L  I  Q  F  A  P  I  G  L  V   715
aacatgctgaactccggtggtgcggttcagtctctggagtttgatgataatgcaaccttg 2340
 N  M  L  N  S  G  G  A  V  Q  S  L  E  F  D  D  N  A  T  L   735
gtcaagattggggttagaggttgtggggagatgagtgtctttgcatctgaaaaaccagtt 2400
 V  K  I  G  V  R  G  C  G  E  M  S  V  F  A  S  E  K  P  V   755
tgctgcaaaattgatggggttagtgtggaatttgtttatgaggacaagatggtgagagtt 2460
 C  C  K  I  D  G  V  S  V  E  F  V  Y  E  D  K  M  V  R  V   775
caaattctgtggcctagttcttcaacattgtctttggtggagtttttatttgatccctc 2520
 Q  I  L  W  P  S  S  S  T  L  S  L  V  E  F  L  F  •        792
atgagattaacttgggatcctatgtatgtgtctatgttttaaagtactttatataagtgt 2580

aatgtgtatatattttcacttttaacactgtgtttgttctaataggattatgcagttag 2640

gtggttaattagttatttgtttgtgaagtaactaacttacttttgttgtaagcttatata 2700

tacttgtatgcataataaatagctaatggatttgatcaattcctaacaaaaaaaaaaaaa 2760

aaaaaactcgaggggg
```

Figure 5.2 (Continued): *LcRS* sequence

Nucleotide and deduced amino acid sequences encoding putative *LcRS* in lentil. The deduced protein sequence of *LcRS* is indicated below nucleotide sequence ('•' indicates translation termination site). Nucleotides coloured in blue indicate 5' and 3' UTR regions flanking the ORF of *LcRS*. Highlighted nucleotide region indicate the 252 bp partial RS fragment obtained by amplification using degenerate primers. Forwards and reverse arrows in the highlighted region indicate primers used for qRT-PCR analysis in this study. Partial gene sequence (1957 bp) obtained by amplification of primers designed in singletons obtained from lentil transcriptome assembly is represented within red arrows in the sequence. Nucleotides with underlined with red in the 5' region of the sequence indicate nucleotides obtained using RACE (Rapid amplification of cDNA ends) approach. Polyadenlyation signal is indicated in the 3'UTR region upstream of the poly (A) site.

classified into GH36 A-K, where four families (GH36A, GH36B, GH36F, and GH36K) have been annotated to α-galactosidases and only one family GH36C has been annotated as RS protein (Naumoff 2011). No resolved structure for RS protein is available in protein data banks. Alignment of RS amino acid sequences from closely related species (Figure 5.3) revealed two important motifs KXD and RXXXD that were conserved in *LcRS*. These results suggest that the isolated sequences code for RS in lentil. No signal peptide (analyzed using Signal P) was present in the *LcRS* sequence confirming that RS is cytosolic.

Only one RS cDNA clone was isolated, however in other plant species more than one isoform of RS has been reported (Dierking and Bilyeu 2008; Egbert et al 2013). Phylogenetic analysis of RS sequences (Figure 5.4) placed the sequences in three major clades indicated as a, b and c. *LcRS, PsRS, GmRS2* and *GmRS3* appear in the same clade (a) suggesting a common ancestor and high identity between sequences. Similarly, *AtRS5, MtRS, CaRS, CsRS* were also placed in the same clade (b) showing high similarity between the sequences and *OsRS* was placed in clade c. *RS* sequences showing high identity to *PsRS* sequence were shown to synthesize raffinose in *Glycine max* and *Arabidopsis thaliana* (Dierking and Bilyeu 2008; Egbert et al 2013).

5.4.3 *LcRS* transcript accumulation

RNA gel blot analyses using RNA from developing lentil seeds and Probe 2 (Figure 5.1C) produced a strong hybridization signal to a 2.7 kb long transcript (Figure 5.5A, B). No hybridization signals were observed prior to 14 DAF during seed development (data not shown). Very low hybridization was observed at 14 and 18 DAF, suggesting that *LcRS* gene expression is initiated around this time. Transcript accumulation of 2.7 kb transcript increased at 22 DAF with maximum transcript accumulation at 24 DAF. Beyond this stage, the 2.7 kb *LcRS* transcript declined sharply towards later stages of seed development (Figure 5.5 A, B). The 2.7 kb transcript correlated with full length *LcRS* cDNA sequence of 2.7 kb. Similar sequence lengths were also observed for *PsRS* (2671 bp) and *Arabidopsis RS5* (2829 bp) (Peterbauer et al 2002a, Egbert et al 2013). Gene specific primers were designed for qRT-PCR analysis (Probe 3 in Figure 5.1C indicates position of amplification) was amplified and DNA sequenced to confirm amplification of *LcRS*. A trend similar to RNA gel blot analysis was observed for *LcRS* gene

Figure 5.3: Amino acid alignment of raffinose synthase genes in closely related species. Raffinose synthase sequences of *P. sativum* (CAD20127.2), *A. thaliana* RS5 (AAL32859.1), *G. max* RS2 (ACD13461.1) and *M. trancatula* (AES71465.1) were used in the alignment. All amino acid residues highlighted in black indicate identical residue whereas similar residues are highlighted residues in grey. Amino acid residue indicated with arrows show conserved motifs KxD, RXXXD.

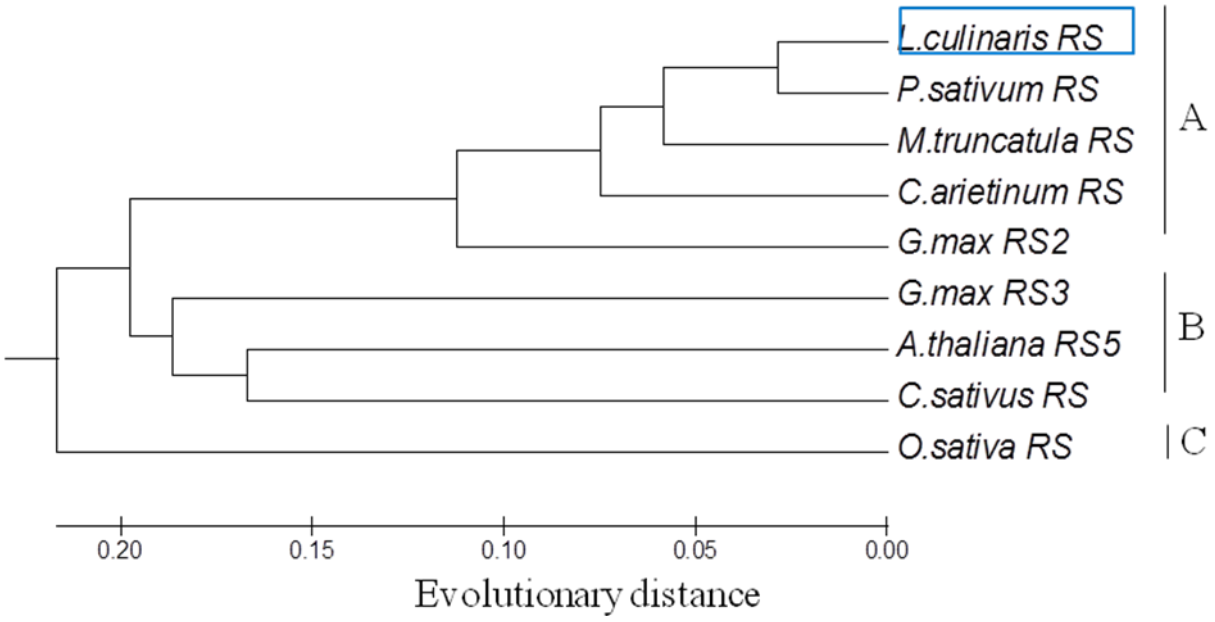


Figure 5.4: Phylogenetic analysis of raffinose synthase genes in different plant species. For phylogenetic analysis protein sequences used were *P. sativum* (CAD20127.2), M. *truncatula* (AES71465.1),*G. max* RS2(ACD13461.1), *A. thaliana* RS5 (NP_198855.1), *C. sativa* (ABD72603.1), *G. max RS3* (ACD13462.1, *O. sativa* (BAD68247.1) and *C. arietinum* (XP_004501888.1).

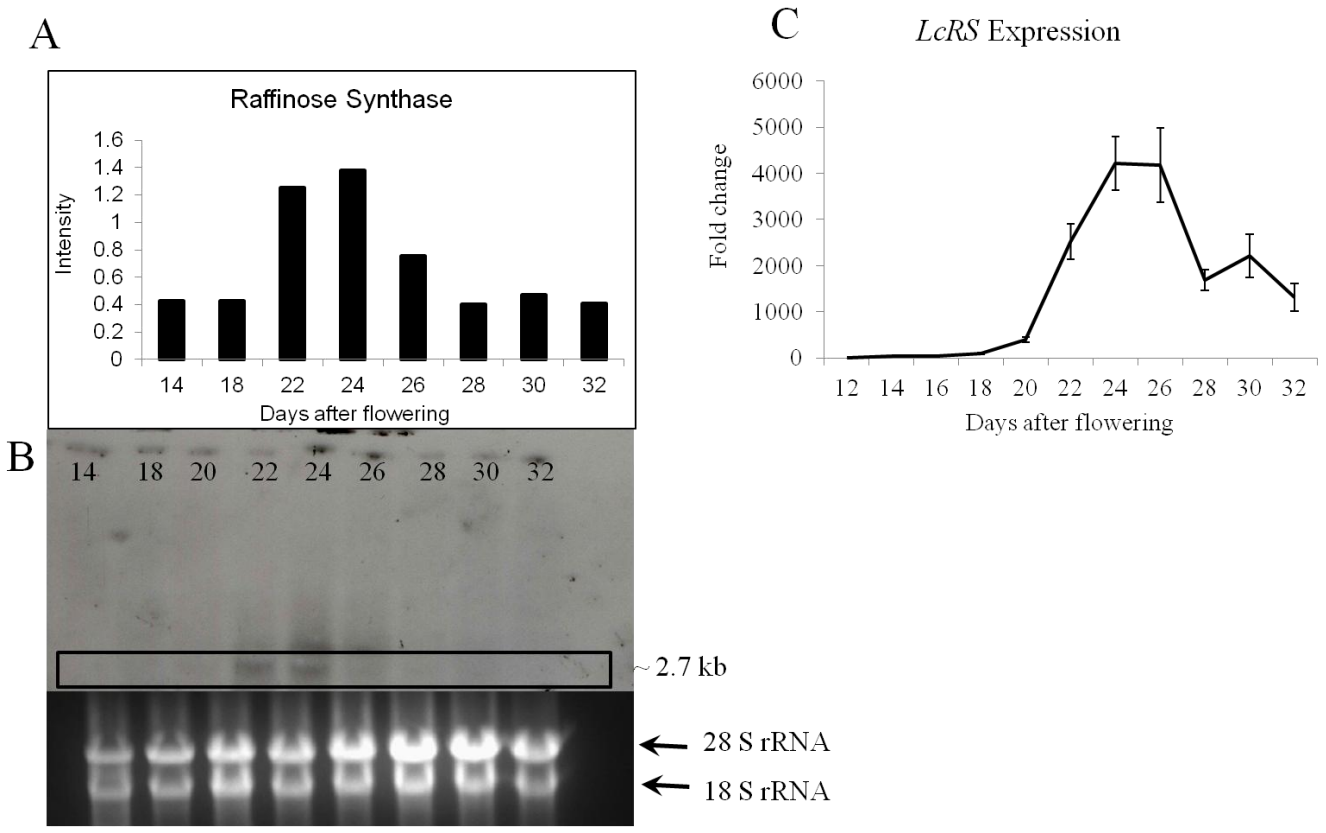


Figure 5.5: Transcript accumulation of raffinose synthase gene in *Lens culinaris* (A) Graph plotted based on intensity of transcript obtained in the RNA blot with background as the reference (B) RNA blot showing raffinose synthase transcript accumulation and 28S and 18S rRNA showing intact RNA observed in the gel used for RNA blot. (C) Gene expression of *LcRS* during seed development by quantitative real time PCR analysis. Fold change normalized using GAPD, EF1-alpha and HSP. Error bars indicate standard error of the mean of four determinations at each time point.

expression by qRT-PCR analysis where, transcript abundance increased from 20 DAF with maximum transcript accumulation at 24 and 26 DAF (Figure 5.5 B). A sharp decline in transcript accumulation was observed at 28 DAF and towards later stages of seed development. Quantitative real time PCR analysis reveals *LcRS* transcript at seed maturity (32 DAF) though this was not evident through RNA gel blot analysis (Figure 5.5). Gene encoding RS in *Pisum sativum* has been reported (Peterbauer et al 2013), however transcript accumulation was not reported in pea during seed development (Peterbauer et al 2001b). *LcRS* transcript was detected in RNA gel blot analysis only when 40 µg of total RNA was used. No transcript was detected when 20 µg total RNA was used though transcript accumulation patterns were obtained for both *LcGolS* and *LcSTS* using 20µg total RNA (Data not shown). This suggests that the LcRS transcripts are present at very low concentrations compared to *LcGolS*, *LcSTS*, or the RNA is very unstable.

5.4.4 Raffinose accumulation and raffinose synthase enzyme activity

Analysis of RS activity in developing lentil seeds from 12 to 38 DAF at two day intervals revealed that RS activity was not detected before 14 DAF (Figure 5.6C). RS activity increased rapidly from 16 DAF reaching maximum activity (11.0 pkat/mg protein) at 24 DAF. Beyond this stage a decrease in RS activity was observed at later stages of seed development. No activity was detected at 32 and 34 DAF, though presence of transcript was observed at these stages using qRT-PCR analysis (Figure 5.5C). RS enzyme activity of *Pisum sativum* showed a similar trend where maximum activity was observed in 25 - 30 DAF for pea line RRRbRb (Line containing high verbascose concentration in the seeds) and 20 DAF for pea SD line (Line containing low verbascose concentration in the seeds) (Peterbauer et al 2001b). Similar to the results obtained in lentil RS enzyme activity, no enzyme activity was observed during early stages of seed development in *Pisum sativum* (Peterbauer et al 2001b).

Carbohydrate accumulation using HPAEC-PAD detected sucrose at 10 DAF and a gradual increase in concentration was observed, showing maximum accumulation at 16 DAF (7.4 g/100g) (Figure 5.6). A gradual decrease in sucrose concentration was observed after 16 DAF and it reached a plateau towards late stages of seed development (36 DAF, 38 DAF-concentration 1.2 g/100g). Raffinose was first detected at 10 DAF (0.1 g/100g) and it remained at similar concentrations till 18 DAF, after which it steadily increased to reach a maximum

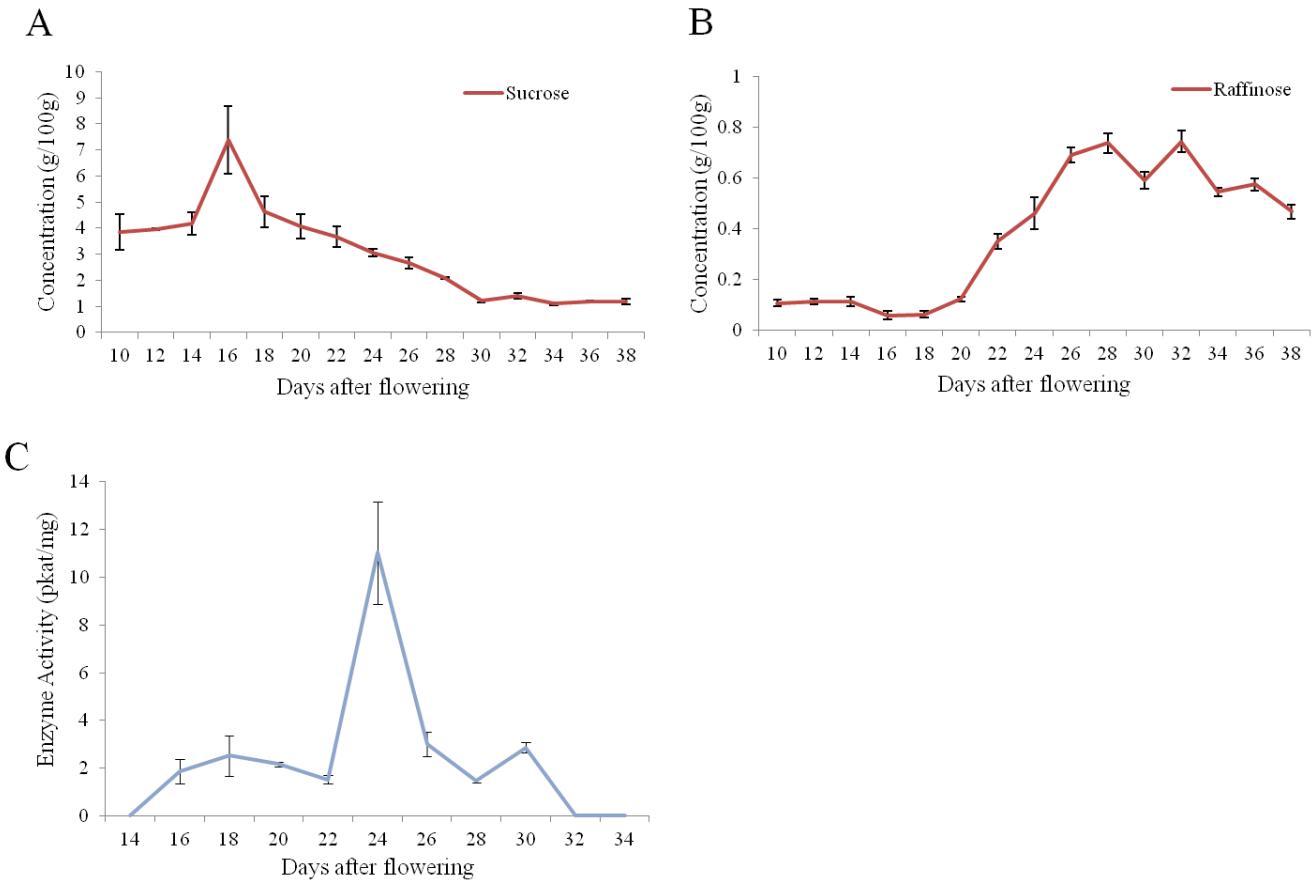


Figure 5.6: Accumulation of sucrose, raffinose and raffinose synthase enzyme activity during seed development. (A) Sucrose accumulation during seed development (g/100 g seed dry weight) (B) Raffinose accumulation during seed development in g/100 g (C) Raffinose synthase (RS) activity of during seed development (pkat / mg protein). Error bars indicate standard error of the mean for six determinations at each time point.

concentration (0.6 g/100g dry weight) at 26 – 32 DAF (Figure 5.6 B). Raffinose concentration decreased slightly in mature seeds (0.5 g/100g dry weight). A strong correlation was observed between sucrose and raffinose concentration in developing seeds; decrease in sucrose accumulation corresponded with raffinose accumulation in seeds (Figure 5.6). Similar trend in raffinose accumulation was observed in *Pisum sativum* during seed development where rapid accumulation of raffinose was observed at 25 - 30 DAF for line RRRbRb (Line with high verbascose concentration) and 20 - 25 DAF for SD1 (Line with low verbascose concentration). Similar accumulation of sucrose was also observed in *P. sativum* where maximum concentration for sucrose was observed at 20 DAF for line RRRRbRb and 15 DAF for SD1 (Peterbauer et al 2001b). A correlation between decrease in sucrose accumulation and increase in raffinose accumulation was also observed in developing pea seeds (Peterbauer et al 2001b).

In conclusion, a gene coding for *LcRS* of length 2776 coding for polypeptide with 792 amino acid residues was obtained. Gene expression of *LcRS* obtained by RNA gel blot analysis and qRT-PCR analysis correlated with RS enzyme activity during seed development. Accumulation of raffinose correlated to RS enzyme activity. These results show that gene isolated might contribute to raffinose accumulation in the seeds. Purification of RS enzyme or expression of the enzyme *in vitro* might facilitate more understanding of enzyme properties in lentil.

6. ANALYSIS OF STACHYOSE AND VERBASCOSE ACCUMULATION IN *Lens culinaris* Medik. DURING SEED DEVELOPMENT.

Study 4

In this study, a gene coding for stachyose synthase (*LcSTS*) coding for 853 amino acids was isolated. *LcSTS* gene expression, enzyme activity of stachyose synthase, verbascose synthase and galactan : galactan galactosyl transferase and accumulation of stachyose and verbascose was studied during lentil seed development.

Kannan U, Sharma R, Gangola M, Khedikar Y, Ganeshan S, Chibbar RN (2014) Analysis of stachyose and verbascose accumulation in *Lens culinaris* Medik. during seed development (to be submitted).

6.1 Abstract

Stachyose synthase (STS; EC 2.4.1.67) catalyzes the synthesis of stachyose from raffinose and galactinol. A STS cDNA clone of length 2669 bp (5'UTR - 30 bp, ORF - 2559 bp, 3'UTR - 80 bp followed by polyA) coding for 853 amino acids with predicted molecular weight of 96 kDa was isolated. During seed development analysis of gene expression by RNA gel blot and qRT-PCR analysis indicate maximum accumulation at 26 and 24 days after flowering (DAF) respectively. Transcript accumulation correlated with STS enzyme activity, where highest STS enzyme activity was observed at 26 DAF. Similarly, highest Verbascose synthase (VS) activity was observed at 26 DAF. However, galactinol independent enzyme (galactan galactan galactosyl transferase) activity for RFO synthesis was observed at later stages of seed development (26 - 32 DAF), corresponding to rapid accumulation of stachyose and verbascose in lentil seeds. During lentil seed development stages, stachyose and verbascose correlate both with transcript accumulation for STS gene and enzyme activity for STS, VS and galactan : galactan galactosyl transferase enzyme activity.

6.2 INTRODUCTION

Raffinose and stachyose is present in most legume seeds, with stachyose being the predominant RFO in several species including soybean (*Glycine max*) (Saravitz et al 1987), common bean (*Phaseolus vulgaris*) (McPhee et al 2002), lupin (*Lupinus albus*) (Martìnez-Villaluenga et al 2005), chickpea (*Cicer arietinum*) (Gangola et al 2013), some genotypes of lentil (*Lens culinaris*) (Tahir et al 2011) and garden pea (*Pisum sativum*) (Peterbauer et al 2001b). The total RFO concentration show limited variation (4.0 to 6.1 nmoles/100g flour) in domesticated *Lens* genotypes, compared to the wild species that show much larger variation (1.7 to 5.2 nmoles/100g flour) (Tahir et al 2011). The three major constituents of RFO, also show much larger variation in the genus Lens. In the domesticated Lens genotypes, raffinose (1.5 to 3.0 g/100g flour) and stachyose (1.5 to 3.0 g/100g flour) are present in similar concentrations, and verbascose is present in lower concentration (1.1 to 1.6 g/100g flour) (Tahir et al 2011). However, *Lens ervoides* genotypes have higher verbascose and lower stachyose concentrations. In contrast, *Lens nigricans* genotypes have high raffinose and low verbascose concentrations (Tahir et al 2011).

Stachyose and verbascose are synthesized by addition of a galactosyl residue from

galactinol to raffinose and stachyose, respectively, catalyzed by the enzymes stachyose synthase (STS; EC 2.4.1.67) and verbascose synthase (VS) respectively. In contrast to raffinose synthase, STS enzyme has been studied in several legume species. STS has been purified to apparent homogeneity from seeds of *Phaseolus vulgaris* (Tanner and Kandler 1968), *Vigna angularis* (Peterbauer and Richter 1998), *Lens culinaris* (Hoch et al 1999), *Pisum sativum* (Peterbauer et al 2002b) and leaves of *Cucumis sativa* and *Cucumis melo* (Huber et al 1990; Holthaus and Schmitz 1991). All purified enzymes show a pH optima between 6.5 and 7.0. Immunoblot analysis of protein from various legume species (seeds from *L. culinaris, G. max, C. arietinum*, various *Vigna* species and leaves of *Cucurbita pepo*) showed a significant difference in STS molecular weights ranging from 85 - 95 kDa (Richter et al 2000).

STS purified from lentils is monomeric with molecular weights of 88.6 kDa and pI of 4.8 (Hoch et al 1999). Lentil STS also catalyzed a range of galactosyl transfer reactions synthesizing galactopinitol A and ciceritol (Hoch et al 1999). STS purified from *P. sativum* seeds with high verbascose content was also shown to catalyze verbascose synthesis by transfer of galactosyl residue from galactinol to stachyose (Peterbauer et al 2002b; Peterbauer et al 2003). However purified lentil and adzuki bean STS did not catalyze verbascose synthesis (Hoch et al 1999). Hence presence of a chain elongation enzyme VS in lentil was suggested.

Though purification of lentil STS has been reported (Hoch et al 1999), no nucleotide sequence for a complete gene coding for STS is available. In a recent transcriptome assembly, a contig (1.3 kb) corresponding to STS gene has been reported (Kaur et al 2011); that does not compare to full length STS genes. In this study a gene coding for STS was isolated and characterized and its expression during lentil seed development was analyzed. Pattern of RFO accumulation during lentil seed development and enzyme activities suggest that in addition to STS and VS, a galactinol independent enzyme also contribute to stachyose and verbascose synthesis in developing lentil seeds.

6.3 MATERIALS AND METHODS

6.3.1 Plant Material

Lens culinaris cv CDC Redberry plants were grown in growth chambers with temperatures 21 ^{o}C during the day and 15 ^{o}C during the night with 18 h photoperiod and light intensity [250 mmol, $m^{-2}s^{-1}$ PPFD (photosynthetic photon flux density)]. Young leaves were

collected for genomic DNA extractions. Fully opened flowers were tagged and developing seeds were collected from 10 DAF to 38 DAF at two day intervals. All samples were frozen in liquid nitrogen and stored at -80 ^{o}C until used for analysis.

6.3.2 Isolation of partial *LcSTS* sequence, cDNA library screening

Genomic DNA was extracted from young lentil leaves (200 mg) using CTAB based DNA isolation procedure (Doyle and Doyle 1990). Partial DNA sequence of *LcSTS* was isolated by PCR amplification using degenerate primers Forward DG/STS_F1 and DG/STS_R1 (Appendix 4.2). PCR reaction comprised of 20 mM Tris-HCl (pH 8.4), 50 mM KCl, $MgCl_2$ (1.5mM or 2.5 mM or 3.5 mM), primer concentrations (0.2 µM and 0.4 µM), 0.2mM each dNTP and 3 units Taq Polymerase (Life Technologies Inc., Burlington, Ontario, Canada). Thermocycler conditions were: initial denaturation at 95 ^{o}C 4 min 15 sec, followed by 32 cycles of 95 ^{o}C for 45 sec, 20 sec annealing temperature (tested at 54 ^{o}C, 57 ^{o}C and 60 ^{o}C), 1 min 30 sec extension at 72 ^{o}C and final cycle at 72 ^{o}C at 10 min. The amplified DNA fragments were analyzed by agarose (1% w/v) gel electrophoresis (Appendix 6.1). The amplified fragment was compared with *Pisum sativum* STS1 for nucleotide identity and presence of introns in the sequence. The amplified sequence was used to search for sequence similarity in the *Lens culinaris* transcriptome assembly. Sequence similarity search in the transcriptome assembly in the database yielded a 1.3 kb fragment which was cloned (CloneJET PCR Cloning Kit, Thermo Fisher Scientific Inc., Pittsburgh, PA, USA) and DNA sequenced. Spidey (NCBI database) was used to identify intron / exon splice site junctions in the sequence between genomic DNA fragment and 1.3 kb contig from transcriptome assembly (Kaur et al 2011). Plasmids containing the 1.3 kb insert were digested with restriction enzymes *BsrG1* and *XbaI* to obtain a fragment of 663 bp (Figure 6.1A) and was used as a probe for cDNA library screening.

Screening of cDNA library ($1x10^9$ pfu/mL) with the 663 bp fragment as the probe was done as per manufacturer's instructions (Stratagene Agilent Technologies Inc., Santa Clara, CA, USA). Plaque lifting and hybridization was performed as per standard protocols (Sambrook and Russell 2001). Partial STS was labelled with alpha ^{32}P- dCTP using random primer labelling kit (Life Technologies Inc., Burlington, Ontario, Canada). Positive plaques were selected based on appearance of spots in replicate autoradiographs. Plaques were cored from agar plates and secondary screening was performed on the clones as per manufacturer's protocol. Positive clones

obtained from secondary screening were excised using ExAssist SOLR system (Agilent Technologies Inc., Santa Clara, CA). Phagemids were digested with restriction enzymes to check the length of the insert and sent for DNA sequencing (National Research Council, Saskatoon, Canada).

6.3.3 RNA gel blot and quantitative PCR analysis

Total RNA was isolated from developing seeds (200 mg) using CTAB-LiCl method (Kannan et al 2014). For RNA gel blot analysis, RNA (20µg) was electrophoresed in agarose gel with formaldehyde. 28S and 18S rRNA was visualized under UV for examining quality and equal loading of total RNA. The RNA was transferred to Hybond-N^+ membrane (GE Healthcare, Baie d'Urfe, Quebec, Canada) using a positive pressure blot, PosiBlot^{TM} apparatus (Agilent Technologies Inc., Santa Clara, CA, USA) for 2 h at 75mm Hg pressure. RNA was UV cross linked and hybridized with partial STS sequence (Figure 6.1A, Probe 1) radiolabelled with α-^{32}P-dCTP using random primer labelling kit (Life Technologies, Inc., Burlington, Ontario, Canada). Hybridization and washing of the membranes was performed using standard protocols (Sambrook and Russell 2001). Membranes were exposed to Biomax MR film (KODAK® BioMax® Autoradiography Film) at -80 °C for 24 h to several days depending on observed signal intensities.

For qPCR, total RNA (5µg) was DNAse treated (Life Technologies, Inc., Burlington, Ontario, Canada) and 1µg RNA was used for cDNA synthesis. cDNA was synthesized using 500 ng Oligo $dT_{(12\text{-}18)}$ (Life Technologies, Inc., Burlington, Ontario, Canada), 900 ng of random primer (Life Technologies Inc., Burlington, Ontario, Canada). 5µl of a 1:15 dilution cDNA was used a template with 300 mM primers and 2x Maxima^{TM} SYBR Green I qPCR Master Mix (Fisher Scientific Company, Ottawa, Ontario, Canada). Gene specific primers for *LcSTS* forward STS1RT_F and reverse STS1RT_R (Appendix 6.3) were used to amplify a STS amplicon. GAPD (Gyceraldehyde-3-phosphate dehydrogenase), HSP (Heat Shock Protein 90) and EF1-α (Elongation Factor 1 alpha) were used as reference genes (Appendix 6.3). No template control and no reverse transcriptase control were included in the analysis. Relative expression was calculated by $2^{-\Delta\Delta Ct}$ method (Livak et al 2001) using 12 DAF sample as the calibrator. The experiment was repeated twice with two technical replicates.

6.3.4 Phylogenetic Analysis

Sequence of STS genes from different plant species were obtained from GenBank (NCBI) database. Sequence alignments were performed using Clustal Omega (Sievers et al 2011). A phylogenetic tree was constructed using MEGA 6 (Tamura et al 2013). The UPGMA (Unweighted Pair Group Method with Arithmetic Mean) and Poisson correction method were used to calculate evolutionary distances.

6.3.5 STS enzyme activity assay

STS enzymatic activity was determined in crude cell extracts incubated with substrates and the stachyose or verbascose synthesis was assayed using HPAEC-PAD (Peterbauer et al 1998; 2001b; Hitz et al 2002) with slight modifications (Gangola 2014). To extract cellular proteins, ground seeds (200 mg) were added to 2 mL of extraction buffer (50mM HEPES (2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid) - NaOH, pH 7.0 and 5mM 2-β mercaptoethanol). The cell extracts were centrifuged at 10000 x g for 10 min at 4 ^{o}C. The supernatant was desalted by passing the supernatant in a pre-equilibrated Sephadex G25 column. Protein concentration was estimated by dye binding assay (Bradford 1976) using a commercial kit (Bio-Rad protein assay, Bio-Rad Laboratories Inc., Mississauga, Ontario, Canada) with bovine serum albumin as standard. The STS assay conditions, pH, protein concentration, temperature and substrate concentration were same as previously optimized for *Cicer arietinum* (Gangola 2014). Substrate concentrations used were: STS - 10 mM galactinol, 10 mM DTT, 20 mM raffinose; VS - 10 mM galactinol, 10 mM DTT, 20 mM stachyose; galactinol independent activity (galactan galactan galactosyl transferase) - 10 mM DTT, 20 mM raffinose. Thirty microgram of crude protein extract was used for STS enzyme assay and 40 µg was used for VS and GGT enzyme assays. Control reactions were: (i) no substrate with cell extract (ii) substrate with no cell extract. Reaction mixtures were incubated at 25 ^{o}C for 60 min after which the assay was terminated by adding 50 µL of 95% (v/v) ethanol. To denature the proteins, reaction mixture was incubated for 1 min in a boiling water bath. The tubes were centrifuged at 10,000 x g for 10 min and supernatant was transferred to a new tube. Ten mg of Dowex 50WX8 100 ion exchange resin (Sigma-Aldrich Inc., St. Louis, Missouri, USA) was added to the supernatant and shaken at 200 rpm for 30 min. The filtrate was transferred to a new tube and supernatant was dried using a Speedvac concentrator and universal vacuum system (Thermo Savant, Holbrook, NY, USA).

The pellet was dissolved in 500 µL water and mixed well. The tubes were subsequently centrifuged at 10,000 x g for 10 min and 125 µL of the supernatant was added to 375 µL of water. The substrates and products were analyzed using Dionex ICS 5000 (Dionex ion chromatography system 5000) as described in Gangola et al 2014.

6.3.6 Determination of stachyose and verbascose concentration

RFO concentrations were determined using HPAEC-PAD (high performance anion exchange chromatography with pulse amperometric detection) as described in (Gangola et al 2014). Developing lentil seed samples were freeze dried and ground. Flour samples (100 mg) were extracted with 4 mL 80% (v/v) ethanol. The supernatants were passed through pre-equilibrated C18 solid phase columns (Honeywell Burdick and Jackson, MI, USA) and then dried using speedvac concentrator and universal vacuum system (Thermo Savant, Holbrook, NY, USA). The dried samples were resuspended in 500 µL nanopure water and centrifuged at 12,000 x g for 10 min. 125 µL of the sample was added to 375 µL of water and analyzed using Dionex ICS 5000 Columns (Dionex Canada Ltd., Oakville, ON, Canada) ; CarboPac PA 100 and CarboPac PA guard column (Thermo Fisher Scientific Inc., Stevens Point, WI, USA).

6.4 RESULTS AND DISCUSSION

6.4.1 Isolation of *LcSTS* sequence

Degenerate oligonucleotide primers designed from conserved regions in the STS gene nucleotide sequences in data base, were used in a PCR reaction with lentil genomic DNA as a template. PCR amplification yielded 884 bp (Appendix 6.1) fragment that on DNA sequencing showed high sequence identity to STS nucleotide sequences. The amplified *LcSTS1* showed 96% sequence identity to *Pisum sativum* RS sequence in the GenBank database. However 49% query coverage was obtained suggesting the presence of introns in the amplified sequence. The amplified sequence aligned from 1617 - 1687bp and 1686 - 2054 bp to *Pisum sativum* STS1 sequence. The 884 bp genomic sequence of *LcSTS* when searched against the *Lens culinaris* transcriptome shotgun assembly database (Kaur et al 2011) revealed a contig (# 06676 accession Number gb|JI852941.1|) of 1373 bp corresponding to partial STS sequence. The contig sequence was identical to the isolated *LcSTS* sequence with the exclusion of introns present in the amplified sequence from genomic DNA (Sequence alignment between genomic fragment and

LcSTS contig shown in Appendix 6.2). From lentil cDNA, PCR amplification resulted in a 1.3 kb fragment that on DNA sequencing was confirmed to be partial *LcSTS*. Partial cDNA fragments of the 1.3 kb *LcSTS* were used as probes for RNA gel blot analysis and cDNA library screening (Probe indicated in Figure 6.1).

Primary screening of the cDNA library with partial *LcSTS* yielded 53 putative positive clones carrying STS cDNA insert. Presence of insert was confirmed by amplification using gene specific primers (amplicon size 1337 bp) and 38 of the 53 clones revealed the presence of insert. Secondary screening using 38 clones revealed 29 positive clones. Positive clones obtained by secondary screening were *in-vivo* excised and 62 clones (phagemids) were DNA sequenced. Sequence analysis of 62 phagemids (obtained from cDNA library) showed only four with STS cDNA insert while all other clones showed similarity to *Pisum sativum* starch synthase gene, *Pisum sativum* aldose gene, CVC gene convicilin from *Lens culinaris*, mRNA for legumin B in *Vicia sativa* and uncharacterized mRNA. A *LcSTS* sequence (2693 bp) was isolated, consisting of 5'UTR - 30bp, ORF - 2559 bp and 3' - UTR - 101bp (Figure 6.1A). *LcSTS* coded for 853 amino acids with predicted molecular weight of 96 kDa and pI of 5.57. Pfam search (Punta et al 2014) for conserved domains showed active sites for raffinose synthase or seed imbibition protein belonging to the Clan 0058 proteins, which represent Tim barrel glycosyl hydrolase superfamily.

LcSTS1 showed 91% sequence identity to *Pisum sativum STS*, 82% similarity to putative *Cicer arietinum* STS, and 81% sequence similarity to *Medicago truncatula* STS, and 76% similarity to putative *G. max* STS isoforms 1 and 2, 74% identity to predicted *Phaseolus vulgaris* STS and *V. angularis* STS at amino acid level (Figure 6.2). Phylogenetic analysis of *LcSTS* with STS genes from diverse plant genes (Figure 6.3) showed three distinct clades A, B, C. *Leguminoseae* species *Pisum sativum*, *LcSTS1*, *Medicago truncatula*, *Vigna angularis* and *Cicer arietinum* clustered in clade A. *LcSTS* showed high identity to *Pisum sativum STS* suggesting close evolutionary relationship and similar function to *Pisum sativum* STS.

STS purified from adzuki bean (*Vigna angularis*) was a single polypeptide with molecular mass of 90 kDa (Peterbauer and Richter 1998). Hoch et al 1999 purified STS from lentil and size exclusion chromatography estimated its size to be 99 kDa, however SDS-PAGE showed the presence of a protein of molecular mass of 88.6 kDa (Hoch et al 1999; Richter et al 2000). Melon leaf STS showed the presence of two polypeptide of sizes 45 and 50 kDa with pI

```
gcacgaggctcttaaccatttcttgtttccatggctcctcccttgaattcccccacccat 60
                                M  A  P  P  L  N  S  P  T  H  10
gatctcatcaaaactgagtctttgtttgatttatctgaaagaaagtttaaggtgagaggt 120
 D  L  I  K  T  E  S  L  F  D  L  S  E  R  K  F  K  V  R  G  30
gttccttttgttccatgatgttcctgaaaatgttactttcagctcgttcttttctatatgc 180
 V  P  L  F  H  D  V  P  E  N  V  T  F  S  S  F  F  S  I  C  50
aaaccttctgaatcaaatgcaccaccttcacttcttcaaaaagttctttccttttcacac 240
 K  P  S  E  S  N  A  P  P  S  L  L  Q  K  V  L  S  F  S  H  70
aaaggtggtttctttggattctctcatgaaaccccatcagatagattgatgaattcctta 300
 K  G  G  F  F  G  F  S  H  E  T  P  S  D  R  L  M  N  S  L  90
ggtagtttcaatggaaaagactttgttagtattttcaggttcaaaacatggtggtctact 360
 G  S  F  N  G  K  D  F  V  S  I  F  R  F  K  T  W  W  S  T  110
caatggataggaaaatctggttctgatttacaaatggaaactcaatggattctcattgaa 420
 Q  W  I  G  K  S  G  S  D  L  Q  M  E  T  Q  W  I  L  I  E  130
gttcctgaaaccaaatcctatgttgtaatcattcctattatagaaaaatgtttcaggtct 480
 V  P  E  T  K  S  Y  V  V  I  I  P  I  I  E  K  C  F  R  S  150
gcacttttttcctggttttaatgatcatgtcatgatttgtgctgaaagtggttcaacaaag 540
 A  L  F  P  G  F  N  D  H  V  M  I  C  A  E  S  G  S  T  K  170
gtgaaagaatcaacattcaattcaattgcttatgttcatttttctgaaaatccttatgat 600
 V  K  E  S  T  F  N  S  I  A  Y  V  H  F  S  E  N  P  Y  D  190
ttgatgaaagaagcttatagtgttataagggttcatctcaatacttttaggcttttggaa 660
 L  M  K  E  A  Y  S  V  I  R  V  H  L  N  T  F  R  L  L  E  210
gagaaaacaatcccaaatttagttgataaatttggttggtgcacatgggatgctttctat 720
 E  K  T  I  P  N  L  V  D  K  F  G  W  C  T  W  D  A  F  Y  230
ttaactgttaatcctattggtattttccatggtcttgatgagttcttgaaaggcggtgtc 780
 L  T  V  N  P  I  G  I  F  H  G  L  D  E  F  L  K  G  G  V  250
gaaccgaggtttgttatcattgatgatggatggcaaagcattaattttgatggttgtgat 840
 E  P  R  F  V  I  I  D  D  G  W  Q  S  I  N  F  D  G  C  D  270
cctaatgaagatgcaaagaatcttgttcttggtggtgaacaaatgacgggtagacttcat 900
 P  N  E  D  A  K  N  L  V  L  G  G  E  Q  M  T  G  R  L  H  290
aggtttgaagaatgtgacaagttcaaaaaatatgaaagtgggttgttattaggtcctaat 960
 R  F  E  E  C  D  K  F  K  K  Y  E  S  G  L  L  L  G  P  N  310
tcacctacttatgatcccaaaaaatttagggatttgattttaaagggtattgaacatgag 1020
 S  P  T  Y  D  P  K  K  F  R  D  L  I  L  K  G  I  E  H  E  330
aaattgaagaagaaaaaagagaaagctattttatcaaagagttctgatttagttgaaata 1080
 K  L  K  K  K  K  E  K  A  I  L  S  K  S  S  D  L  V  E  I  350
gagtcaaagatcaaggaagtagtaaaagaaattgatgatctctttggtggagaacaactt 1140
 E  S  K  I  K  E  V  V  K  E  I  D  D  L  F  G  G  E  Q  L  370
agcagtgttcaaaaaagtgagttgaagagtgagtatggattgaaggctttcacaaaggat 1200
 S  S  V  Q  K  S  E  L  K  S  E  Y  G  L  K  A  F  T  K  D  390
ttgaggactaaattcaaaggtttggatgatatttatgtttggcatgcactttgtggtgct 1260
 L  R  T  K  F  K  G  L  D  D  I  Y  V  W  H  A  L  C  G  A  410
tggggtggtgtgaggccagaaaccacacaccttaataccaaaattgtcccttgcaaaagc 1320
 W  G  G  V  R  P  E  T  T  H  L  N  T  K  I  V  P  C  K  S  430
tcacctggtcttgaagggacaatggaggatactgcagtggttgagattttcaaaggttca 1380
 S  P  G  L  E  G  T  M  E  D  T  A  V  V  E  I  F  K  G  S  450
attgggctagttcatcctagtcaagctaatgaactttatgattccatgcattcttatctt 1440
 I  G  L  V  H  P  S  Q  A  N  E  L  Y  D  S  M  H  S  Y  L  470
gctgaatctggtatcacaggagtcaaagttgatgtgatgcattctcttgaatatgtttgt 1500
 A  E  S  G  I  T  G  V  K  V  D  V  M  H  S  L  E  Y  V  C  490
```

Figure 6.1: *LcSTS* sequence

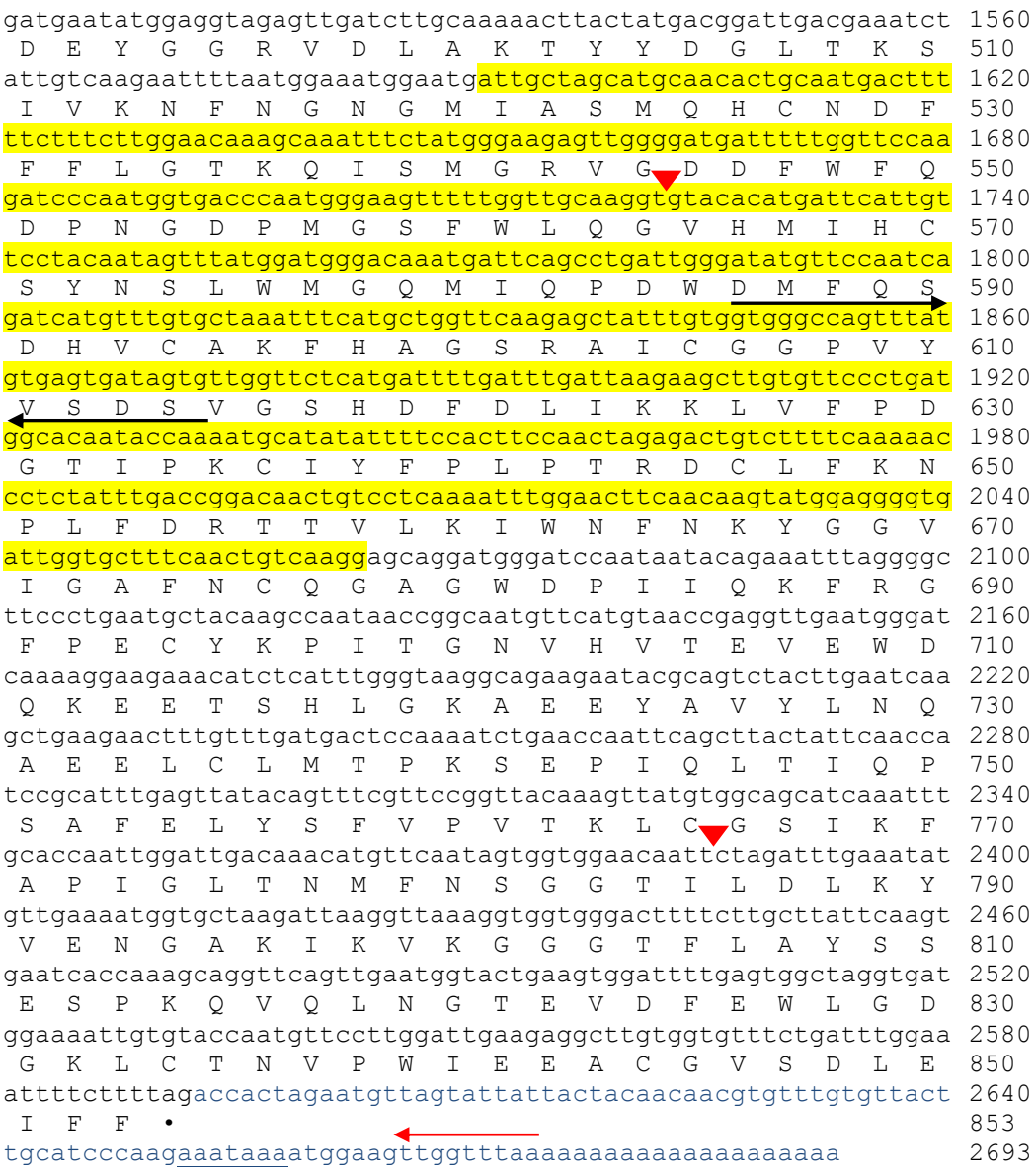


Figure 6.1: *LcSTS* sequence

Nucleotide and deduced amino acid sequences encoding putative LcSTS in lentil. The deduced protein sequence of *LcSTS* is indicated below nucleotide sequence ('•' indicates translation termination site). Nucleotides coloured in blue indicate 5' and 3' UTR regions flanking the ORF of *LcSTS*. Highlighted nucleotide region indicate partial *LcSTS* fragment obtained by amplification using degenerate primers. Forwards and reverse arrows (black) in the highlighted region indicate primers used for qRT-PCR analysis in this study. Partial gene sequence (1.3 kb) obtained from lentil transcriptome assembly is represented within red arrows in the sequence. '▼' in red indicate nucleotide region of LcSTS used as probes cDNA library screening and RNA gel blot analysis. Polyadenlyation signal is indicated in the 3'UTR region upstream of the poly(A) site.

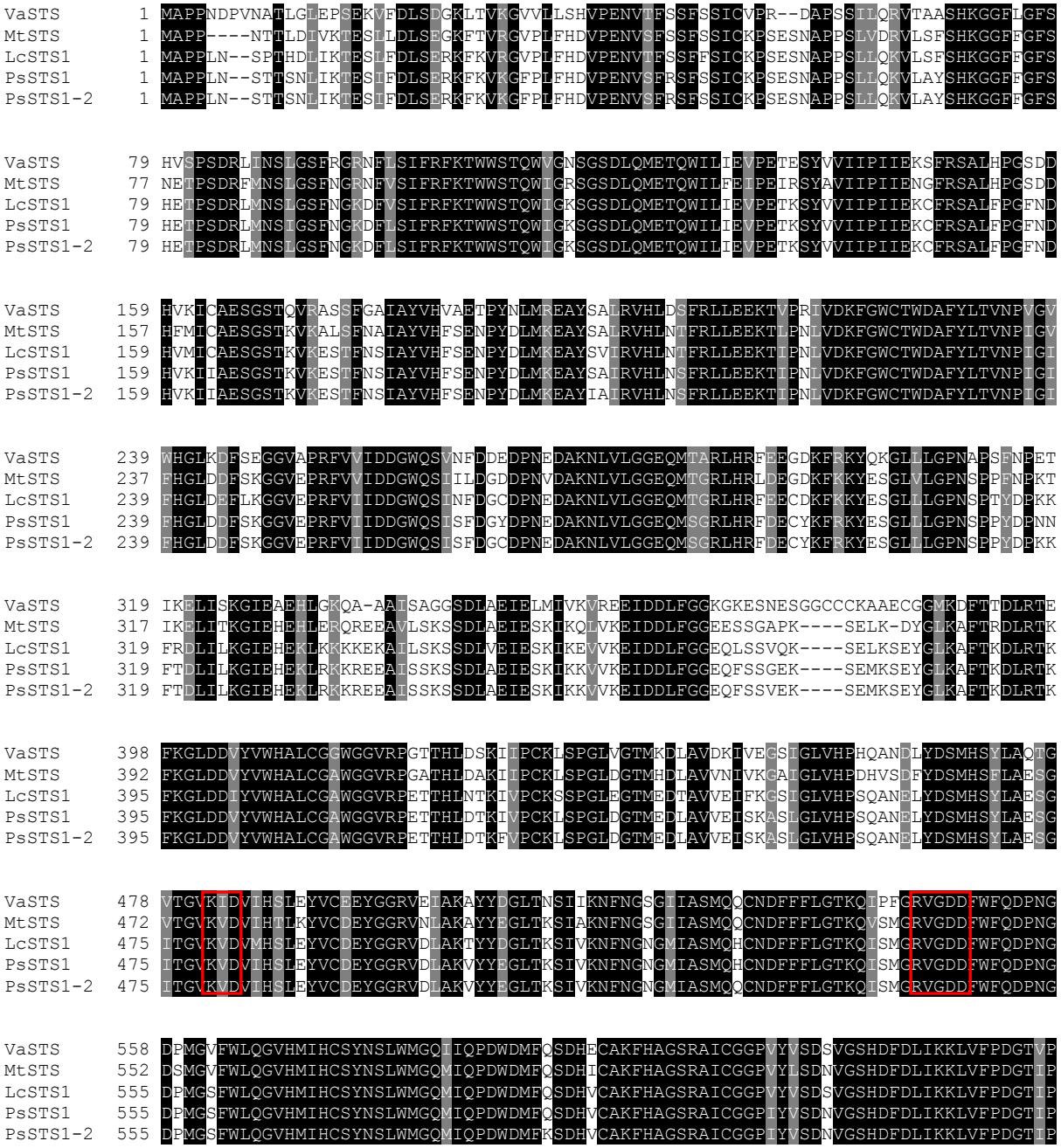


Figure 6.2: Sequence alignment of amino acids of STS gene of closely related species

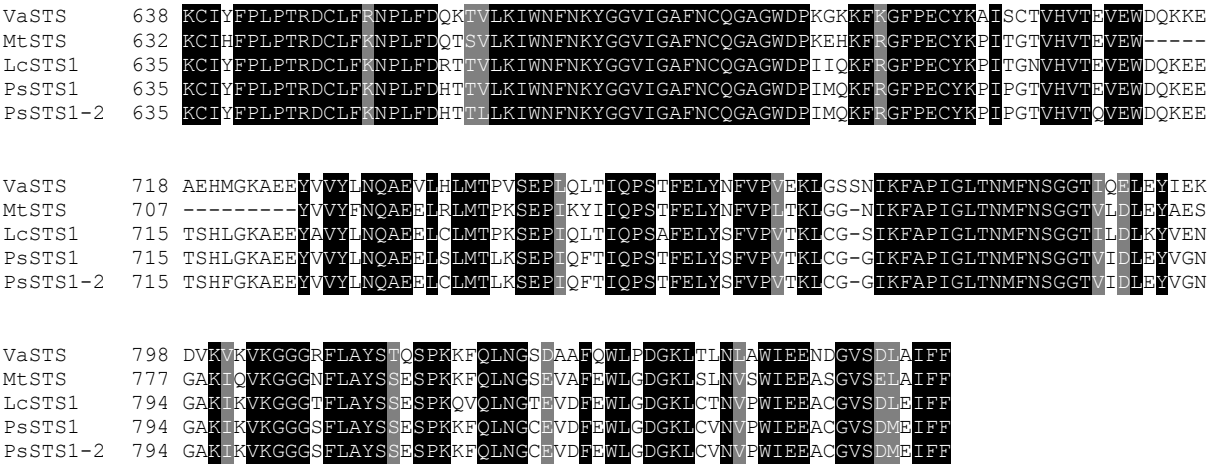


Figure 6.2 Continued: Sequence alignment of amino acids of STS gene of closely related species. Highly conserved amino residues are highlighted in black and similar residues are highlighted in gray. Amino acid residue indicated within red boxes indicate conserved motifs KxD, RXXXD.

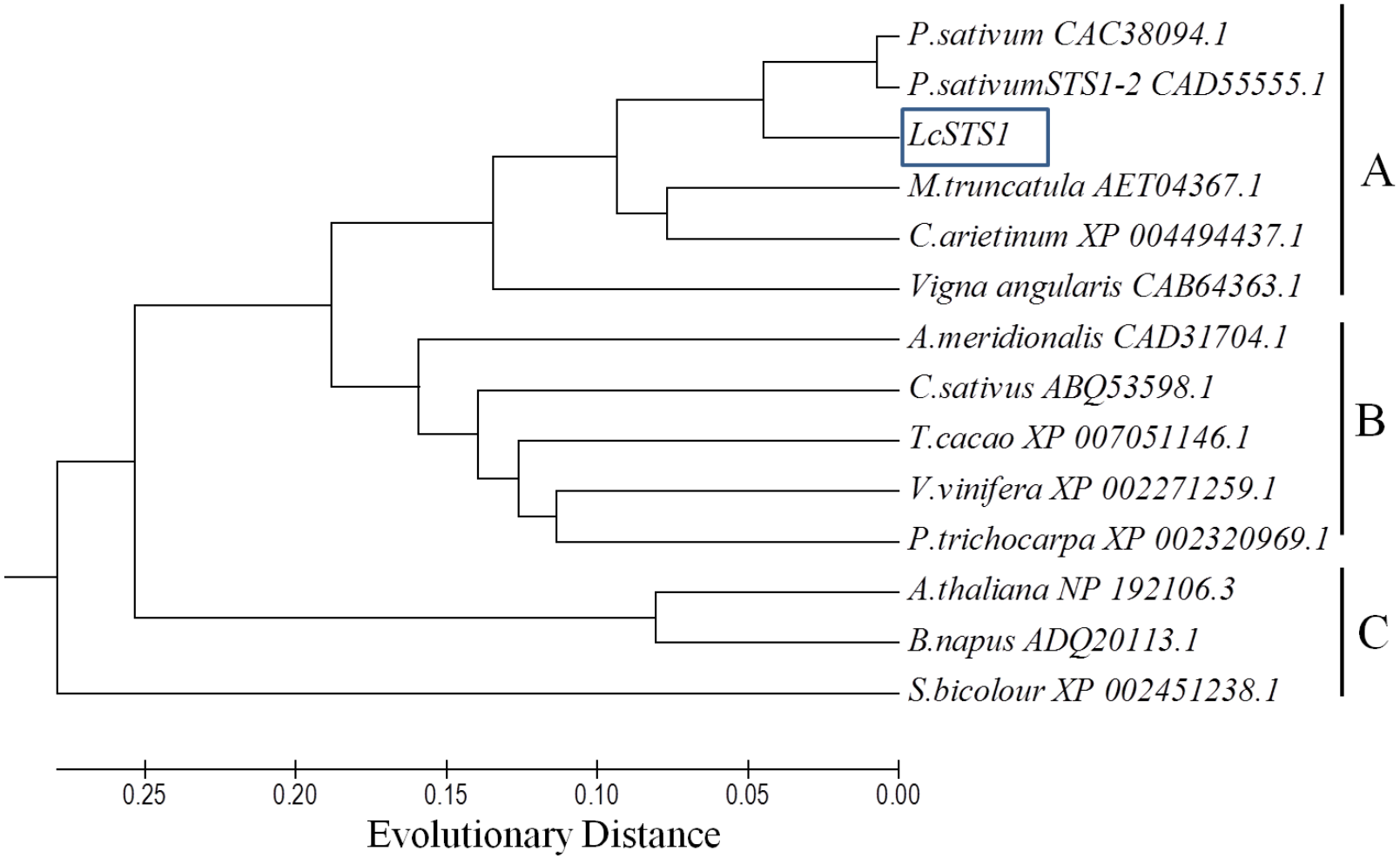


Figure 6.3: Phylogenetic analysis of STS genes from different plant species. Accession numbers were presented along with name of the species.

of 5.1 (Holthaus and Schmitz 1991; Richter et al 2000).

6.4.2 Transcript accumulation of *LcSTS*

RNA gel blot analysis using partial *LcSTS* used as a probe (Figure 6.1) produced a strong hybridization signal to a 2.7 kb transcript (Figure 6. 4 A, B). A low intensity hybridization signal of *LcSTS* was observed at early stages of seed development (18, 20 and 22 DAF). A sharp increase in transcript accumulation of *LcSTS1* (Figure 6.4 B) was observed at 24 DAF with maximum accumulation at 26 DAF, after which a gradual decrease in transcript accumulation was observed from 26 DAF to 30 DAF. No hybridization signal for 2.7 kb transcript was observed at 32 and 34 DAF (Figure 6.4B). Transcript accumulation of *LcSTS* using quantitative real time PCR analyses showed slight increase in transcript accumulation from 18 DAF, with a sharp increase in transcript accumulation at 22 DAF with maximum accumulation at 24 DAF (Figure 6.4C). Beyond this stage, gene expression declined and only a very small amount of transcript could be detected in mature seeds (32 and 34 DAF). This trend in transcript accumulation observed using qRT-PCR analysis was similar to RNA gel blot results though, it showed maximum accumulation at 26 DAF in RNA gel blot analysis. This difference could be attributed to difference in biological replicates with samples obtained from plants grown in different batches. Seed samples used for oligosaccharides analysis, RNA blot and activity assays were obtained from the same batch (grown at the same time). The *LcSTS* transcript accumulation observed was similar to *STS* transcript accumulation in *Vigna angularis*, where maximum accumulation was observed at 23 - 25 DAF (Peterbauer et al 1999). Similar transcript accumulation pattern (highest at 26 DAF) was also reported in *Pisum sativum* line RRRbRb, line with higher verbascose concentration (Peterbauer et al 2001b**).**

6.4.3 Enzyme activity of stachyose synthase

STS enzyme activity was analyzed during seed development (Figure 6.5) with galactinol as the galactosyl donor in the reaction. Low enzymatic activity was obtained at early stages of seed development increasing at 18 DAF and rapid increase was observed at 24 DAF with maximum STS activity at 26 DAF (74 ± 10.2 pkat/mg protein). Beyond this stage, activity decreased towards later stages of seed development. However STS activity was detected in

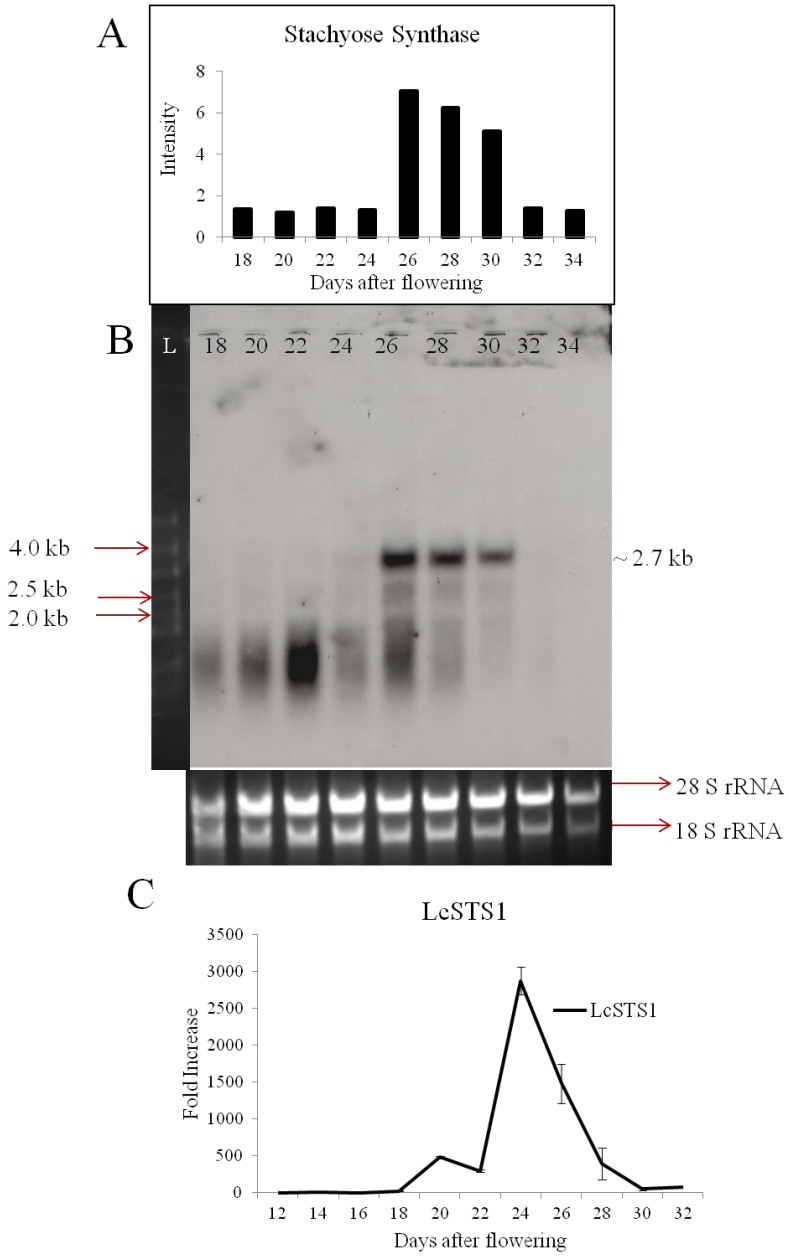


Figure 6.4: Transcript accumulation of STS gene in *Lens culinaris* (A) Graph plotted based on intensity of transcript obtained in the RNA blot with background as the reference (B) STS gene expression and 28S and 18S rRNA showing intact RNA observed in the gel used for RNA blot (C) Expression of *LcSTS* by qRT-PCR. Error bars indicate standard error of the mean of four determinations at each time point.

mature seeds (36 DAF activity - 6.2 ± 1.7 pkat/mg). Similar trend in STS activity was also reported in *Pisum sativum* (Peterbauer et al 2001). STS activity also corresponded well with *LcSTS* transcript accumulation, as maximum gene expression and enzyme activity was observed at 26 DAF.

STS is a multifunctional enzyme with wide range of substrate specificities and this characteristic is also species dependant. In *Phaseolus vulgaris* STS catalyzed synthesis from raffinose and galactinol but did not synthesize verbascose (Tanner and Kandler 1968). STS from *Vicia faba* catalyzed synthesis of stachyose and verbascose (Tanner et al 1967). STS from *Vigna angularis* seeds catalyzed synthesis of stachyose and also galactosyl ononitol by transfer of a galactosyl residue from galactinol to D-ononitol (Peterbauer and Richter 1998; Peterbauer et al 1999). Previous reports suggest lentil STS is capable of utilizing galactinol as well as galactosyl ononitol, galactopinitolA as galactosyl donors for stachyose synthesis (Hoch et al 1999). In addition to synthesis of stachyose lentil STS also catalyzed synthesis of many galactosyl cyclitols which include galactopinitolA, galactopinitolB, ciceritol and fagopyritolB1 (Hoch et al 1999). Concentration / presence of galactosyl cyclitols were not analyzed in this study due to unavailability of standards as the method employed in this study showed no separation of cyclitol compounds (Gangola et al 2014).

6.4.4 Verbascose synthase enzyme activity

Previous reports suggest inability of lentil STS to synthesize verbascose (Hoch et al 1999). In this study VS activity was detected in developing lentil seeds. VS activity in lentil was obtained using desalted crude extract with stachyose and galactinol as substrates (Figure 6.5). No VS activity was observed at early stages of seed development at 14, 16, 18 and 20 DAF. An increase in VS activity occurred from 22 DAF with maximum activity observed at 26 DAF (11 ± 1.5 pkat/mg protein). VS activity decreased towards later stages of seed development and no activity was detected at 34 DAF. Highest enzymatic activity for both STS and VS was observed at same stage (26 DAF) of seed development. Similar observation was reported in developing seeds of *Pisum sativum*, where maximum activity for both STS and VS was observed at 30 DAF (Peterbauer et al 2001b).

6.4.5 Galactan: Galactan galactosyl transferase enzyme activity

Activity of galactinol independent enzyme, galactan:galactan galactosyl transferase (GGT) was also detected in lentil during seed development with raffinose as the substrate (Figure 6.5 C). No GGT activity was observed during early stages (14, 16 and 18 DAF of seed development) of seed development. A slight increase in GGT activity was observed at 20, 22 and 24 DAF, with a rapid increase from 24 to 26 DAF and further gradual increase to reach the maximum (21.3 pkat/mg protein) at 32 DAF. A reduction in enzyme activity was observed towards later stages of development, however enzyme activity was detected in mature seeds (34 DAF).

To isolate a gene coding for GGT in lentil, gene sequence for *Ajuga reptans* GGT was used to screen a lentil cDNA library from developing lentil seeds. However, no positive signals were observed. No gene coding for GGT in lentil could be isolated.

6.4.4 Accumulation of stachyose and verbascose in developing seeds

Lentil seeds accumulate RFO during the later part of seed development. Stachyose concentration started to increase at 20 DAF and gradually increased to two percent of the dry flour at 32 DAF. However, it decreased to 1.5 % in the mature seed at 38 DAF (Figure 6.6 A). Increase in verbascose concentration was observed at 24 DAF and the highest concentration of 0.8% of flour dry weight was observed at 32 DAF. The verbascose concentration also decreased to 0.6% in mature seeds (Figure 6.6 B). Both the raffinose and stachyose concentrations are lower than reported earlier from field grown material (Tahir et al 2011). The plant material analyzed in this study was grown in growth chambers. A similar observation of reduction in RFO concentration in greenhouse as compared to field grown material has been reported in chickpea (Gangola et al 2013).

Maximum STS activity was obtained (at seed stages prior to maximum stachyose accumulation) at 26 DAF where rapid accumulation of stachyose was observed during seed development. Similarly maximum VS activity was observed prior to maximum verbascose accumulation. Maximum VS activity was observed at 26 DAF where rapid increase in verbascose concentration was observed. GGT activity compared to oligosaccharide accumulation where maximum accumulation of oligosaccharides was observed at 32 DAF Accumulation of verbascose correlated with stachyose (substrate) accumulation. A delay in verbascose accumulation was observed (24 DAF) when compared to stachyose accumulation at 20 DAF.

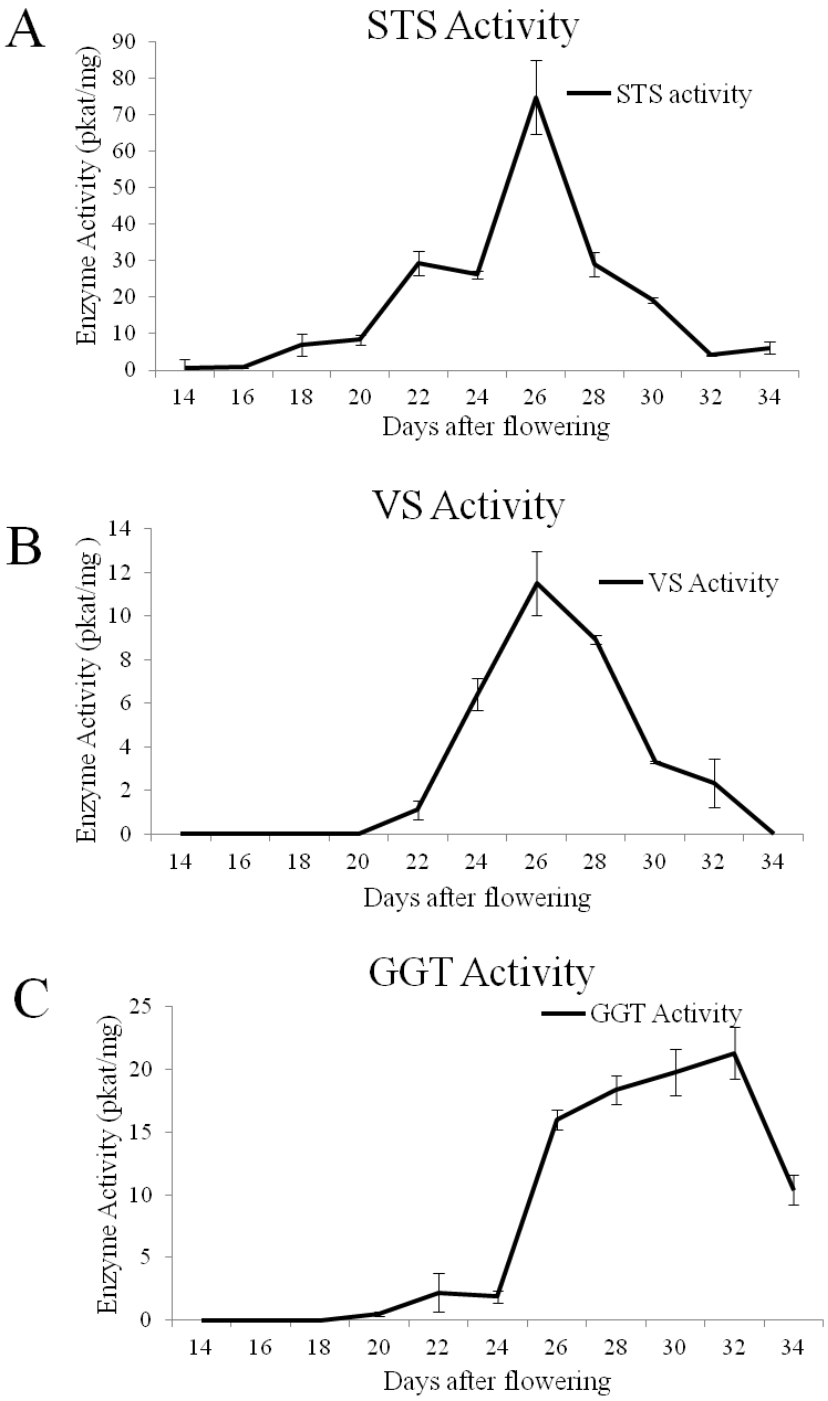


Figure 6.5: Enzyme activity during seed development (A) STS activity with raffinose and galactinol as the substrates (B) VS activity with stachyose and galactinol as the substrates (C) Galactinol independent activity obtained with raffinose as the substrate. Error bars indicate standard error of the mean for six determinations at each time point.

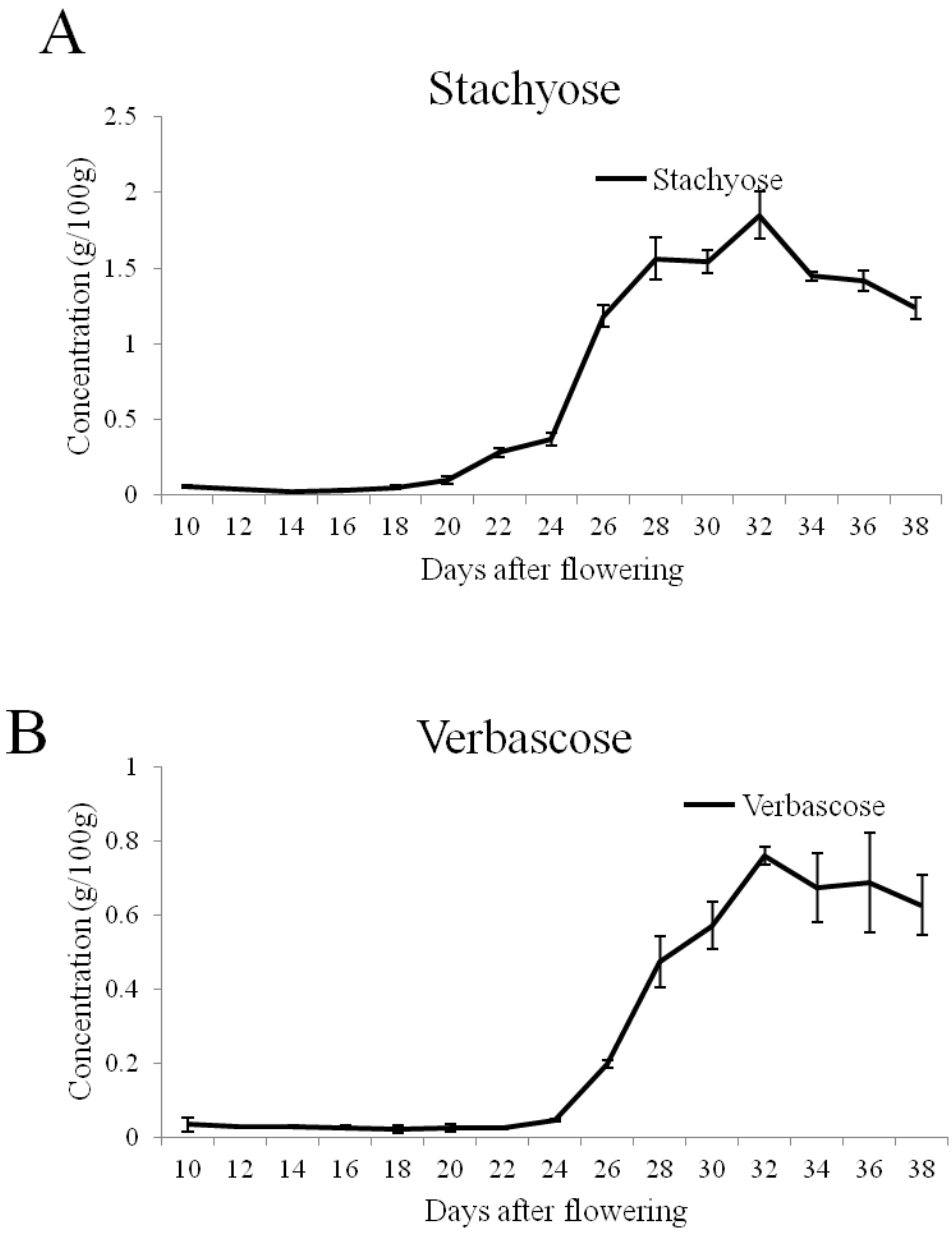


Figure 6.6: Accumulation (g/100g) of (A) stachyose and (B) verbascose during lentil seed development. Error bars indicate standard error of the mean for six determinations at each time point.

Similar RFO accumulation patterns were also observed in *Pisum sativum*, where rapid increase in accumulation was observed from 23- 26 DAF (Peterbauer et al 2001b).

In conclusion, a gene coding for *LcSTS* of length 2693 bp coding for 853 amino acids residues was obtained. Expression of *LcSTS* isolated from *Lens culinaris* correlated with STS activity where maximum gene expression and activity was observed at 26 DAF. Verbascose accumulation was initiated after increase in substrate stachyose. This trend was also observed in activity where VS activity was seen to increase from 20 DAF and STS activity increased from 16 DAF (Figure 6.5). Though purification of STS has been previously reported (Hoch et al 1999), this is first study to report *LcSTS* gene expression, STS activity, VS activity and GGT activity in *Lens culinaris* Medik. during seed development.

7. GENERAL DISCUSSION AND CONCLUSIONS

7.1 Inferences

Lentil (*Lens culinaris* Medik.) is one of the founder crops and its cultivation dates back to ancient times associated with domestication of einkorn, emmer and hulled barley (Zohary and Hopf 1973). Pulse crops in general have gained a great significance in crop rotation, due to their nitrogen fixing capability that enriches soil with nitrogen. Pulses also have superior nutritional quality as they have a higher amount of protein compared to cereal grains, but are also enriched in the essential amino acid lysine that is deficient in cereal grains. Pulse seeds are also rich in dietary fiber and starch that is higher in amylose concentration compared to cereal grains, which may provide a better source of slow digestible or resistant starch (Chibbar et al 2010). A major limitation to increase human consumption of pulses is higher concentration of raffinose family oligosaccharides (RFO) that remain undigested in the small intestine and can be fermented in the large intestine by its microbiota. This process results in stomach discomfort, flatulence and in extreme cases diarrhea. In animals, it also causes low absorption of energy from the feed. Therefore, there is a need to reduce the RFO concentration in pulse seeds so the complete potential of pulse seeds such as lentils can be realized for human diet and animal feed.

The hypothesis of the present thesis was that RFO biosynthetic enzymes gene expression and consequent enzyme activity influence RFO concentration in lentil seeds. Complete cDNA clones were characterized for four genes associated with RFO biosynthesis in lentil seeds. During lentil seed development, the accumulation of RFO constituents and their substrates corresponded to RFO biosynthetic genes expression and corresponding enzyme activities. These results supported the hypothesis postulated in this thesis. The major findings during the course of this research are summarized in the following paragraphs.

Optimization of RNA isolation procedure and reference gene validation for qPCR analysis: Crucial for gene expression analyses

Transcriptome studies to improve seed quality require RNA of high purity. With small lentil seed size, it is not only essential to extract RNA of high purity, but also to obtain adequate quantities of RNA. During seed development, starch, proteins, oligosaccharides and phenolic compounds accumulate and these compounds interfere with RNA isolation. The phenol guanidine method that has been optimized for efficient RNA isolation from developing wheat

seeds (Ganeshan et al 2010) failed to yield adequate amount of RNA in developing lentil seeds. Commercially available Trizol and lysis buffer based methods also provided low yields and co-purified starch and oligosaccharides. A CTAB-LiCl based procedure provided RNA of high yield and purity. The RNA obtained using this procedure was utilized for cDNA library construction and was also used to study transcript accumulation by RNA gel blot analysis and quantitative real time PCR analysis (Chapters 4, 5, 6)

Identification of reference genes for use in qPCR determination of gene expression studies

Stable reference genes are crucial to accurately interpret qRT-PCR results and draw valid conclusions. The nature of the samples and experimental treatments can alter the transcript level of housekeeping genes, thus influencing the relative expression results obtained with genes of interest. In this thesis (Chapter 4), eight reference genes were tested and the most stable reference genes during seed development were identified as EF1α (Elongation factor 1-alpha), GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) and HSP (Heat shock protein). The importance of reference gene validation for 'a sample set' is evident with results obtained for stable reference genes during circadian rhythm or in various tissues tested (Appendices 4.10, 4.11 and 4.12). The most stable reference genes obtained for leaf samples during circadian rhythm were ATUB (α-Tubulin), BTUB (β-Tubulin) and ARF (Adipose ribosylation factor); whereas stable reference genes obtained across various tissues (leaf, roots and stem) were EF1α, HSP and GAPD. Similar results were obtained in chickpea where EF1α and HSP90 were reported as the most stable reference genes during seed development (Garg et al 2010).

Sequence alignment and phylogenetic analysis reveals functional role of RFO biosynthetic genes.

Full length coding sequences of RFO biosynthetic genes GS, RS and STS were isolated by screening a lentil cDNA library. *LcGolS1* and *LcGolS2* of sequence lengths 1336 and 1324 bp, respectively were isolated (Chapter 4). The open reading frames of *LcGolS1* and *LcGolS2* genes were 1002 bp and 975 bp with predicted molecular weights of 38.8 and 37.6 kDa, respectively. A *LcRS* gene coding coding for 792 amino acids with a predicted molecular weight of 88.2 kDa was isolated (Chapter 5). In Chapter 6, a 2669 bp *LcSTS* gene coding for 853 amino acids with a predicted molecular weight of 96 kDa was isolated.

All isolated RFO biosynthetic genes showed high similarity to previously annotated genes in the database. Both *LcGolS* isoforms revealed high sequence identity to other GS sequences in the GenBank database. Sequence identity between *LcGolS1* and *LcGolS2* was 76% at the nucleotide level. Translated *LcGolS1* and *LcGolS2* revealed 73% similarity between the polypeptide sequences. Phylogenetic analysis of *LcGolS1* and *LcGolS2* clustered them in two different clades, with *LcGolS1* showing high similarity to *P. sativum GolS* and *LcGolS2* showing high similarity to *G.max GolS* and *B. hygrometrica GolS*. The *P. sativum GolS* was identified in developing seed samples and contributed to galactinol accumulation in developing seeds (Peterbauer et al 2001). The *B. hygrometrica GolS* was expressed during drought stress suggesting its role in abiotic stress (Wang et al 2009). These results suggest differential functional roles for the two GS isoforms, where *LcGolS1* might contribute to galactinol accumulation in the seeds and *LcGolS2* might contribute to abiotic stress tolerance in lentil.

Gene sequences isolated for *LcRS* and *LcSTS* also revealed high sequence similarity to *P. sativum RS* and *P. sativum STS*. Genes with high sequence similarity might perform similar functions. This was evident in *A. thaliana,* where the presence of more than six genes of putative RS were reported (Peters et al 2010). *AtRS5* showed maximum sequence similarity to *P. sativum RS.* The *AtRS5* was characterized later and shown to catalyze raffinose synthesis (Egbert et al 2013). A similar observation was made in *G. max*, where *GmRS2* (gene later characterized to code for RS) was reported to show high sequence similarity to *PsRS2* when compared to other putative RS genes in *G. max* (Dierking and Bilyeu 2008; Skoneczka et al 2009).

Differences in 3' UTR region in GS genes might indicate regulation at post transcriptional level

A total of 1.8 x 10^5 phage particles were screened to isolate galactinol synthase sequences. Sequencing of phagemids showed differences in 5' and 3'UTR lengths for *LcGolS1* and *LcGolS2*. Nucleotide results for *LGolS1* and *LcGolS2* transcripts showed differences in 3' UTR lengths (Fig 4.2). The 3'UTR lengths for *LcGolS1* were 10 bp, 203 bp and 279 bp and 3' UTR differences for LcGolS2 were 182 bp, 211 bp and 240 bp. Though differences in 3'UTR lengths were observed, the sequences were identical and no nucleotide changes were observed. The differential polyadenylation of these transcripts suggest possible post transcriptional regulation. Transcripts with shorter 3'UTR reduced chances of transcriptional regulation which

includes miRNA binding, whereas longer sequences might promote chances of post transcriptional regulation (Wu et al 2011).

Correlation of gene expression and enzyme activity to oligosaccharide accumulation in developing lentil seeds

Maximum gene expression of *LcGolS1* and *LcGolS2* was found at 32 DAF and 24 DAF respectively. GS enzyme activity correlated with galactinol accumulation, where maximum activity was obtained at 26 and 28 DAF and maximum galactinol accumulation was obtained at 26 DAF. Concentration of its precursors, *Myo*-inositol and glucose/galactose concentrations were detected as early as 10 DAF. A correlation was obtained between glucose/galactose concentrations and galactinol accumulation in developing lentil seeds.

Expression of *LcRS* was found to be highest between 24 to 26 DAF during lentil seed development. Transcript accumulation of *LcRS* correlated with RS enzyme activity where maximum activity was observed at 24 DAF. Rapid accumulation of raffinose was observed between 22 to 24 DAF correlating with *LcRS* expression and enzyme activity. Maximum accumulation of raffinose was found at 32 DAF. Sucrose concentration was found to be maximum at 16 DAF decreased gradually and reached a plateau at 32 DAF, indicating the utilization of substrate for raffinose synthesis. These results suggest that RS gene *LcRS* isolated in this study contributes to raffinose accumulation in lentil seeds.

Expression of *LcSTS* obtained by qRT-PCR and RNA gel blots indicated maximum expression at 24 and 26 DAF respectively. Differences in gene expression results obtained through qRT-PCR and RNA gel blot analysis were attributed to biological replicates used for analysis. Two different sample batches were used in the study, where RNA gel blot, enzyme activity and oligosaccharide accumulation analyses were done using samples collected from the same batch. A correlation was obtained between *LcSTS* expression and STS enzyme where maximum activity was obtained at 26 DAF. No gene for verbascose was isolated; however verbascose synthase (VS) activity was detected in developing lentil seeds. Maximum VS activity was observed at 26 DAF, similar to stachyose synthase activity. In *P. sativum* lines with high verbascose concentrations STS was reported to catalyze stachyose and verbascose synthesis (Peterbauer et al 2002a; Peterbauer et al 2002b, Peterbauer et al 2003). The maximum concentration of verbascose was observed at 32 DAF, similar to stachyose accumulation.

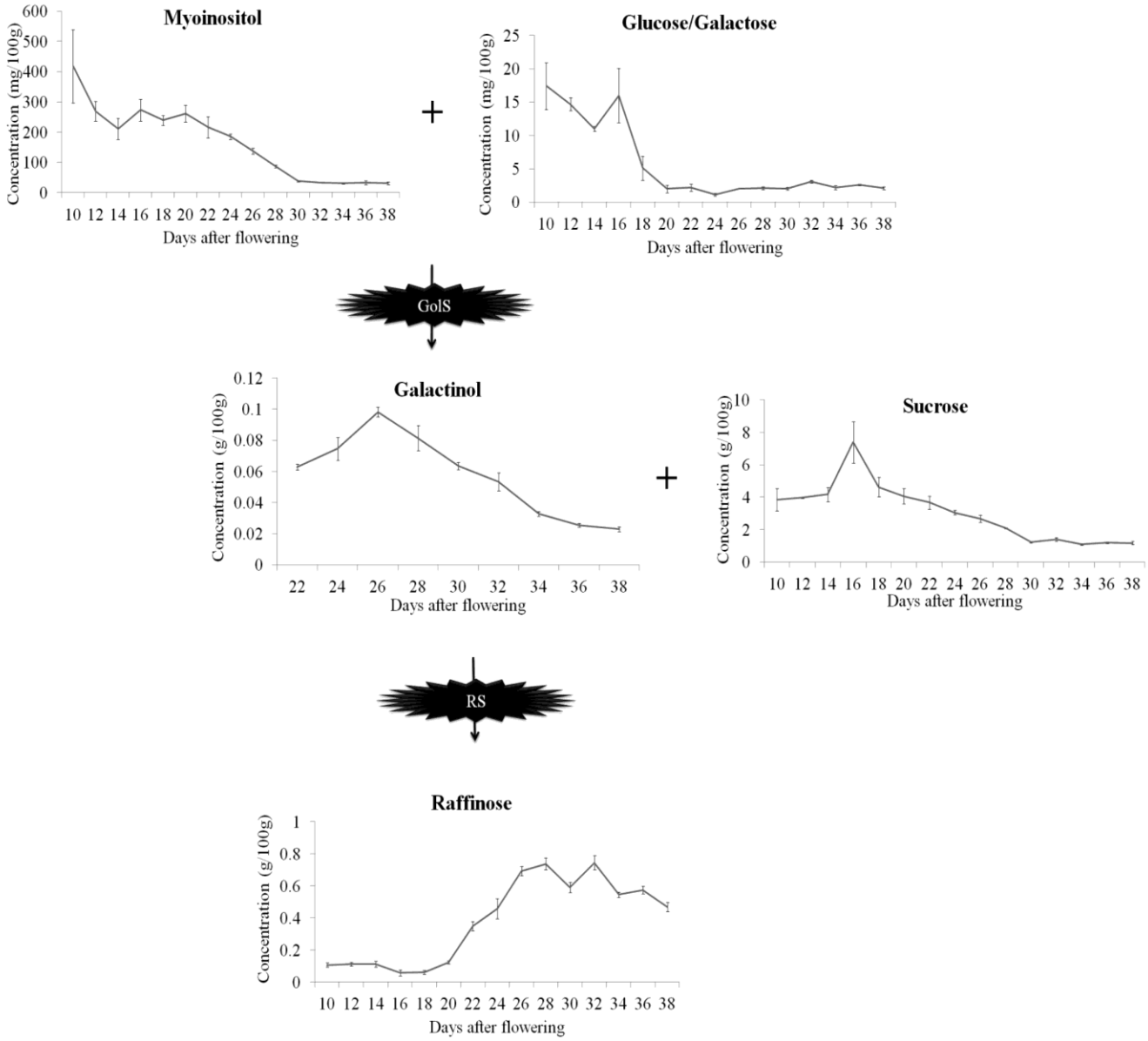


Figure 7.1: Accumulation of raffinose and its precursors (Myo-inositol, glucose / galactose, galactinol and sucrose) during lentil seed development. GolS and RS is galactinol synthase and raffinose synthase respectively (Error bars indicate standard error of the mean of six determinations at each time point).

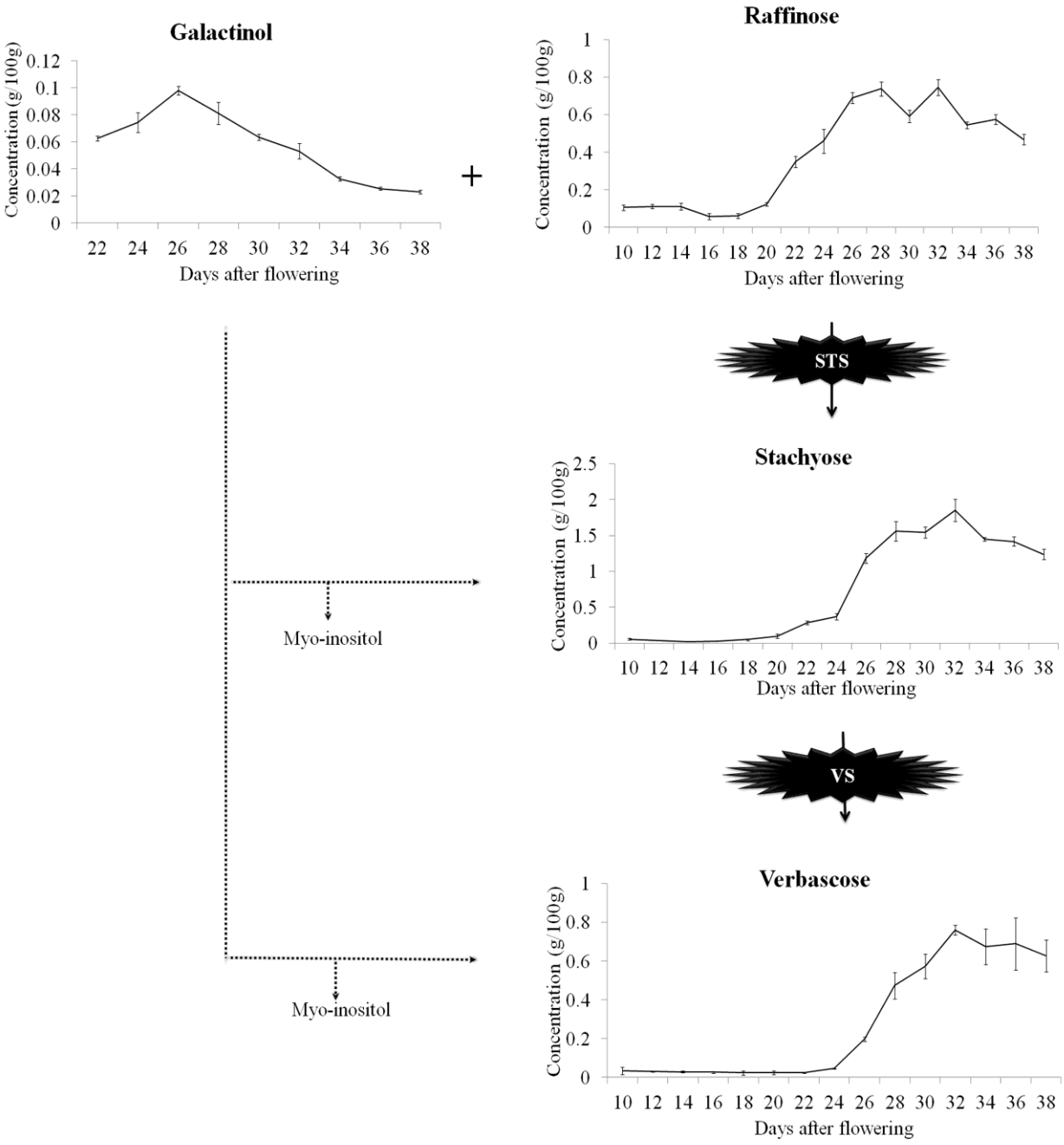


Figure 7.2: Accumulation of stachyose and verbascose during seed development.STS and VS is stachyose synthase and verbascose synthase respectively (Error bars indicate standard error of the mean for six determinations at each time point).

Galactan : galactan galactosyl transferase (GGT activity) was detected at neutral pH during seed development with maximum activity obtained at 26 - 32 DAF. Maximum RFO accumulation was observed 26-32 DAF.

RFO utilized as substrate for higher RFO synthesis

Raffinose and stachyose act as substrates for the synthesis of higher RFO. The concentrations of *myo*-inositol and glucose / galactose started to decrease 20 DAF (Figure 7.1), indicating initiation of substrate utilization in the raffinose biosynthetic pathway. GS utilized substrates and increased its enzyme activity from 20 DAF. As a result of GS activity, galactinol started to accumulate and reached it maximum at 26 DAF. At 24 DAF, which is closer to 26 DAF where galactinol concentration is at its maximum, RS enzyme activity reached its maximum. RS catalyzed raffinose accumulation from 22 DAF with a steady increase in raffinose accumulation that reached its peak at 32 DAF.

Raffinose is utilized as a substrate to synthesize stachyose. STS enzyme activity was at its maximum at 26 DAF, synthesizing stachyose to its maximum at 32 DAF. VS enzyme activity followed a similar trend as STS enzyme activity. In this study, when verbascose concentration reached its maximum at 32 DAF, STS enzyme activity started to decline. Enzyme activity correlated with RFO accumulation where STS and VS activity reached its maximum two days after RS maximum activity.

7.2 Novel Findings

- A CTAB-LiCl based RNA isolation procedure was optimized for high yield and quality in developing lentil (*Lens culinaris* cv CDC Redberry seeds).
- EF1α (Elongation factor 1- alpha), GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) and HSP (Heat shock protein) were found to be the most stable reference genes during seed development for quantitative real time PCR experiments.
- Three RFO biosynthetic genes were characterized: cDNA clones for galactinol synthase (*LcGolS1*, *LcGolS2*), raffinose synthase (*LcRS*) and stachyose synthase (*LcSTS*) were isolated.

The isolated *LcRS* gene is the major contributor to raffinose accumulation in lentil seeds: Maximum expression of *LcRS* and enzyme activity was obtained at 24 DAF. *LcRS* expression correlated with RS activity and raffinose concentration during seed development.

- The isolated *LcSTS* gene is the major contributor to stachyose accumulation in lentil seeds. Gene expression of *LcSTS* correlated with STS activity and stachyose concentration during seed development.
- Accumulation of verbascose correlated with accumulation of its substrate (stachyose): During lentil seed development, STS and VS activity increased at 16 and 20 DAF respectively. This increase in enzyme activity correlated with increase in stachyose and verbascose concentration which was observed in the seeds from 20 and 24 DAF respectively. The trend in verbascose accumulation correlated to its substrate (stachyose) accumulation in developing lentil seeds.
- Activity of enzyme GGT which catalyzes galactinol independent RFO synthesis was also detected in lentil during seed development.

7.3 Future Research

- The cDNA isolated and characterized in this study can be used to produce recombinant RFO biosynthetic enzymes in adequate quantities and purity.
- Detailed enzyme kinetic analyses of the three RFO biosynthetic enzymes to elucidate the biochemical mechanism of RFO biosynthesis in lentils.
- Use the homogenous recombinant proteins to produce antibodies in rabbits. The antibodies can be used to study, existence of monomeric / heteromeric protein conformation, or the existence of any enzyme complexes for RFO biosynthesis.
- Analysis of RFO biosynthetic pathway in lentil genotypes with with contrasting RFO concentrations to identify reactions determining RFO concentration or composition.
- Identification of allelic variation in RFO biosynthetic genes influencing RFO concentration and composition.
- Purify Verbascose synthase and (GGT) and / or genes for these enzymes in lentils.

- Characterize galactosyl cyclitols to understand carbon partitioning between RFO and cyclitols.

The isolation and characterization of cDNA clones for key RFO biosynthetic genes and their expression during seed development has shown that RFO are synthesized at later stages of lentil seed development. The characterized gene sequences can be utilized to study allelic variation in the genus *Lens* with contrasting concentrations of RFO and its constituents (Tahir et al 2012). Allelic variation in RFO biosynthetic genes can be used to develop gene based perfect DNA markers that can be used in marker assisted selection to accelerate the development of lentil cultivars with desired RFO concentration and composition. Alternatively, the cDNA clones can be used to produce recombinant RFO biosynthetic enzymes that can be used to study enzyme characteristics, and / or produce antibodies that can be used to precisely determine the cellular location of RFO biosynthesis and post-translational regulation of enzyme activity, if any. A better understanding of RFO biosynthesis, its biochemical and molecular basis can help to develop practical strategies to reduce RFO in lentils and other pulse crops.

8. Literature Cited

Abdel-Gawad AS (1993) Effect of domestic processing on oligosaccharide content of some dry legume seeds. Food Chemistry 46: 25-31.

Agriculture Agri-Food Canada (2014) Crop profile for lentil in Canada. <http://www.agr.gc.ca/pmc-cropprofiles > Retrieved 20-10-2014.

Allan SM, Hitz WD (2000) Plant raffinose synthase homologs. International Patent Publication WO 00/24915, PCT/US99/24923.

Allen GC, Flores-Vergara MA, Krasnyanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyl trimethyl ammonium bromide. Nature Protocols 1: 2320-2325.

Amuti KS, Pollard CJ (1977) Soluble carbohydrates of dry and developing seeds. Phytochemistry 16: 529-532.

Anderson JW, Major AW (2002) Pulses and lipaemia, short- and long-term effect: Potential in the prevention of cardiovascular disease. British Journal of Nutrition 88: S263-S271.

Aranda R, Dineen SM, Craig RL, Guerrieri RA, Robertson JM (2009) Comparison and evaluation of RNA quantification methods using viral, prokaryotic and eukaryotic RNA over a 104 concentration range. Analytical Biochemistry 387: 122-127.

Aranda P, Dostalova J, Frias J, Lopez-Jurado M, Kozlowska H, Vidal-Valverde C, Zdyunczk Z (2001) Nutrition. In: Carbohydrates in grain legumes: Improving nutritional quality and agronomic characteristics. Eds: Hedley CL. CAB International, Wallingford, UK, pp 61-87.

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research 38: W529-W533.

Avigad G, Dey PM (1997) Carbohydrate metabolism: Storage carbohydrates. In: Plant biochemistry. Eds: Dey PM, Harbourne J. Academic Press, San Diego, USA, pp 143-204.

Bachmann M, Matile P, Keller F (1994) Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L. Cold acclimation, translocation, and sink to source transition: Discovery of a chain elongation enzyme. Plant Physiology 105: 1335-1345.

Bachmann M, Keller F (1995) Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L. Inter- and intracellular compartmentation. Plant Physiology 109: 991-998.

Båga M, Chodaparambil SV, Limin AE, Pecar M, Fowler DB, Chibbar RN (2007) Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. Functional & Integrative Genomics 7: 53-68.

BeMiller JN (2001) Classification, structure and chemistry of polysaccharides of food. In: Handbook of dietary fiber. Eds: Cho SS, Dreher ML. Marcel Dekker, NY, USA, pp 603-611.

Bhatty RS (1988) Composition and quality of lentil (*Lens culinaris* Medik): a review. Canadian Institute of Food Science and Technology Journal 21: 144-160.

Blair MW, Fernandez AC, Ishitani M, Moreta D, Seki M, Ayling S, Shinozaki K (2011) Construction and EST sequencing of full-length, drought stress cDNA libraries for common beans (*Phaseolus vulgaris* L.). BMC Plant Biology 11: 171.

Blöchl A, Peterbauer T, Richter A (2007) Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. Journal of Plant Physiology 164: 1093-1096.

Bock C, Ray H, Georges F (2009) Down-regulation of galactinol synthesis in oilseed *Brassica napus* leads to significant reduction of antinutritional oligosaccharides. Botany 87: 597-603.

Boye J, Zare F, Pletch A (2010) Pulse proteins: Processing, characterization, functional properties and applications in food and feed. Food Research International 43: 414-431.

Bradford (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72: 248-254.

Brenac P, Horbowicz M, Downer SM, Dickerman AM, Smith ME, Obendorf RL (1997) Raffinose accumulation related to desiccation tolerance during maize (*Zea mays* L.) seed development and maturation. Journal of Plant Physiology 150: 481-488.

Breton C, Šnajdrová L, Jeanneau C, Koča, Imberty A (2006) Structures and mechanisms of glycosyltransferases. Glycobiology 16: 29R-37R.

Brown TA (2006) Gene cloning and DNA analysis: An Introduction. Fifth Edition. Blackwell Publishing, UK.

Bryant G, Wolfe J (1992) Can hydration forces induce lateral phase separations in lamellar phases? Cryo-Letters 13: 23-36.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaff MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55: 611-22.

Cai R, McCurdy A, Baik BK (2002) Textural property of 6 legume curds in relation to their protein constituents. Journal of Food Science 67: 1725-1730.

Castillo EM, De Lumen BO, Reyes PS, De Lumen HZ (1990) Raffinose synthase and galactinol synthase in developing seeds and leaves of legumes. Journal of Agricultural and Food Chemistry 38: 351-355.

Cathala G, Savouret JF, Mendez B, West BL, Karin M, Martial JA, Baxter JD (1983) A method for isolation of intact, translationally active ribonucleic acid. DNA 2: 329-335.

Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N (2013) ConSurf: Using evolutionary data to raise testable hypotheses about protein function. Israel Journal of Chemistry 53: 199-206.

Chappell AS, Scaboo AM, Wu X, Nguyen H, Pantalone VR, Bilyeu KD (2006) Characterization of the MIPS gene family in *Glycine max*. Plant Breeding 125: 493-500.

Chibbar RN, Ganeshan S, Baga M, Khandelwal RL (2004) Carbohydrate metabolism. In: Encyclopedia of grain science. Eds: Wrigley CW, Walker CE, Corke H. Elsevier, Amsterdam, Netherlands, pp 168-179.

Chibbar RN, Ambigaipalan P, Hoover R (2010) Molecular diversity in pulse seed starch and complex carbohydrates and its role in human nutrition and health. Cereal Chemistry 87: 342-352.

Chiera JM, Grabau EA (2007) Localization of myo-inositol phosphate synthase (GmMIPS-1) during the early stages of soybean seed development. Journal of Experimental Botany 58: 2261–2268.

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. Analytical Biochemistry 162: 156-159.

Chompreeda PT, Fields ML (1984) Effects of heat and natural fermentation on amino acids, flatus producing compounds, lipid oxidation and trypsin inhibitor in blends of soybean and cornmeal. Journal of Food Science 49: 563-565.

Chiu CP, Watts AG, Lairson LL Gilbert M, Lim D, Wakarchuk WW, Withers SG, Strynadka NC (2004) Structural analysis of the sialyltransferase Cst II from *Campylobacter jejuni* in complex with a substrate analog. Nature Structure and Molecular Biology 11: 163-170.

Claros MG, Canovas FM (1998) Rapid high quality RNA preparation from pine seedlings. Plant Molecular Biology Reports 16: 9-18.

Coon CN, Leske KL, Akavanichan O, Cheng TK (1990) Effect of oligosaccharides-free soybean meal on true metabolizable energy and fibre digestion in adult roosters. Poultry Science 69: 787-793.

Costa GED, Queiroz-Monici KDS, Reis SMPM, de Oliveira AC (2006) Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. Food Chemistry 94: 327-330.

Cristofaro E, Mottu F, Wuhrmann JJ (1973) Study of the effect on flatulence of leguminous seeds oligosaccharides. Nestle Research News 102-104.

Cunningham SM, Nadeaub P, Castonguayb Y, Labergec S, Volenec JJ (2003) Raffinose and stachyose accumulation, galactinol synthase expression, and winter injury of contrasting alfalfa germplasms. Crop Science 43: 562-570.

Dash PK (2013) High quality RNA isolation from ploy phenol-, polysaccharide- and protein-rich tissues of lentil (*Lens culinaris*). 3 Biotech 3: 109-114.

Davis KR (1981) Effect of processing on composition and tetrahymena relative nutritive-value of green and yellow peas, lentils, and white pea beans. Cereal Chemistry 58: 454-460.

de Faria SM, Lewis GP, Sprent JI, Sutherland JM (1989) Occurrence of nodulation in the leguminosae. New Phytologist 111: 607-619.

DeVries JW (2003) On defining dietary fibre. Proceedings of the Nutrition Society 62: 37-43.

Dey P.M. (1985) D-Galactose containing oligosaccharides. In: Biochemistry of storage carbohydrates in green plants. Eds: Dey PM, Dixon R. Academic Press, London, UK pp 53-129.

Die JV, Román B, Nadal S, González-Verdejo CI (2010) Evaluation of candidate reference genes for expression studies in *Pisum sativum* under different experimental conditions. Planta 232: 145-53.

Dierking EC, Bilyeu KD (2008) Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. The Plant Genome 1: 135–145.

Dierking EC, Bilyeu KD (2009) Raffinose and stachyose metabolism are not required for efficient soybean seed germination. Journal of Plant Physiology 166: 1329-1335.

dos Santos TB, Budzinski IG, Marur CJ, Petkowicz CL, Pereira LF, Vieira LG (2011) Expression of three galactinol synthase isoforms in *Coffea arabica* L. and accumulation of raffinose and stachyose in response to abiotic stresses. Plant Physiology and Biochemistry 49: 441-448.

dos Santos R, Vergauwen R, Pacolet P, Lescrinier E, Van den Ende W (2013) Manninotriose is a major carbohydrate in red deadnettle (*Lamium purpureum*, Lamiaceae). Annals of Botany 111: 385-393.

Downie B, Bewley JD (2000) Soluble sugar content of white spruce (*Picea glauca*) seeds during and after germination. Physiologia Plantarum 110: 1-12.

Downie B, Gurusinghe S, Dahal P, Thacker RR, Snyder JC, Nonogaki H, Yim K, Fukanaga K, Alvarado V, Bradford KJ (2003) Expression of a galactinol synthase gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. Plant Physiology 131: 1347-59.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15.

Dry I, Smith A, Edwards A, Bhattacharyya M, Dunn P, Martin C (1992) Characterization of cDNAs encoding 2 isoforms of granule-bound starch synthase which show differential expression in developing storage organs of pea and potato. The Plant Journal 2: 193-202.

Dumont E, Fontaine V, Vuylsteker C, Sellier H, Bodèle S, Voedts N, Devaux R, Frise M, Avia K, Hilbert JL, Bahrman N, Hanocq E, Lejeune-Hénaut I, Delbreil B (2009) Association of sugar content QTL and PQL with physiological traits relevant to frost damage resistance in pea under field and controlled conditions. Theoretical and Applied Genetics 118: 1561-1571.

Duszkiewicz-Reinhard W, Gujska E, Khan K (1994) Reduction of stachyose in legume flours by lactic acid bacteria. Journal of Food Science 59: 115-117.

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnology Journal 8: 2-9.

Egbert A, Keller F and Peters S (2013) Abiotic stress-induced accumulation of raffinose in Arabidopsis leaves is mediated by a single raffinose synthase (RS5, At5g40390). BMC Plant Biology 13: 218.

El-Adawy TA, Rahma EH, El-Bedawey AL, El-Beltagy AE (2003) Nutritional potential and functional properties of germinated mung bean, pea and lentil seeds. Plant Foods for Human Nutrition 58: 1-13.

Ereifej KI, Haddad SG (2000) Chemical composition of selected Jordanian cereals and legumes as compared with the FAO, Moroccan, East Indian and Latin American tables for use in Middle East. Trends in Food Science and Technology 11: 374-378.

Erskine W, Muehlbauer FJ, Short RW (1990) Stages of development in lentil. Experimental Agriculture 26: 297-302.

FAO/WHO (1998) Carbohydrates in human nutrition, Report of a joint FAO/WHO expert consultation, FAO Food and Nutrition Paper 66 Rome: FAO.

FAOSTAT: FAO Statistical Database <http://faostat.fao.org/default.aspx?lang=en> Retrieved on 20-10-2014.

Fasina O, Tyler B, Pickard M, Zheng GH, Wang N (2001) Effect of infrared heating on the properties of legume seeds. International Journal of Food Science and Technology 36: 79-90.

Ford R, Mustafa B, Inder P, Shaikh R, Materne M, Taylor P (2007) Lensomics: Advances in genomics and molecular techniques for lentil breeding and management. In: Lentil: An ancient crop for modern times. Eds: Yadav SS, McNeil DL, Stevenson PC. Springer Netherlands, pp 275-289.

Frias J, Vidal-Valverde C, Bakhsh A, Arthur AE, Hedley C (1994) An assessment of variation for nutritional and non-nutritional carbohydrates in lentil seeds *(Lens culinaris)*. Plant Breeding. 113: 170-173.

Frias J, Prodanov M, Sierra I, Vidal-Valverde C (1995) Effect of light on carbohydrates and hydrosoluble vitamins of lentils during soaking. Journal of Food Protection 58: 692-695.

Frias J, Diaz-Pollan C, Hedley CL, Vidal-Valverde C (1996) Evolution and kinetics of monosaccharide, disaccharide and α-galactosides during germination of lentils. Zeitschrift für Lebensmittel-Untersuchung und Forschung 202: 35-39.

Frias J, Bakhsh A, Jones DA, Arthur AE, Vidal-Valverde C, Rhodes MJC, Hedley CL (1999) Genetic analysis of the raffinose oligosaccharide pathway in lentil seed. Journal of Experimental Botany 50: 469-476.

Frias J, Vidal-Valverde C, Sotomayor C, Diaz-Pollan C, Urbano G (2000) Influence of processing on available carbohydrate content and antinutritional factors of chickpeas. European Food Research and Technology 210: 340-345.

Frias J, Doblado R, Vidal-Valverde C (2003) Kinetics of soluble carbohydrates by action of endo / exo α-galactosidase enzyme in lentils and peas. European Food Research and Technology 216: 199-203.

Ganeshan S, Vitamvas P, Fowler DB, Chibbar RN (2008) Quantitative expression analysis of selected COR genes reveals their differential expression in leaf and crown tissues of wheat (*Triticum aestivum* L.) during an extended low temperature acclimation regimen. Journal of Experimental Botany 59: 2393-2402.

Ganeshan S, Drinkwater JM, Repellin A, Chibbar RN (2010) Selected carbohydrate metabolism genes show coincident expression peaks in grains of in vitro-cultured immature spikes of wheat (*Triticum aestivum* L.). Journal of Agricultural Food Chemistry 58: 4193-4201.

Ganeshan S, Leis M, Drinkwater JM, Madsen LT, Jain JC, Chibbar RN (2012) In vitro-cultured wheat spikes provide a simplified alternative for studies of cadmium uptake in developing grains. Journal of the Science of Food and Agriculture 92: 1740-1747.

Gangola MP, Khedikar YP, Gaur PM, Båga M, Chibbar RN (2013) Genotype and growing environment interaction shows a positive correlation between substrates of raffinose family oligosaccharides (RFO) biosynthesis and their accumulation in chickpea (*Cicer arietinum* L.) seeds. Journal of Agricultural and Food Chemistry 6: 4943-4952.

Gangola MP, Jaiswal S, Khedikar YP, Chibbar RN (2014) A reliable and rapid method for soluble sugars and RFO analysis in chickpea using HPAEC-PAD and its comparison with HPLC-RI. Food Chemistry 154: 127-33.

Garg R, Sahoo A, Tyagi AK, Jain M (2010) Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). Biochemical and Biophysical Research Communications 396: 283-288.

Gibbons BJ, Roach PJ, Hurley TD (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. Journal of Molecular Biology 319: 463-477.

Gil L, Ben-Ari J, Turgeon R, Wolf S (2012) Effect of CMV infection and high temperatures on the enzymes involved in raffinose family oligosaccharide biosynthesis in melon plants. Journal of Plant Physiology 169: 965-970.

Graham PH, Vance CP (2000) Nitrogen fixation in perspective: an overview of research and extension need. Field Crops Research 65: 93-106.

Graham PH, Vance CP (2003) Importance and constraints to greater use. Plant Physiology 131: 872-877.

Granito M, Champ M, Guerra M, Frias J (2003) Effect of natural and controlled fermentation on flatus-producing compounds of beans (*P. vulgaris*). Journal of the Science of Food and Agriculture 83: 1004-1009.

Granito M, Torres A, Frías J, Guerra M, Vidal-Valverde C (2005) Influence of fermentation on the nutritional value of two varieties of *Vigna sinensis*. European Food Research Technology 220: 176-181.

Griga M, Kosturkova G, Kuchuk N, Ilieva-Stoilova M (2001) Biotechnology. In: Carbohydrates in grain legume Seeds. Improving nutritional quality and agronomic characteristics. Eds: Hedley CL. CAB International, Wallingford, UK, pp 145-207.

Grusak MA (2009) Nutritional and health-beneficial quality. In: The lentil: Botany, production and uses. Eds: Erskine W, Muehlbauer FJ, Sarker A, Sharma B. CAB International, Wallingford, UK, pp 368-390.

Guillon F, Champ MM J (2002) Carbohydrate fractions of legumes: uses in human nutrition and potential for health. British Journal of Nutrition 88: S293-S306.

Gulewicz P, Szymaniec S, Bubak B, Frias J, Vidal-Valverde C, Trojanowska K, Gulewicz K (2002) Biological activity of α-galactoside preparations from *Lupinus angustifolius* L. and *Pisum sativum* L. seeds. Journal of Agricultural and Food Chemistry 50: 384-389.

Gulewicz P, Martínez-Villaluenga C, Kasprowicz-Potocka M, Frias J (2014) Non-nutritive compounds in Fabaceae family seeds and the improvement of their nutritional quality by traditional processing – a review. Polish Journal of Food and Nutrition Sciences 64: 75-89.

Haab CI, Keller F (2002) Purification and characterization of the raffinose oligosaccharide chain elongation enzyme, galactan:galactan galactosytransferase (GGT), from *Ajuga reptans* leaves. Physiologia Plantarum 114: 361-371.

Hannah M A, Zuther E, Buchel K, Heyer AG (2006) Transport and metabolism of raffinose family oligosaccharides in transgenic potato. Journal of Experimental Botany 57: 3801-3811.

Hao JH, Yang R, Fang KF, Wang JL, Zhang Q, Shen YY, Li TL (2014) Low night temperatures inhibit galactinol synthase gene expression and phloem loading in melon leaves during fruit development. Russian Journal of Plant Physiology 61: 178-187.

Haritatos E, Ayre BG, Turgeon R (2000) Identification of phloem involved in assimilate loading in leaves by the activity of the galactinol synthase promoter. Plant Physiology 123: 929-938

Hayakawa K, Mizutani J, Wada K, Masai T, Yoshiara I, Mitsuoka T (1990) Effects of soybean oligosaccharides on human fecal microflora. Microbial Ecology in Health and Disease 3: 293-303.

Hedley CL (2001) Carbohydrates in grain legume seeds. Improving nutritional quality and agronomic characteristics. CAB International, Wallingford, UK.

Hegeman CE, Good LL, Grabau EA (2001) Expression of D-myo-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. Plant Physiology 125: 1941-1948.

Hitz WD, Carlson TJ, Kerr PS, Sebastian SA (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiology 128: 650-660.

Hoagland DR (1918) The relation of the plant to the reaction of the nutrient solution. Science 48: 422-425.

Hoch G, Peterbauer T, Richter A (1999) Purification and characterization of stachyose synthase from lentil (*Lens culinaris*) seeds: galactopinitol and stachyose synthesis. Archives of Biochemistry and Biophysics 366: 75-81.

Holthaus U, Schmitz K (1991) Distribution and immunolocalization of stachyose synthase in *Cucumis melo* L. Planta 185: 479-486.

Horbowicz M, Obendorf RL (1994) Seed desiccation tolerance and storability: Dependence on flatulence-producing oligosaccharides and cyclitols - review and survey. Seed Science Research 4: 385-405

Hu R, Fan C, Li H, Zhang Q, Fu YF (2009) Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. BMC Molecular Biology 10: 93.

Huber JLA, Pharr DM, Huber SC (1990) Partial purification and characterization of stachyose synthase in leaves of *Cucumis sativus* and *Cucumis melo*: utilization of a rapid assay for myoinositol. Plant Science 69: 179-188.

Hunt AG (2012) RNA regulatory elements and polyadenylation in plants. Frontiers in Plant Science. 2:109.

Ibrahim SS, Habiba RA, Shatta AA, Embaby HE (2002) Effect of soaking, germination, cooking and fermentation on antinutritional factors in cowpeas. Nahrung 46: 92-95.

Iqbal A, Khalil IA, Ateeq N, Khan MS (2006) Nutritional quality of important food legumes. Food Chemistry 97: 331-335.

Jenkins DJ, Wolever TM, Taylor RH, Barker H, Fielden H, Baldwin JM, Bowling AC, Newman HC, Jenkins AL, Goff DV (1981) Glycemic index of foods: a physiological basis for carbohydrate exchange. American Journal of Clinical Nutrition 34: 362-366.

Jones DA, DuPont MS, Ambrose MJ, Frias J, Hedley CL (1999) The discovery of compositional variation for the raffinose family of oligosaccharides in pea seeds. Seed Science Research 9: 305-310.

Jood S, Mehta U, Singh R, Bhat CM (1985) Effect of processing on flatus-producing factors in legumes. Journal of Agricultural Food Chemistry 33: 268-271.

Jood S, Bishnoi S, Sharma A (1998) Chemical analysis and physico-chemical properties of chickpea and lentil cultivars. Nahrung 42: 71-74.

Jorge AR-P, Omar Z (2011) Isolation of total RNA from tissues rich in polyphenols and polysaccharides of mangrove plants. Electronic Journal of Biotechnology 14.

Kannan U, Ganeshan S, Chibbar RN (2014) A hexadecyltrimethylammonium bromide (CTAB) based protocol to isolate high quality RNA in adequate quantities for gene expression analyses in developing seeds of lentils (*Lens culinaris* Medik.). Gene Expression to Genetical Genomics 7: 7-16.

Karner U, Peterbauer T, Raboy V, Jones DA, Hedley CL, Richter A (2004) *Myo*-inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds. Journal of Experimental Botany 55: 1981-1987.

Kaur S, Cogan NO, Pembleton LW, Shinozuka M, Savin KW, Materne M, Forster JW (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genomics 12: 265.

Keller F, Matile P (1985) The role of the vacuole in storage and mobilization of stachyose in tubers of *Stachys sieboldii*. Journal of Plant Physiology 119: 369-380.

Keller F, Pharr DM (1996) Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides. In: Photoassimilate distribution in plants and crops. Eds: Zamski E, Schaffer AA. Marcel Dekker, New York, USA, pp 115-184.

Klotke J, Kopka J, Gatzke N, Heyer AG (2004) Impact of soluble sugar concentrations on the acquisition of freezing tolerance in accessions of *Arabidopsis thaliana* with contrasting

cold adaptation – evidence for a role of raffinose in cold acclimation. Plant, Cell & Environment 27: 1395-1404.
Koster KL (1991) Glass formation and dessication tolerance in seeds. Plant Physiology 96: 302-304.
Koplik R, Borkova M, Mestek O, Komınkova J, Suchanek M (2002) Application of size-exclusion chromatography – inductively coupled plasma mass spectrometry for fractionation of element species in seeds of legumes. Journal of Chromatography B 775: 179-187.
Knaupp M, Mishra KB, Nedbal L, Heyer AG (2011) Evidence for a role of raffinose in stabilizing photosystem II during freeze–thaw cycles. Planta 234: 477-486.
Kuo TM, Lowell CA, Smith PT (1997) Changes in soluble carbohydrates and enzymic activities in maturing soybean seed tissues. Plant Science 125: 1-11.
Kylen AM, McCready RM (1975) Nutrition in seeds and sprouts of alfalfa, lentils, mung beans and soybeans. Journal of Food Science 40: 1008-1009.
Laudencia-Chingcuanco DL, Stamova BS, You FM, Lazo GR, Beckles DM, Anderson OD (2007) Transcriptional profiling of wheat caryopsis development using cDNA microarrays. Plant Molecular Biology 63: 651-668.
Lan T, Yao B, Shen Y, Wang XA (2013) Isolation of high-quality total RNA from lipid rich seeds. Analytical Biochemistry 438: 11-13.
Leopold AC, Sun WQ, Bernal-Lugo I (1994) The glassy state in seeds: analysis and function. Seed Science Research 4: 267-274.
Leterme P (2002) Recommendations by health organizations for pulse consumption. British Journal of Nutrition 88: S239-S242.
Lehle L, Tanner W (1973) The function of myo-inositol in the biosynthesis of raffinose. Purification and characterization of galactinol:sucrose 6-galactosyltransferase from *Vicia faba* seeds. European Journal of Biochemistry 38: 103-10.
Li S, Kim WD, Kaneko S, Prema PA, Nakajima M, Kobayashi H (2007) Expression of rice (*Oryza sativa* L. var. Nipponbare) alpha-galactosidase genes in *Escherichia coli* and characterization. Bioscience, Biotechnology and Biochemistry 71: 520–526.
Li ZW, Trick HN (2005) Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. BioTechniques 38: 872-876.
Lin TP, Huang NH (1994) The relationship between carbohydrate composition of some tree seeds and their longevity. Journal of Experimental Botany 45: 1289-1294.
Liu JJ, Odegard W, Lumen BO (1995) Galactinol synthase from kidney bean cotyledon and zucchini leaf (Purification and N-terminal sequences). Plant Physiology 109: 505-511.
Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2–2^{-\Delta\Delta Ct}$ method. Methods 25: 402-40.
Lowell CA, Kuo TM (1989) Oligosaccharide metabolism and accumulation in developing soybean seeds. Crop Science 29: 459-465.
Ludt C, Kindl H (1990) Characterization of a cDNA encoding *Lens culinaris* glycolate oxidase and developmental expression of glycolate oxidase mRNA in cotyledons and leaves. Plant Physiology 94: 1193-1198.
Machaiah JP, Pednekar MD, Thomas P (1999) Reduction in flatulence factors in mung beans (*Vigna radiata*) using low-dose $^{\gamma}$-irradiation. Journal of the Science of Food and Agriculture 79: 648-652.

Macrae E (2007) Extraction of plant RNA. In: Protocols for nucleic acid analysis by non radioactive probes. Eds: Hilario E, Mackay JF. Humana Press Inc, Totowa, USA, pp 15-24.

Manoli A, Sturaro A, Trevisan S, Quaggiotti S, Nonis A (2012) Evaluation of candidate reference genes for qPCR in maize. Journal of Plant Physiology 169: 807- 815.

Mansour EH, Khalil AH (1998) Reduction of raffinose oligosaccharides in chickpea (*Cicer arietinum*) flour by crude extracellular fungal α-galactosidase. Journal of the Science of Food and Agriculture 78: 175-181.

Martínez-Villaluenga C, Frias J, Vidal-Valverde C (2005) Raffinose family oligosaccharides and sucrose contents in 13 Spanish lupin cultivars. Food Chemistry 91: 645-649.

Martínez-Villaluenga C, Frias, J, Vidal-Valverde C (2008a) Alpha-galactosides: antinutritional factors or functional ingredients? Critical Reviews in Food Science and Nutrition 48: 301–316.

Martínez-Villaluenga C, Chicholoska J, Kliber A, Gulewicz K (2008b) Raffinose family oligosaccharides of lupin (*Lupinus albus* L. cv Multolupa) as a potential prebiotic. Proceedings of the Nutrition Society 67: E55.

Masai T, Wada K, Hayakawa K (1987) Effects of soybean oligosaccharides on human intestinal flora and metabolic activities. Japanese Journal of Bacteriology 42: 313.

McCutchen CM, Duffaud GD, Leduc P, Petersen ARH, Tayal A, Khan SA, Kelly RM (1996). Characterization of extremely thermostable enzymatic breakers (alpha-1,6-galactosidase and beta-1,4-mannanase) from the hyperthermophilic bacterium *Thermotoga neapolitana* 5068 for hydrolysis of guar gum. Biotechnology and Bioengeering 52: 332-339.

McIntosh S, Watson L, Bundock P, Crawford A, White J, Cordeiro G, Barbary, Rooke L, Henry R (2007) SAGE of the developing wheat caryopsis. Plant Biotechnology Journal 5: 69-83.

McPhee KE, Zemetra RS, Brown J, Myers JR (2002) Genetic analysis of the raffinose family oligosaccharides in common bean. Journal of the American Society for Horticultural Science 127: 376-382.

Meis SJ, Fehr WR, Schnebly SR (2003) Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. Crop Science 43: 1336-1339.

Muehlbauer FJ, Cubero JI, Summerfield RJ (1985) Lentil (*Lens culinaris* Medik.). In: Grain legume crops. Eds: Summerfield RJ, Roberts EF. Collins, UK, pp 266-311.

Muehlbauer FJ, Cho S, Sarker A, McPhee KE, Coyne CJ, Rajesh PN, Ford R (2006) Application of biotechnology in breeding lentil for resistance to biotic and abiotic stress. Euphytica 147: 149-165.

Murray MG, Thompson WF (1980) Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Research 8: 4321-4325.

Naumoff DG (2011) Hierarchical classification of glycoside hydrolases. Biochemistry (Moscow) 76: 622-635.

Nishizawa A, Yabuta Y, Shigeoka S (2008) Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. Plant Physiology 147: 1251-1263.

Nunes ACS, Vianna GR, Cuneo F, Amaya-Farfan J, Rech EL, Aragao FJL (2006) RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. Planta 224: 125-132.

Obendorf RL, Odorcic S, Ueda T, Coseo MP, Vassallo E (2004) Soybean galactinol synthase forms fagopyritol B1 but not galactopinitols: substrate feeding of isolated embryos and heterologous expression. Seed Science Research 14: 321-333.

Obendorf RL, Zimmerman AD, Ortiz PA, Taylor AG, Schnebly SR (2008) Imbibitional chilling sensitivity and soluble carbohydrate composition of low raffinose, low stachyose soybean seed. Crop Science 48: 2396-2403.

Obendorf RL, Zimmerman AD, Zhang Q, Castillo A, Kosina SM, Bryant EG, Sensenig EM, Wu J, Scnebly SR (2009) Accumulation of soluble carbohydrates during seed development and maturation of low-raffinose, low-stachyose soybean. Crop Science 49: 329-341.

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

Ogawa D, Yamaguchi K and Nishiuchi T (2007) High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased thermotolerance but also salt/osmotic stress tolerance and enhanced callus growth. Journal of Experimental Botany 58: 3373-3383.

Oosumi C, Nozaki J, Kida T (1998) Raffinose synthetase gene, process for reducing raffinose, and transformed plant. International Patent Publication WO98/49273, CT/JP97/03879.

Panikulangara TJ, Eggers-Schumacher G, Wunderlich M, Stransky H, Schoffl F (2004) Galactinol synthase1. A novel heat shock factor target gene responsible for heat-induced synthesis of raffinose family oligosaccharides in Arabidopsis. Plant Physiology 136: 3148-3158.

Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M (2009) Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC Molecular Biology 10(1): 11.

Pattanagul W, Madore MA (1999) Water deficit effects on raffinose family oligosaccharide metabolism in coleus. Plant Physiology 121: 987-993.

Pedersen A, Sandstrom B, Amelsvoort JMM (1997). The effect of ingestion of inulin on blood lipids and gastrointestinal symptoms in healthy females. British Journal of Nutrition 78: 215-222.

Peterbauer T, Richter A (1998) Galactosylononitol and stachyose synthesis in seeds of adzuki bean. Purification and characterization of stachyose synthase. Plant Physiology 117: 165-172.

Peterbauer T, Mucha J, Mayer U, Popp M, Glössl J, Richter A (1999) Stachyose synthesis in seeds of adzuki bean (*Vigna angularis*): molecular cloning and functional expression of stachyose synthase. Plant Journal 20: 509-518.

Peterbauer T, Richter A (2001a) Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Science Research 11: 185-197.

Peterbauer T, Lahuta LB, Blochl A, Mucha J, Jones DA, Hedley CL, Gorecki RJ, Richter A (2001b) Analysis of raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. Plant Physiology 127: 1764-1772.

Peterbauer T, Mach L, Mucha J, Richter A (2002a) Functional expression of a cDNA encoding pea (*Pisum sativum* L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. Planta 215: 839-846.

Peterbauer T, Mucha J, Mach L, Richter A (2002b) Chain elongation of raffinose in pea seeds: isolation, characterization, and molecular cloning of a multifunctional enzyme catalyzing the synthesis of stachyose and verbascose. Journal of Biological Chemistry 277: 194-200.

Peterbauer T, Karner U, Mucha J, Mach L, Jones DA, Hedley CL, Richter A (2003) Enzymatic control of the accumulation of verbascose in pea seeds. Plant Cell and Environment 26: 1385-1391.

Peters S, Mundree SG, Thomson JA, Farrant JM, Keller F (2007) Protection mechanisms in the resurrection plant *Xerophyta viscosa* (Baker): both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit. Journal of Experimental Botany 58: 1947-1956.

Peters S and Keller F (2009) Frost tolerance in excised leaves of the common bugle (*Ajuga reptans* L.) correlates positively with the concentrations of raffinose family oligosaccharides (RFOs). Plant, Cell & Environment 32: 1099-1107.

Peters S, Egert A, Stieger B, Keller F (2010) Functional identification of Arabidopsis ATSIP2 (At3g57520) as an alkaline α-galactosidase with a substrate specificity for raffinose and an apparent sink-specific expression pattern. Plant and Cell Physiology 51: 1815-1819.

Polowick PL, Baliski DS, Bock C, Ray H, Georges F (2009) Over-expression of α-galactosidase in pea seeds to reduce raffinose oligosaccharide content. Botany 87: 526-532.

Porres JM, Urbano G, Fernandez-Figares I, Prieto C, Perez L, Aguilera JF (2002) Digestive utilisation of protein and amino acids from raw and heated lentils by growing rats. Journal of the Science of Food and Agriculture 82: 1740-1747.

Price KR, Lewis J, Wyatt GM, Fenwick GR (1988) Flatulence causes, relation to diet and remedies. Nahrung 32: 609-626.

Puckaka S, Ratajczak E, Kalemba E (2009) Non-reducing sugar levels in beech (*Fagus sylvatica*) seeds as related to withstanding desiccation and storage. Journal of Plant Physiology 166: 1381-1390.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2014) Pfam: the protein families database. Nucleic Acids Research 42: D222-D230.

Quemener B, Brillouet JM (1983) Ciceritol, a pinitol digalactoside from seeds of chickpea, lentil and white lupin. Phytochemistry 22: 1745-1751.

Ramdath DD, Tsao R (2012) Bioactives and health benefits of lentils (*Lens culinaris L.*). In: Cereals and pulses: Nutraceutical properties and health benefits. Eds: Yu L, Tsao R Shahidi F. Wiley-Blackwell, Oxford, UK, pp 217-228.

Reddy NR, Pierson MD, Sathe SK, Salunkhe DK (1984) Chemical, nutritional and physiological aspects of dry bean carbohydrates: a review. Food Chemistry 13: 25-68.

Rennie EA, Turgeon R (2009) A comprehensive picture of phloem loading strategies. Proceedings of the National Academy of the Sciences of the United States of America 106: 14162-14167.

Richter A, Hoch G, Puschenreiter M, Mayer U, Peterbauer T (2000) The role of stachyose synthase in the oligosaccharide metabolism of legume seeds. In: Seed biology: Advances and applications. Eds: Black M, Bradford KJ, Vásquez-Ramos J. CAB International, Oxfordshire, USA, pp 75-84.

Roberfroid M B, Van Loo JAE, Gibson GR (1998) The bifidogenic nature of chicory inulin and its hydrolysis products. Journal of Nutrition 128: 11-19.

Robertson LD, Erskine W (1997) Lentil. In: Biodiversity in trust. Eds. Fuccillo D, Sears L, Stapleton P. Cambridge University Press, Cambridge, UK, pp 128-138.
Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant-tissues. Plant Molecular Biology 5: 69-76.
Rohde P, Hincha DK, Heyer AG (2004) Heterosis in the freezing tolerance of crosses between two *Arabidopsis thaliana* accessions (Columbia-0 and C24) that show differences in non-acclimated and acclimated freezing tolerance. The Plant Journal 38: 790-799.
Roy F, Boye JI, Simpson BK (2010) Bioactive proteins and peptides in pulse crops: pea, chickpea and lentil. Food Research International 43: 432-442.
Roy S, Kucukural A, Zhang Y (2010) I-Tasser: a unified platform for automated protein structure and function prediction. Nature Protocols 5: 725-738.
Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA
Saravitz D M, Pharr DM, Carter TE (1987) Galactinol synthase activity and soluble sugars in developing seeds of four soybean genotypes. Plant Physiology 83: 185-189.
Sarker A, Erskine W (2006) Recent progress in the ancient lentil. Journal of Agricultural Science 144: 19-29.
Saxena MC (2009) Plant Morphology, Anatomy and Growth Habit. In: The lentil: Botany, production and uses. Eds: Erskine W, Muehlbaur FJ, Sarker A and Sharma B. CAB International, Oxfordshire, UK, pp 34-46.
Schley PD, Field CJ (2002) The immune-enhancing effects of dietary fibres and prebiotics. British Journal of Nutrition 87: 221–230.
Schneider T, Keller F (2009) Raffinose in chloroplasts is synthesized in the cytosol and transported across the chloroplast envelope. Plant and Cell Physiology 50: 2174-2182.
Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology 7: 3.
Sebastian SA, Kerr PS, Pearlstein RW, Hitz WD (2000) Soybean germplasm with novel genes for improved digestibility. In: Soy in animal nutrition. Eds: Drackley JK. Federation of Animal Science Societies, Savoy, USA, pp 56–74.
Sengupta S, Mukherjee S, Parween S, Majumder AL (2012) Galactinol synthase across evolutionary diverse taxa: Functional preference for higher plants? FEBS Letters 586: 1488-1496.
Sharma A, Jood S, Sehgal S (1996) Antinutrients (phytic acid, polyphenols) and minerals (Ca, Fe) availability (in vitro) of chickpea and lentil cultivars. Nahrung 40: 182-184.
Sharpe AG, Ramsay L, Sanderson L, Fedoruk MJ, Clarke WE, Li R, Kagale S, Vijayan P, Vandenberg A, Bett KE (2013) Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. BMC Genomics 14: 192.
Sievenpiper JL, Kendall CW, Esfahani A, Wong JM, Carleton AJ, Jiang HY, Bazinet RP, Vidgen E, Jenkins DJ (2009) Effect of non-oil-seed pulses on glycaemic control: a systematic review and meta-analysis of randomised controlled experimental trials in people with and without diabetes. Diabetologia 52: 1479-95.
Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Molecular Systems Biology 7: 539.

Sika M, Terrab A, Swan PB, Hegarty PVJ (1995) Composition of selected Moroccon cereals and legumes: comparison with FAO table use in Africa. Food and Composition Analysis 8: 62-70.

Singh R, Hucl P, Båga M, Chibbar RN (2012) Validation of molecular markers for pre-harvest sprouting resistance in bread wheat. Cereal Research Communications 40: 194-203.

Skoneczka JA, Saghai Maroof MA,Shang C, Buss GR (2009) Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Science 49: 247-255.

Solanki IS, Kapoor AC, Singh U (1999) Nutritional parameters and yield evaluation of newly developed genotypes of lentil (*Lens culinaris* Medik.). Plant Foods for Human Nutrition 54: 79-87.

Sotomayor C, Frias J, Fornal J, Sadowska J, Urbano G, Vidal-Valverde C (1999) Lentil starch content and its microscopical structure as influenced by natural fermentation. Starch 51: 152-156.

Sprenger N, Keller F (2000) Allocation of raffinose family oligosaccharides to transport and storage pools in *Ajuga reptans*: the roles of two distinct galactinol synthases. The Plant Journal 21: 249-258.

Statistics Canada < http://www.statcan.gc.ca/start-debut-eng.html> Retrieved on 20-10-2014

Sui XL, Menga FZ, Wanga HY,Weia YX, Li RF, Wanga ZY, Hua LP, Wang SH, Zhanga ZX (2012) Molecular cloning, characteristics and low temperature response of raffinose synthase gene in *Cucumis sativus* L. Journal of Plant Physiology 169: 1883-1891.

Sun WQ, Irving TC, Leopold AC (1994) The role of sugar, vitrification and membrane phase transition in seed desiccation tolerance. Physiologia Plantarum 90: 621-628.

Tahir M, Vandenberg A, Chibbar RN (2011a) Influence of environment on seed soluble carbohydrates in selected lentil cultivars. Journal of Food Composition and Analysis 24: 596-602.

Tahir M, Lindeboom N, Båga M, Vandenberg A, Chibbar RN (2011b) Composition and correlation between major seed constituents in selected lentil (*Lens culinaris* Medik.) genotypes. Canadian Journal of Plant Science 91: 825-835.

Tahir M, Båga M, Vandenberg A, Chibbar RN (2012) An assessment of raffinose family oligosaccharides and sucrose concentration in genus Lens. Crop Science 52: 1713-1720.

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. The Plant Journal 29: 417-426.

Takasoye M, Inoue N, Knuma C (1990) Clinical investigation of feces improvements by soybean oligosaccharides. Rinsho Toh Kenkyu (Clinics and Research), (in Japanese) 67: 304-310.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30: 2725-2729.

Tanner W, Lehle L, Kandler O (1967) Myo-Inositol, a cofactor in the biosynthesis of verbascose. Biochemical and Biophysical Research Communications 29: 166-171.

Tanner W and Kandler O (1968) Myo-Inositol, a cofactor in the biosynthesis of stachyose. European Journal of Biochemistry 4: 233-239.

Tapernoux-Lüthi EM, Schneider T, Keller F (2007) The C-terminal sequence from common bugle leaf galactan:galactan galactosyltransferase is a non-sequence-specific vacuolar sorting determinant. FEBS Letters 581: 1811-1818.

Thavarajah D, Ruszkowski J and Vandenberg A (2008) High potential for selenium biofortification of lentils (*Lens culinaris* L.). Journal of Agricultural and Food Chemistry 56: 10747-10753.

Torres A, Frias J, Granito M, Vidal-Valverde C (2007) Germinated *Cajanus cajan* seeds as ingredients in pasta products: Chemical, biological and sensory evaluation. Food Chemistry 101: 202-211.

Turgeon R (1996) Phloem loading and plasmodesmata. Trends in Plant Science 1: 418-423.

Ueda T, Coseo MP, Thomas JH, Obendorf RL (2005) A multifunctional galactinol synthase catalyzes the synthesis of fagopyritol A1 and fagopyritol B1 in buckwheat seed. Plant Science 168: 681-690.

Unda F, Canam T, Preston L, Mansfield SD (2012) Isolation and characterization of galactinol synthases from hybrid poplar. Journal of Experimental Botany 63: 2059-2069.

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research 40: 115.

Urbano G, López-Jurado M, Hernandez M, Moreu MC, Frias J, Diaz- Pollan C, Prodanov M, Vidal-Valverde C (1995) Nutritional assessment of raw, heated, and germinated lentils. Journal of Agricultural Food Chemistry 43: 1871-1877.

Urbano G, Porres JM, Frías J, Vidal-Valverde C (2007) Nutritional value. Lentil: An ancient crop for modern times, Eds: Yadav S, McNeil DL, Stevenson PC. Springer, Netherlands, pp 47-93.

Van Barneveld RJ (1999) Understanding the nutritional chemistry of lupin (Lupinus spp.) seed to improve livestock production efficiency. Nutrition Research Reviews 12: 203-230.

Vandenberg A, Banniza S, Warkentin TD, Ife S, Barlow B, McHale S, Brolley B, Gan Y, McDonald C, Bandara M, Dueck S (2006) CDC Redberry lentil. Canadian Journal of Plant Science 86: 497-498.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002**)** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3: 0034.

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era. Current Opinion in Plant Biology 12: 202-210.

Verdier J, Lalanne D, Pelletier, Torres-Jerez Ivone, Righetti K, Bandyopadhyay K, Leprince O, Chatelain E, Vu E L, Gouzy J, Gamas P, Udvardi MK, Buitink J (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. Plant Physiology 163: 757-774.

Verma P, Shah N, Bhatia S (2013) Development of an expressed gene catalogue and molecular markers from the de novo assembly of short sequence reads of the lentil (*Lens culinaris* Medik.) transcriptome. Plant Biotechnology Journal 11: 894-905.

Vertucci CW, Farrant JM (1995) Acquisition and loss of desiccation tolerance. In: Seed development and germination. Eds: Kigel J, Galili G. Marcel Dekker Press, New York, USA, pp 237-271.

Vidal-Valverde C, Frias J (1992) Changes in carbohydrates during germination of lentils. Zeitschrift für Lebensmittel-Untersuchung und Forschung 194: 461-464.

Vidal-Valverde C, Frias J, Prodanov M, Tabera J, Ruiz R, Bacon J (1993a) Effect of natural fermentation on carbohydrates, riboflavin and trypsin inhibitor activity of lentils. Zeitschrift fur Lebensmitteluntersuchung und Forschung A 197: 449-452.

Vidal-Valverde C, Frias J, Valverde S (1993b) Changes in the carbohydrate composition of legumes after soaking and cooking. Journal of the American Dietetic Association 93: 547-550.

Vidal-Valverde C, Sierra JFI, Lambein IBF, Kuo Y (2002) New functional legume foods by germination: effect on the nutritive value of beans, lentils and peas. European Food Research Technology 215: 472-477.

Vijayakumari K, Siddhuraju P, Jonardhaman K (1996) Effect of soaking, cooking and autoclaving on phytic acid and oligosaccharide contents of the tribal pulse, *Mucuna monosperma* DC. ex. Wight. Food Chemistry 55: 173-177.

Voitsekhovskaja OV, Koroleva OA, Batashev DR, Knop C, Tomos AD, Gamalei YV, Heldt HW, Lohaus G (2006) Phloem loading in two Scrophulariaceae species. What can drive symplastic flow via plasmodesmata? Plant Physiology 140: 383-395.

Voitsekhovskaja OV, Rudashevskaya EL, Demchenko KN, Pakhomova MV, Batashev DR, Gamalei YV, Lohaus G, Pawlowski K (2009) Evidence for functional heterogeneity of sieve element–companion cell complexes in minor vein phloem of *Alonsoa meridionalis*. Journal of Experimental Botany 60: 1873-1883.

Volk GM, Haritos EE, Turgeon R (2003) Galactinol synthase gene expression in melon. Journal of the American Society or Horticultural Science 128: 8-15.

Wang GF, Wang G, Zhang XW, Wang F, Song RT (2012) Isolation of high quality RNA from cereal seeds containing high levels of starch. Phytochemical Analysis 23: 159-163.

Wada K, Watabe J, Mizutani J, Suzuki H, Kiriku N, Hayakawa K, Yamaguchi C (1991) Effects of soybean oligosaccharides intake on fecal microflora, enzyme activity, ammonia and frequency of evacuation in elderly persons. Bifidus 4: 135-140.

Wang N, Daun JK (2004) The chemical composition and nutritive value of Canadian pulses. Canadian Grain Commission, Winnipeg, Canada.

Wang N, Daun JK (2006) Effects of variety and crude protein content on nutrients and anti-nutrients in lentils (*Lens culinaris*). Food Chemistry 95: 493-502.

Wang N, Hatcher DW, Toews R, Gawalko EJ (2009) Influence of cooking and dehulling on nutritional composition of several varieties of lentils (*Lens culinaris*). LWT- Food Science and Technology 42: 842-848.

Wang N, Warkentin TD, Vandenberg B, Bing DJ (2014) Physiochemical properties of starches and various pea and lentil varieties, and characteristics of their noodles prepared by high temperature extrusion. Food Research International 55: 119-127.

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

Wang Z, Zhu Y,Wang L, Liu X, Liu Y, Phillips J, Deng X (2009) A WRKY transcription factor participates in dehydration tolerance in *Boea hygrometrica* by binding to the W-box elements of the galactinol synthase (BhGolS1) promoter. Planta 230: 1155-1166.

Watanabe E, Oeda K (1998) Raffinose synthetase genes and use thereof. European Patent Application EP0849359, 97122417.5.

Wiggins HS (1984) Nutritional value of sugars and related compounds undigested in the small intestine. Proceedings of the Nutrition Society 43: 69-75.

Wilfinger WW, Mackey K, Chomczynski P (1997) Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. BioTechniques 22: 474-481.

Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt A G (2011) Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative

polyadenylation. Proceedings of the National Academy of Sciences of the United States of America 108: 12533-12538.

Wu X, Gaffney B, Hunt AG, Li QQ (2014). Genome-wide determination of poly(A) sites in *Medicago truncatula*: evolutionary conservation of alternative poly(A) site choice. BMC Genomics 15: 615.

Xing D, Li QQ (2011) Alternative polyadenylation and gene expression regulation in plants. Wiley Interdisciplinary Reviews RNA 2: 445-458.

Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y (2010) Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. Plant Physiology 153: 1729-1746.

Ying SY (2004) Complementary cDNA libraries. Molecular Biotechnology 27: 245-252.

Zhang L, ZhiKe Z, Shunquan L, TingTing Z, XiangHui Y (2013) Evaluation of six methods for extraction of total RNA from loquat. Notulae Botanicae Horti Agrobotanici Cluj-Napoca. 41: 313-316.

Zhao T, Martin D, Meeley RB, Downie B (2004) Expression of the maize galactinol synthase gene family: (II) Kernel abscission, environmental stress and myo-inositol influences accumulation of transcript in developing seeds and callus cells. Physiologia Plantarum 121: 647-655.

Zhao TY, Meeley RB, Downie B (2003) Aberrant processing of a maize galactinol synthase transcript is caused by heat stress. Plant Science 165: 245-256.

Zhou J, Yang Y, Yu J, Wang LK, Yu X, Ohtani M, Kusano M, Saito K, Demura T, Zhuge Q (2014) Responses of Populus trichocarpa galactinol synthase genes to abiotic stresses. Journal of Plant Research 127: 347-358.

Zhuo CL, Wang T, Lu SY, Zhao YQ, Li XG, Guo ZF (2013) A cold responsive galactinol synthase gene from *Medicago falcata* (MfGolS1) is induced by myo-inositol and confers multiple tolerances to abiotic stresses. Physiologia plantarum 149: 67-78.

Zohary D (1972) The wild progenitor and the place of origin of the cultivated lentil: *Lens culinaris.* Economic Botany 26: 326-332.

Zohary D, Hopf M (1973) Domestication of pulses in the old world. Science 182: 887-894.

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Evolving Genes and Proteins. Eds: Bryson V, Vogel HJ. Academic Press, New York, USA, pp 97-166.

Zuther E, Büchel K, Hundertmark M, Stitt M, Hincha DK, Heyer AG (2004) The role of raffinose in the cold acclimation response of *Arabidopsis thaliana*. FEBS Letters 576: 169-73.

APPENDIX:

Appendix 4.1: Materials and methods used in the study

Chemicals used in this study

Sucrose was purchased from EMD Chemicals (Mississauga, ON, Canada). Standards of UDP-galactose (Uridine 5'-diphosphogalactose disodium salt), *myo*-inositol, galactinol, glucose, fructose, raffinose, stachyose and verbascose were purchased from Sigma-Aldrich (Sigma-Aldrich Inc., St. Louis, Missouri, USA). All / other chemicals used in this study were purchased from Sigma-Aldrich.

DNA Extraction from leaves

DNA was extracted from young leaves (two week old plants) of *Lens culinaris* cv CDC Redberry based on the method of Doyle and Doyle 1990. Briefly, fresh leaf tissue (0.5 g -1.0 g) was ground in liquid nitrogen, and the ground powder was transferred to 7.0 mL of pre-warmed (60 ^{o}C) extraction buffer [50 mM Tris-HCl pH 8.0, 20 mM EDTA (Ethylenediaminetetraacetic acid) pH 8.0, 3% CTAB, 1.4 M NaCl (Sodium chloride), 0.5% PVP (polyvinylpyrrolidone 360,000), 0.2% β-mercaptoethanol]. The powdered leaf material was gently mixed with the extraction buffer by inverting the tube several times. The leaf slurry was incubated at 60 ^{o}C for at least 30 min and inverted a few times to ensure equal mixing. An equal volume of chloroform/isoamylalcohol (24:1) was added to the extract and tubes were inverted ~50 times to obtain a homogenous solution followed by centrifugation at 2500 x g for 10 min at 23 ^{o}C (room temperature). The upper organic phase was transferred to a 15 mL centrifuge tube, and the above chloroform/ isoamylalcohol extraction step was repeated. The upper phase was transferred to a 15 mL centrifuge tube and approximately 2/3 volume of ice-cold isopropanol was added to the supernatant and mixed well by inverting the tube. The resulting mixture was incubated at -20 ^{o}C overnight and centrifuged for 2500 x g 15 min at 4^{o}C. Supernatant was discarded and pellet was treated with 10 mL of wash buffer (75% (v/v) ethanol, 10 mM NH_4Ac), incubated at 23 ^{o}C for 10 min and centrifuged at 2500 x g for 10 min at 23^{o}C. Supernatant was discarded and 10 mL of wash buffer was added to the pellet and centrifuged at 2500 x g for 10 min at 23^{o}C. Supernatant was the discarded and the resulting pellet was air dried and subsequently dissolved in 200-1000 µL TE buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA).

Degenerate primers design for isolation of partial sequence of RFO biosynthetic genes

Degenerate primers for amplification of partial RFO biosynthetic genes, nucleotide sequences for *GolS*, *RS* and *STS* from closely related species were aligned to determine conserved nucleotide sequences (Appendix 4.2). Primers were designed at conserved nucleotide sequences and primer characteristics were calculated using OligoEvaluator™ software accessible online (http://www.sigmaaldrich.com/life-science/custom-oligos/custom-dna/learning-center/calculator.html). Partial gene sequences of *GolS*, *RS* and *STS* amplified using degenerate primers were cloned using CloneJet PCR cloning kit (Thermo Fisher Scientific Inc., PA, USA) and plasmids were sequenced (National Research Council, Saskatoon, Canada) to confirm the nucleotide sequences.

Screening of cDNA library

A cDNA library was constructed using RNA from developing lentil seeds (18, 20, 22 and 24 DAF) to isolate full length sequences of RFO biosynthetic genes. The titre of the cDNA library was $1x10^9$ pfu/mL. XL1- Blue MRF cells grown in LB (Luria- Bertani) media with supplements (0.2% (w/v) maltose and 10 mM $MgSO_4$) to reach or achieve an OD of 0.7 at A_{600nm}. The cells were pelleted by centrifuging at 1000 x g for 10 min at 4 ^oC. The pelleted cells were diluted with 10 mM $MgS0_4$ to an OD (A_{600nm}) of 0.5. The equivalent of 2.5 x 10^4 pfu/plate was combined with 300 µL freshly prepared XL 1- Blue MRF cells and the mixture was incubated at 37 ^oC for 15 min. The mixture was added to 3 mL of NZY top agar (~48 ^oC) and poured into a dry, prewarmed NZY plate. The top agar was spread evenly and it was incubated at room temperature for 10 min. The plates were inverted and incubated at 37 ^oC for 5 - 8 h. The plates were stored at 4 ^oC for a minimum of 2 h before plaque lifting and hybridization.

Plaque lifting and hybridization was performed as per standard protocols (Sambrook and Russell 2001). Briefly, a nitrocellulose membrane was placed onto each NZY agar plate for 2 min to transfer phage particles. The plate along with membrane was pierced with a sterile needle at four random spots towards the periphery of the plate. This procedure was performed to identify the orientation of the membrane to the plate after hybridization. The membrane was removed carefully from the plate without removing top agar and a second membrane was placed onto the same agar plate for 4 min and pierced at same spots as the previous membrane. The membranes were incubated in denaturing solution (1.5 M NaCl and 0.5 M NaOH) for 2 min to

denature the DNA on the membrane. The membranes were then submerged in neutralization solution (1.5 M NaCl and 0.5 M Tris-HCl pH 8.0) for 5 min. The membranes were the transferred to a solution containing 0.2 M Tris-HCl (pH 7.5) and 2X SSC buffer, and incubated for no longer than 30 sec. The membranes were then transferred onto a Whatman™ filter paper and cross-linked under UV (UVP Inc., Upland, CA, USA). Hybridization was done using gene specific probes labelled with ^{32}P-dCTP using a random primer labelling kit (Life Technologies, Inc., Burlington, Ontario, Canada) as per manufacturer's instructions. The membranes were hybridized in Church hybridization buffer (0.5 M Na_2HPO4, 7% SDS (w/v) and 1 mM EDTA) at 65 ^{o}C O/N probe labelled with [α^{32}-P] dCTP. The membrane was washed two times in 1 × SSPE (150 mM NaCl; 10mM NaH2PO4. H2O; 1 mM EDTA), 0.1% SDS (w/v) for 10 min each, and then in 0.1 × SSPE, 0.1% SDS (w/v) until background activity was negligible. Membranes were exposed to KODAK® BioMax® film (Kodak) at -80 °C for 24 h to several days depending on signal intensities observed.

Positive plaques were selected based on the appearance of spots in replicate autoradiographs. Plaques were cored from the plate and transferred to a sterile microcentrifuge tube with 500 µL SM Buffer (100 mM NaCl, 8 mM $MgSO_4$. $7H_2O$, 50 mM Tris-Cl pH 7.5, gelatin 0.002% (w/v)) and 20 µL chloroform and mixed by vortexing to release phage particles. One microliter of the SM buffer along with phage particles was used as the template to amplify using gene specific primers to confirm presence of insert. Primary clones which showed the presence of insert (amplification of expected size) were chosen for secondary screening.

Phagemids were excised using ExAssist SOLR system (Agilent Technologies Inc., Santa Clara, CA) as per manufacturer's instructions. Briefly, XL1-Blue and SOLR positive cells (single colony was inoculated) were grown in LB broth with supplements (0.2% (w/v) maltose and 10mM $MgSO_4$) at 30 ^{o}C to an 0.7 at OD A_{600nm}. The cells were pelleted by centrifugation at 1000 x g for 10 min at 4^{o}C. The pelleted cells were resuspended in 10 mM $MgSO_4$ and diluted with 10 mM $MgSO_4$ to an OD of 1.0 at A_{600nm} (8 X 10^8 cells/mL). A 200 µL aliquot of XL 1-Blue MRF cells, 250 µL of phage stock (Phage particles suspended in SM buffer obtained from primary screening) and 1 µL of ExAssist helper phage (>1 X 10^6 pfu/µL) were combined in a 15 mL conical centrifuge tube and incubated at 37 ^{o}C for 15 min. Three milliliters of LB broth with supplements (0.2% (w/v) maltose and 10mM $MgSO_4$) was added to the centrifuge tube and incubated at 37 ^{o}C with shaking for 2.5 - 3 h. The centrifuge tubes were then incubated at 65 - 70

^{o}C for 20 min to lyze the lamda phage particles followed by centrifugation at 1000 x g for 15 min at 23 ^{o}C to pellet cell debris. The supernatant was transferred to a new centrifuge tube and two reactions were performed, 200 µL of freshly grown SOLR cells was combined with (i) 100 µL and (ii) 10 µL of phage supernatants. The two tubes were incubated at 37 ^{o}C for 15 min. A 100 µL aliquot of the cell mixture from each tube was transferred onto individual LB agar plates containing ampicillin (100 µg/mL) and incubated overnight at 37 ^{o}C. Three clones from each plates were sub-cultured overnight in LB media with ampicillin (100 µg/mL) and plasmids were isolated using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific Inc., Pittsburgh, PA, USA). Plasmids were digested with *EcoRI* and *XhoI* to determine insert lengths and isolated plasmids were DNA sequenced (National Research Council, Saskatoon, Canada) to obtain nucleotide sequences of the insert.

RFO constituent analysis

Lentil seeds at different development stages (developing seeds collected at two day intervals from 10 - 38 DAF) were freeze dried and ground. One hundred milligrams of dried seed sample was used for oligosaccharide extraction. *Myo*-inositol, galactinol, glucose, sucrose, fructose, raffinose, stachyose and verbascose concentration was determined using HPAEC- PAD (Gangola et al 2014). Briefly, soluble sugars were extracted by adding 2 mL of 80% (v/v) ethanol to 100 mg of lentil seed meal in a 15 mL conical tube. The slurry was thoroughly mixed by vortex for 30 sec and incubated at 60 ^{o}C for 45 min in a shaking water bath (Shaking Water Bath 25, Precision Scientific, Chicago, IL, USA) with intermittent mixing every 15 min. The slurry was centrifuged at 12,100 x g for 10 min and the supernatant was transferred to a new conical tube and the pellet was re-extracted as described above. The supernatants were pooled and purified by passage through a pre-equilibrated C_{18} solid phase column (Honeywell Burdick and Jackson, Muskegon, MI, USA) attached to a vacuum manifold so as to remove proteins and lipids present in the soluble extract. For equilibration, the C_{18} column was first washed with 5 mL of 99% (v/v) methanol followed by 5 mL of distilled water. The supernatant was then passed through the C_{18} column and 1.6 mL of the filtered extract was dried using Speedvac concentrator attached to a universal vacuum system (SPD111V, Thermo Savant, Holbrook, NY, USA). The pellet were dissolved in 500 µL of nanopure water, mixed vigorously and centrifuged at 10,000 x g for 10 min. A 125 µL of the supernatant was added to 375 µL water and this dilution was used for RFO constituent analysis. A Dionex ICS 5000 consisting of an autosampler, single gradient

pump, electrochemical detection cell with a disposable gold electrode and Ag/AgCl reference electrode (Dionex Canada Ltd., Oakville, ON, Canada) was used for carbohydrate analysis. A CarboPac PA 100 (4 x 250mm, Dionex) in series with a CarboPac PA 100 guard column (4x 50mm) maintained at 30 ^{o}C was employed for RFO analysis. The following gold standard PAD waveform with four different potentials (E) were used for the gold electrode as follows: +0.1 V for 400 millisecond (msec); -2.0 V for 20 msec (E2), 0.6 V for 10 msec (E3) and -0.1 V for 70 msec (E4). Filter sterilized, degassed solutions were used; solvent A (water) and solvent B (200mM NaOH). The gradient mobile phase maintained at 1 mL/min used for RFO analysis was as follows:

0 min : 90 % Solvent A and 10 % Solvent B

25 min : 0 % Solvent A and 100 % Solvent B

25 min : 90 % Solvent A and 10 % Solvent B

35 min : 90 % Solvent A and 10 % Solvent B

Carbohydrate concentration was calculated using Chromeleon 7.0 software employing standard curves (Dionex Canada Ltd., Oakville, ON, Canada). Standards for myo-inositol, galactinol, glucose, fructose, sucrose, raffinose, stachyose and verbascose were run in five concentration each 62.5, 125, 250, 500 and 1000 ng / injection volume (10µL). Coefficient of determination (R^2) values > 0.99 confirmed linearity of calibration curves. Peak area was used to calculate concentration of sugars in lentil seed meal.

Enzyme activity of RFO biosynthetic genes

Raffinose synthase activity was determined based on previously published methods (Peterbauer et al 2001a; Hitz et al 2002). The enzyme activity procedure optimized for pH, protein concentration, temperature and substrate concentration in *Cicer arietinum* (Unpublished data) was followed in this study. Two hundred milligrams of lentil seeds were ground with liquid nitrogen and 2 mL of extraction buffer (50 mM HEPES [2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid)-NaOH], pH 7.0; 5 mM 2-β mercaptoethanol). This mixture was centrifuged at 10,000 x g for 10 min at 4 ^{o}C and the supernatant was removed and filtered through a Sephadex G25 column pre-equilibrated with 5 mL of extraction buffer. The filtrate was quantified for protein concentration using a commercial dye binding assay (Bradford 1976) using a commercial kit (Bio-Rad protein assay, Bio-Rad Laboratories Inc., Mississauga, Ontario, Canada) with bovine serum albumin as the standard.

The enzyme assay was performed using 10 µg of protein for GS 30 µg protein concentration for RS and STS, 40 µg for VS and galactinol independent/GGT activity. Substrate concentrations for enzyme assays were:

(i) GS 20 mM *myo*-inositol, 10 mM UDP-galactose, 10 mM DTT and 5 mM $MnCl_2$

(ii) RS 10 mM galactinol, 10 mM DTT and 40mM sucrose

(iii) STS 10 mM galactinol, 10 mM DTT and 20 mM raffinose

(iv) VS 10 mM galactinol, 10 mM DTT and 20 mM stachyose

(v) GGT 10 mM DTT and 20 mM raffinose

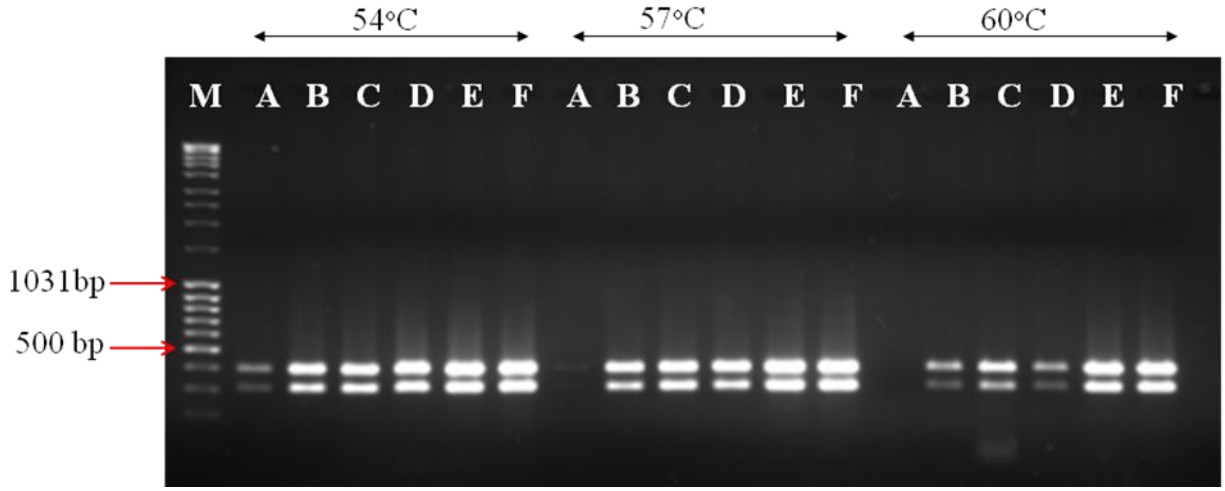
The reaction pH and temperature for all enzyme assays was 7.0 and 25 ^{o}C, respectively. Final volume for GS was 90 µL and 50 µL for RS, STS, VS and GGT. Incubation time was 10 min for GS and 60 min for all other enzyme assays. Two controls (i) protein and no substrate (ii) substrate and no protein was included for each protein sample. The reaction was terminated by the addition of 50 µL of 95% (v/v) ethanol and heating in a boiling water bath for 1 min. The tubes were centrifuged at 12,000 x g for 10 min and the supernatant was transferred to a new tube containing 10 mg of Dowex 50WX8 100 ion exchange resin (Sigma-Aldrich Inc., St. Louis, Missouri, USA) followed by mixing in a shaker at 200 rpm for 30 min. The tubes were then centrifuged for 5 min at 12000 x g for 10 min. The supernatant was transferred to a new tube and dried using a Speedvac concentrator and universal vacuum system. The pellet obtained was dissolved in 500 µL of nano pure water and mixed well. The tubes were then centrifuged at 10000 x g for 10 min. A 125 µL aliquot of the supernatant was added to 375 µL of water and this sample was analyzed by HPAEC-PAD as outlined in the previous section.

Appendix 4.2: Nucleotide sequences used for alignment to design degenerate primers to amplify RFO biosynthetic genes

Gene	**Species and accession numbers of sequences**	**Degenerate primer nucleotide sequences**
Galactinol Synthase	*A.mongolicus* (DQ519361.1)	DG/GS_F1 (Forward) 5'-CKGTGATGGAYTGTTTCTGTG-3' DG/GS_R1 (Reverse) 5'- CAGGRTGACGCCACAACAT-3'
	G.max (AY126715.1)	
	M. sativa (AY126615.1)	
	P.sativum (AJ243815.1)	
Raffinose synthase	*G.max* (EU651889)	DG/RS_F (Forward) 5'-ACRATCTTATCCACCGCYAAATCC-3' DG/RS_R (Reverse) 5'-GTCMTAATCGACGACGGKTG-3'
	P.sativum (AJ426475)	
	V.faba (E24423.1)	
Stachyose synthase	*P.sativum* (AJ512932.1)	DG/STS_F1 (Forward) 5'- ATYGCTAGCATGCAACAGTG-3' DG/STS_R1 (Reverse) 5'- CCCTTGACAGTTGAAAGCAC-3'
	V.angularis (YI9024)	

Appendix 4.3: Genes with accession numbers and primers used to amplify reference genes

Gene	Accession No	Tm	Start Base	Primer Sequence 5'-3'	Product Size
Actin (ACT)	GT626574.1	57.4	56	F- TCGTTCTTTGCCTTCCATCACC	199
		57.7	254	R- ATGCCAACCATCACGCCAGT	
Adenosyl Ribosylation Factor (ARF)	GT619582.1	59	374	F- AGTGGTTGAGGCAAGGGATGAGTT	153
		60.7	526	R- TGGCGGAGTGAATGAAGACCAAGT	
Alpha Tubulin (α-TUB)	GT619630.1	60.7	240	F- AAGCGTGCTTTCGTGCATTGGT	86
		60.5	325	R- TCAAGTGCAGCAAGATCCTCACGA	
Beta Tubulin (β-TUB)	GT624885.1	59.8	51	F- TCACCATGCGTCAAATCCTCCA	73
		59.3	123	R- ACTTCCCAGAATTTGGCACCGA	
Cyclophilin (CYP)	GT626020.1	59.6	333	F-TCAAGAAGCACACTGGTCCAGGAA	192
		59.8	524	R-AGGTCTTGCCAGATCCAGATCCAA	
Elongation Factor 1 alpha (EF1-α)	GT623512.1	58.2	492	F- AGCTGCCAACTTCACATCCCAA	198
		57.2	689	R- ACCATACCTGCATCGCCATTCT	
Gyceraldehyde-3-phosphate dehydrogenase (GAPD)	GT627530.1	60.5	167	F-ATGGACCATCAAGCAAGGACTGGA	169
		59.3	335	R- AGGTCAACAACGGAGACATCCACA	
Heat Shock Protein 90 (HSP)	GT619770.1	59.1	656	F- TGGATTCAGCCTTGATGAGCCA	109
		58.9	764	R- TGGCATGTCAGCATCAGCTTCA	
Phosphoprotein phosphatase 2A (PPA)	GT626594.1	59.2	239	F- AAGTCATGGCGCGTTCGTTACA	157
		58.5	395	R- TCACTTTCCCAGCAGCAGCAAT	
Transcription factor II A (TFIIA)	GT621745.1	60.2	452	F- GTGGAAAGCGGAAACGCAATGA	124
		59.8	579	R-AGATGCTCCCTCCACAAACCTCAA	



Appendix 4.4: PCR optimization for amplification of GS gene using degenerate primers. Lane M indicates DNA marker. Different Mg^{2+} concentrations and primer concentrations indicate lanes A (Primer 0.2mM, Mg^{2+} 1.5mM), B (Primer 0.2mM, Mg^{2+} 1.5mM), C (Primer 0.2mM, Mg^{2+} 1.5mM), D (Primer 0.2mM, Mg^{2+} 1.5mM), E (Primer 0.2mM, Mg^{2+} 1.5mM), F (Primer 0.2mM, Mg^{2+} 1.5mM). Reactions A-F were performed at different melting temperatures 54 ^{o}C, 57 ^{o}C and 60 ^{o}C.

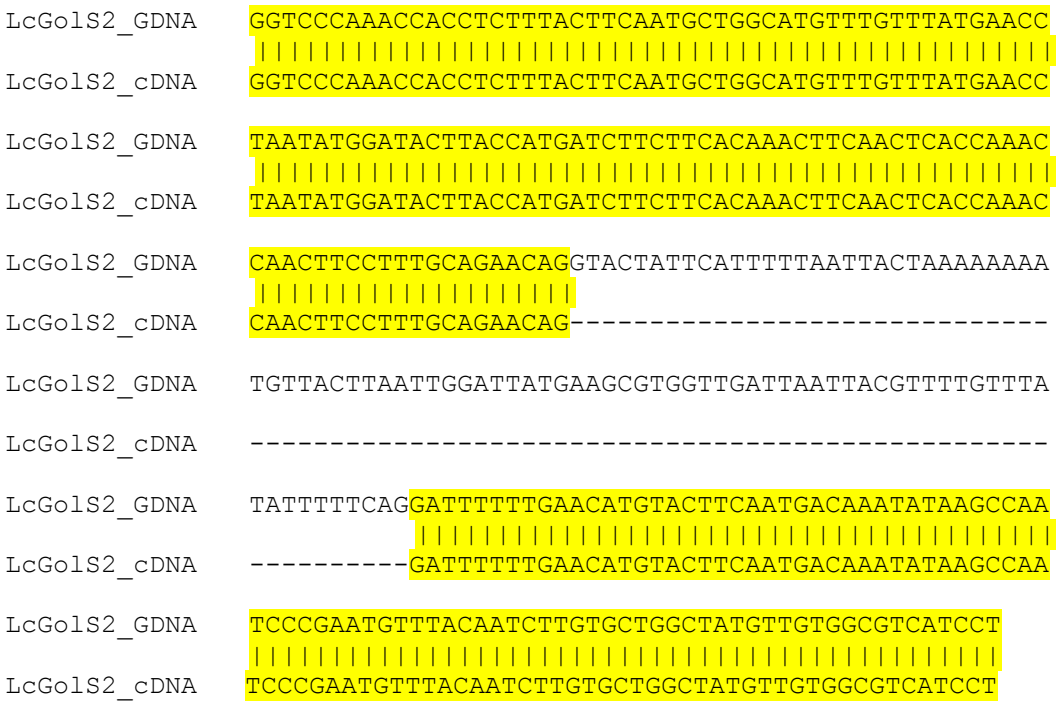
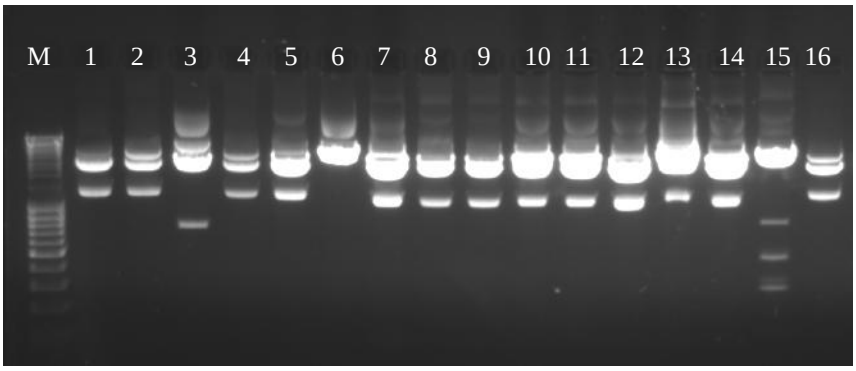


Figure 4.5: Sequence alignment of partial *LcGolS2*. GDNA indicates partial *LcGolS2* amplified from genomic DNA; cDNA indicates sequence amplified from cDNA from developing seed stages. Highlighted region indicate exons in the sequence (Intron / exon splicing sites were identified using Spidey software).



Appendix 4.6: Restriction digestion of *GolS1* phagemids obtained through secondary screening to check presence of inserts. Phagemids digested with enzymes *EcoR1* and *Xho1* were run on a 1% agarose gel. Lanes M indicates DNA ladder. Only one phagemid of the three replicates was digested to check insert sizes.

```
ataccccttacctccacaattcaactcacaatcatcaaacaaactcttaatcattctct    60
ctcacacttttccaaactctctaaaataaacagtttttagttttatcactaaaaaaaatg   120
                                                          M      1
gcaccggagatcgttccaacctcggcaaaatcagttaccactttcacgaaaccgaaacgt   180
 A  P  E  I  V  P  T  S  A  K  S  V  T  T  F  T  K  P  K  R     21
gcatacgtaacgtttctcgccggtaacggtgactacgtcaaaggcgtaattggtttggca   240
 A  Y  V  T  F  L  A  G  N  G  D  Y  V  K  G  V  I  G  L  A     41
aaaggtttacgtaaagtaaaaacggcgtacccgctggttgttgccgtgcttccagacgta   300
 K  G  L  R  K  V  K  T  A  Y  P  L  V  V  A  V  L  P  D  V     61
ccggaggagcaccgtgagatgctggaatctcagggatgtatcgtccgtgagattcaaccg   360
 P  E  E  H  R  E  M  L  E  S  Q  G  C  I  V  R  E  I  Q  P     81
gtatacccaccggaaaatcaaactcagtttgctatggcttattacgtcatcaattactcc   420
 V  Y  P  P  E  N  Q  T  Q  F  A  M  A  Y  Y  V  I  N  Y  S    101
aaactccgtatatgggagtttgtggagtacagcaagatgatatatttggatggagacata   480
 K  L  R  I  W  E  F  V  E  Y  S  K  M  I  Y  L  D  G  D  I    121
caggtatatgaaaacatagatcatcttttgatctacctgacggttacttctacgctgtg    540
 Q  V  Y  E  N  I  D  H  L  F  D  L  P  D  G  Y  F  Y  A  V    141
atggattgtttctgtgaaaaaacttggagtcatacaccacagtacaagattgggtactgt   600
 M  D  C  F  C  E  K  T  W  S  H  T  P  Q  Y  K  I  G  Y  C    161
caacagtgtccagagaaggtgcaatggccgaaagaaatgggtcaacctccttcactttac   660
 Q  Q  C  P  E  K  V  Q  W  P  K  E  M  G  Q  P  P  S  L  Y    181
ttcaatgctggcatgtttctgtttgagccaagcattgaaacttatcacgatcttttgaaa   720
 F  N  A  G  M  F  L  F  E  P  S  I  E  T  Y  H  D  L  L  K    201
acatgccaagctactcctcctactccgtttgcggaacaagatttcttgaacatgtatttt   780
 T  C  Q  A  T  P  P  T  P  F  A  E  Q  D  F  L  N  M  Y  F    221
aaggatatttataggccaattcctttggtttataatcttgttcttgctatgttgtggcgt   840
 K  D  I  Y  R  P  I  P  L  V  Y  N  L  V  L  A  M  L  W  R    241
catcctgagaatgttgaggtccggaaagtcaaggttgttcactattgtgcagcgggttca   900
 H  P  E  N  V  E  V  R  K  V  K  V  V  H  Y  C  A  A  G  S    261
aagccttggaggtatacaggtaaagaagagaatatgcagagggaggatataaagatgttg   960
 K  P  W  R  Y  T  G  K  E  E  N  M  Q  R  E  D  I  K  M  L    281
gtgcagaaatggtgggatatatacaatgattcttcacttgactacaagaagaatttgagt  1020
 V  Q  K  W  W  D  I  Y  N  D  S  S  L  D  Y  K  K  N  L  S    301
ggaaatggtgagacacaaaggaatgaggttgaagagccatttgtacaggcgttgtcggag  1080
 G  N  G  E  T  Q  R  N  E  V  E  E  P  F  V  Q  A  L  S  E    321
gttggccgtgttcgatatgtcacagcaccttcagcagcttaactgtggaagcagaaaaaa  1140
 V  G  R  V  R  Y  V  T  A  P  S  A  A  •                     334
gaaggaggaggtgcaacatgcaagactttgctttcggtgttagtagtcaaatagaatata  1200
tatacatttttttttgttgtattgtttttttctttttaatttgggaatttcggtatttgg  1260
ttttgtacatacggggtctgtatcatggtttgtattttcaggtatatattgttgtgcgtt  1320
tattgatgtgtttttatttatttagtttcttctgttgtaatcgaaatatataatcaaata  1380
tacttcgtgataaaggattgcaaaaaaaaaaaaaaaaa                        1419
```

Appendix 4.7: Nucleotide and deduced amino acid sequences encoding putative *LcGolS1* in lentil (Longest 5' UTR of Variant 1 and 3' UTR of variant 3 are shown in the sequence). The deduced protein sequence of *LcGolS1* is indicated below nucleotide sequence ('•' indicates translation termination site). Nucleotides coloured in blue indicate 5' and 3' UTR regions flanking the ORF of *LcGolS1*. Highlighted (Yellow) nucleotide region indicate the 312 bp partial *LcGolS1* fragment obtained by amplification using degenerate primers. Forwards and reverse arrows in the highlighted region indicate primers used for qRT-PCR analysis in this study. Nucleotides between forwards and reverse red arrows indicate nucleotides used as probes for cDNA library screening and RNA gel blot analysis.

```
aaacatcaaaacaatcaccaaatattcttgttcattgtgctttttcaaccttgtttatct   60

attaaccaaacatttctcagtgaaaccatggctcctgatctcataactgccgcaaccaac  120
                            M  A  P  D  L  I  T  A  A  T  N    11
atcaccgatgctcaaaccaagactgcaaaacgtgcctttgtgacgttccttgccgggaac  180
 I  T  D  A  Q  T  K  T  A  K  R  A  F  V  T  F  L  A  G  N    31
ggtgactacgtcaaaggtgttgagggcttagccaaaggtctgcgtaaggtgaaaaccatg  240
 G  D  Y  V  K  G  V  E  G  L  A  K  G  L  R  K  V  K  T  M    51
tgccctctcgtcgtggcagtgttgcccgatgttccggcggaacaccggaatattctgatc  300
 C  P  L  V  V  A  V  L  P  D  V  P  A  E  H  R  N  I  L  I    71
tcacaaggttgcattgttagagagattgaacctgtttacccacctgagaaccaaacccag  360
 S  Q  G  C  I  V  R  E  I  E  P  V  Y  P  P  E  N  Q  T  Q    91
ttcgcaatggcttattacgttatcaactactccaagctacgtatttgggcttttgaggaa  420
 F  A  M  A  Y  Y  V  I  N  Y  S  K  L  R  I  W  A  F  E  E   111
tatgagaagatgatataccttgacggtgacatacaagtttttgaaaacattgaccattta  480
 Y  E  K  M  I  Y  L  D  G  D  I  Q  V  F  E  N  I  D  H  L   131
tttgacttgccaaataactatttctatgccgtgatggattgtttctgtgaggccacatgg  540
 F  D  L  P  N  N  Y  F  Y  A  V  M  D  C  F  C  E  A  T  W   151
ggtcacaccaagcaatatgaaatcggttactgtcagcaatgtcctgacaaggttcaatgg  600
 G  H  T  K  Q  Y  E  I  G  Y  C  Q  Q  C  P  D  K  V  Q  W   171
ccaagtaattttggtcccaaaccacctctttacttcaatgctggcatgtttgtttatgaa  660
 P  S  N  F  G  P  K  P  P  L  Y  F  N  A  G  M  F  V  Y  E   191
cctaatatggatacttaccatgatcttcttcacaaacttcaactcaccaaaccaacttcc  720
 P  N  M  D  T  Y  H  D  L  L  H  K  L  Q  L  T  K  P  T  S   211
tttgcagaacaggatttttgaacatgtacttcaatgacaaatataagccaatcccgaat  780
 F  A  E  Q  D  F  L  N  M  Y  F  N  D  K  Y  K  P  I  P  N   231
gtttacaatcttgtgctggctatgttgtggcgtcacccctgagaatgttgagcttcacaaa  840
 V  Y  N  L  V  L  A  M  L  W  R  H  P  E  N  V  E  L  H  K   251
gttaaagttgttcattactgtgcagctggttctaagccttggagatacactggtgtggag  900
 V  K  V  V  H  Y  C  A  A  G  S  K  P  W  R  Y  T  G  V  E   271
gagaatatgcagagagaagatataaagatgttggtgaagaagtggtgggaggtatatgaa  960
 E  N  M  Q  R  E  D  I  K  M  L  V  K  K  W  W  E  V  Y  E   291
gacgagagtttggattacaagaaagcgcttaatgagaatcgcttaacatcagcaattttg 1020
 D  E  S  L  D  Y  K  K  A  L  N  E  N  R  L  T  S  A  I  L   311
gaagctggtggaatcaactttgtccgcgctccaaacgctgcttaatttgcgtgctttact 1080
 E  A  G  G  I  N  F  V  R  A  P  N  A  A  •                  325
taccaacttcaacaagatcttctgcaatatttttatttcgctaggggatcaaacataatc 1140
aagttttttatttttatttttgtaaagtaattacgaggactttaatttttattacgagccct 1200
cgttagttttttgtttcctatttttatttttgtatttatttccatttgtaattaactatatc 1260
ctttgtcctttattataatactgatctcccatataattttaagataaaaaaaaaaaaaaa 1320
aaaaaaaaaaaa                                                   1333
```

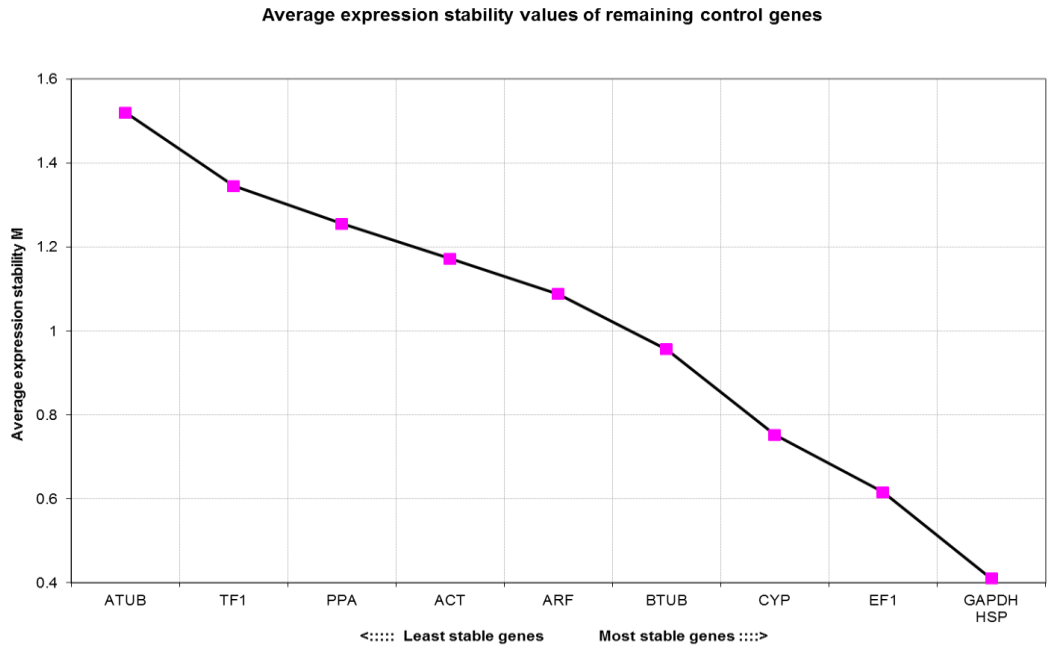
Appendix 4.8: Nucleotide and deduced amino acid sequences encoding putative *LcGolS2* in lentil (Longest 3' UTR of variant 3 is shown in the sequence). The deduced protein sequence of *LcGolS2* is indicated below nucleotide sequence ('•' indicates translation termination site). Nucleotides coloured in blue indicate 5' and 3' UTR regions flanking the ORF of *LcGolS2*. Highlighted nucleotide region indicate the partial *LcGolS2* fragment obtained by amplification using degenerate primers. Red forwards and reverse arrows in the highlighted region indicate primers used for qRT-PCR analysis in this study. Black forward and reverse arrows indicate nucleotides used as primers for amplification of full length *LcGolS2* sequence using RACE (Rapid amplification of cDNA ends) strategy.

```
LcGolS1-a      ATACCCCTTTACCTCCACAATTCAACTCACAATCATCAAACAAACTCTTAATCATTCTCT
LcGolS1-b      ------------------------------------------------------------
LcGolS1-c      ------------------------------------------------------------

LcGolS1-a      CTCACACTTTTCCAAACTCTCTAAAATAAACAGTTTTTAGTTTTATCACTAAAAAAAATG
LcGolS1-b      ---ACACTTTTCCAAACTCTCTAAAATAAACAGTTTTTAGTTTTATCACTAAAAAAAATG
LcGolS1-c      ---ACACTTTTCCAAACTCTCTAAAATAAACAGTTTTTAGTTTTATCACTAAAAAAAATG
                  ************************************************************

LcGolS1-a      GCACCGGAGATCGTTCCAACCTCGGCAAAATCAGTTACCACTTTCACGAAACCGAAACGT
LcGolS1-b      GCACCGGAGATCGTTCCAACCTCGGCAAAATCAGTTACCACTTTCACGAAACCGAAACGT
LcGolS1-c      GCACCGGAGATCGTTCCAACCTCGGCAAAATCAGTTACCACTTTCACGAAACCGAAACGT
               ************************************************************

LcGolS1-a      GCATACGTAACGTTTCTCGCCGGTAACGGTGACTACGTCAAAGGCGTAATTGGTTTGGCA
LcGolS1-b      GCATACGTAACGTTTCTCGCCGGTAACGGTGACTACGTCAAAGGCGTAATTGGTTTGGCA
LcGolS1-c      GCATACGTAACGTTTCTCGCCGGTAACGGTGACTACGTCAAAGGCGTAATTGGTTTGGCA
               ************************************************************

LcGolS1-a      AAAGGTTTACGTAAAGTAAAAACGGCGTACCCGCTGGTTGTTGCCGTGCTTCCAGACGTA
LcGolS1-b      AAAGGTTTACGTAAAGTAAAAACGGCGTACCCGCTGGTTGTTGCCGTGCTTCCAGACGTA
LcGolS1-c      AAAGGTTTACGTAAAGTAAAAACGGCGTACCCGCTGGTTGTTGCCGTGCTTCCAGACGTA
               ************************************************************

LcGolS1-a      CCGGAGGAGCACCGTGAGATGCTGGAATCTCAGGGATGTATCGTCCGTGAGATTCAACCG
LcGolS1-b      CCGGAGGAGCACCGTGAGATGCTGGAATCTCAGGGATGTATCGTCCGTGAGATTCAACCG
LcGolS1-c      CCGGAGGAGCACCGTGAGATGCTGGAATCTCAGGGATGTATCGTCCGTGAGATTCAACCG
               ************************************************************

LcGolS1-a      GTATACCCACCGGAAAATCAAACTCAGTTTGCTATGGCTTATTACGTCATCAATTACTCC
LcGolS1-b      GTATACCCACCGGAAAATCAAACTCAGTTTGCTATGGCTTATTACGTCATCAATTACTCC
LcGolS1-c      GTATACCCACCGGAAAATCAAACTCAGTTTGCTATGGCTTATTACGTCATCAATTACTCC
               ************************************************************

LcGolS1-a      AAACTCCGTATATGGGAGTTTGTGGAGTACAGCAAGATGATATATTTGGATGGAGACATA
LcGolS1-b      AAACTCCGTATATGGGAGTTTGTGGAGTACAGCAAGATGATATATTTGGATGGAGACATA
LcGolS1-c      AAACTCCGTATATGGGAGTTTGTGGAGTACAGCAAGATGATATATTTGGATGGAGACATA
               ************************************************************

LcGolS1-a      CAGGTATATGAAAACATAGATCATCTTTTTGATCTACCTGACGGTTACTTCTACGCTGTG
LcGolS1-b      CAGGTATATGAAAACATAGATCATCTTTTTGATCTACCTGACGGTTACTTCTACGCTGTG
LcGolS1-c      CAGGTATATGAAAACATAGATCATCTTTTTGATCTACCTGACGGTTACTTCTACGCTGTG
               ************************************************************

LcGolS1-a      ATGGATTGTTTCTGTGAAAAAACTTGGAGTCATACACCACAGTACAAGATTGGGTACTGT
LcGolS1-b      ATGGATTGTTTCTGTGAAAAAACTTGGAGTCATACACCACAGTACAAGATTGGGTACTGT
LcGolS1-c      ATGGATTGTTTCTGTGAAAAAACTTGGAGTCATACACCACAGTACAAGATTGGGTACTGT
               ************************************************************

LcGolS1-a      CAACAGTGTCCAGAGAAGGTGCAATGGCCGAAAGAAATGGGTCAACCTCCTTCACTTTAC
LcGolS1-b      CAACAGTGTCCAGAGAAGGTGCAATGGCCGAAAGAAATGGGTCAACCTCCTTCACTTTAC
LcGolS1-c      CAACAGTGTCCAGAGAAGGTGCAATGGCCGAAAGAAATGGGTCAACCTCCTTCACTTTAC
               ************************************************************

LcGolS1-a      TTCAATGCTGGCATGTTTCTGTTTGAGCCAAGCATTGAAACTTATCACGATCTTTTGAAA
LcGolS1-b      TTCAATGCTGGCATGTTTCTGTTTGAGCCAAGCATTGAAACTTATCACGATCTTTTGAAA
LcGolS1-c      TTCAATGCTGGCATGTTTCTGTTTGAGCCAAGCATTGAAACTTATCACGATCTTTTGAAA
               ************************************************************
```

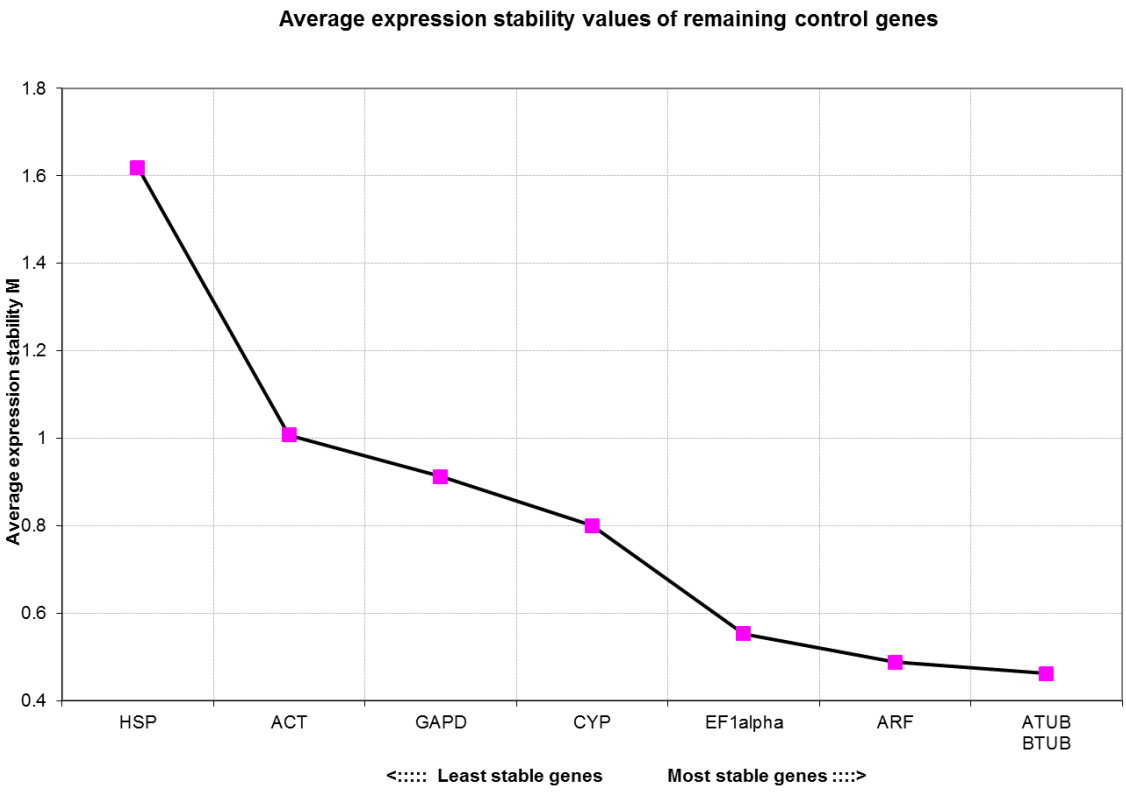
Appendix 4.9: Alignment of *LcGolS1* variants to indicate differences in 5' and 3' UTR lengths. Highlighted nucleotides indicate start and stop codon.

```
LcGolS1-a      ACATGCCAAGCTACTCCTCCTACTCCGTTTGCGGAACAAGATTTCTTGAACATGTATTTT
LcGolS1-b      ACATGCCAAGCTACTCCTCCTACTCCGTTTGCGGAACAAGATTTCTTGAACATGTATTTT
LcGolS1-c      ACATGCCAAGCTACTCCTCCTACTCCGTTTGCGGAACAAGATTTCTTGAACATGTATTTT
               ************************************************************

LcGolS1-a      AAGGATATTTATAGGCCAATTCCTTTGGTTTATAATCTTGTTCTTGCTATGTTGTGGCGT
LcGolS1-b      AAGGATATTTATAGGCCAATTCCTTTGGTTTATAATCTTGTTCTTGCTATGTTGTGGCGT
LcGolS1-c      AAGGATATTTATAGGCCAATTCCTTTGGTTTATAATCTTGTTCTTGCTATGTTGTGGCGT
               ************************************************************

LcGolS1-a      CATCCTGAGAATGTTGAGGTCCGGAAAGTCAAGGTTGTTCACTATTGTGCAGCGGGTTCA
LcGolS1-b      CATCCTGAGAATGTTGAGGTCCGGAAAGTCAAGGTTGTTCACTATTGTGCAGCGGGTTCA
LcGolS1-c      CATCCTGAGAATGTTGAGGTCCGGAAAGTCAAGGTTGTTCACTATTGTGCAGCGGGTTCA
               ************************************************************

LcGolS1-a      AAGCCTTGGAGGTATACAGGTAAAGAAGAGAATATGCAGAGGGAGGATATAAAGATGTTG
LcGolS1-b      AAGCCTTGGAGGTATACAGGTAAAGAAGAGAATATGCAGAGGGAGGATATAAAGATGTTG
LcGolS1-c      AAGCCTTGGAGGTATACAGGTAAAGAAGAGAATATGCAGAGGGAGGATATAAAGATGTTG
               ************************************************************

LcGolS1-a      GTGCAGAAATGGTGGGATATATACAATGATTCTTCACTTGACTACAAGAAGAATTTGAGT
LcGolS1-b      GTGCAGAAATGGTGGGATATATACAATGATTCTTCACTTGACTACAAGAAGAATTTGAGT
LcGolS1-c      GTGCAGAAATGGTGGGATATATACAATGATTCTTCACTTGACTACAAGAAGAATTTGAGT
               ************************************************************

LcGolS1-a      GGAAATGGTGAGACACAAAGGAATGAGGTTGAAGAGCCATTTGTACAGGCGTTGTCGGAG
LcGolS1-b      GGAAATGGTGAGACACAAAGGAATGAGGTTGAAGAGCCATTTGTACAGGCGTTGTCGGAG
LcGolS1-c      GGAAATGGTGAGACACAAAGGAATGAGGTTGAAGAGCCATTTGTACAGGCGTTGTCGGAG
               ************************************************************

LcGolS1-a      GTTGGCCGTGTTCGATATGTCACAGCACCTTCAGCAGCTTAACTGTGGAAGCAAAAAAAA
LcGolS1-b      GTTGGCCGTGTTCGATATGTCACAGCACCTTCAGCAGCTTAACTGTGGAAGCAGAAAAAA
LcGolS1-c      GTTGGCCGTGTTCGATATGTCACAGCACCTTCAGCAGCTTAACTGTGGAAGCAGAAAAAA
               ************************************************** ******

LcGolS1-a      AAA-------------------------------------------------------
LcGolS1-b      GAAGGAGGAGGTGCAACATGCAAGACTTTGCTGTCGGTGTTAGTAGTCAAATAGAATATA
LcGolS1-c      GAAGGAGGAGGTGCAACATGCAAGACTTTGCTTTCGGTGTTAGTAGTCAAATAGAATATA
                **

LcGolS1-a      ------------------------------------------------------------
LcGolS1-b      TATACATTTTTTTTTGTTGTATTGTTTTTTTCTTTTTAATTTGGGAATTTCGGTATTTGG
LcGolS1-c      TATACATTTTTTTTTGTTGTATTGTTTTTTTCTTTTTAATTTGGGAATTTCGGTATTTGG


LcGolS1-a      ------------------------------------------------------------
LcGolS1-b      TTTTGTACATACGGGGTCTGTATCATGGTTTGTATTTTCAGGTATATATTGTTGTGCGTT
LcGolS1-c      TTTTGTACATACGGGGTCTGTATCATGGTTTGTATTTTCAGGTATATATTGTTGTGCGTT


LcGolS1-a      ------------------------------------------------------------
LcGolS1-b      TATTGAAAAAAAAAAAAAAAAAAAAA----------------------------------
LcGolS1-c      TATTGATGTGTTTTTATTTATTTAGTTTCTTCTGTTGTAATCGAAATATATAATCAAATA


LcGolS1-a      ----------------------------------------
LcGolS1-b      ----------------------------------------
LcGolS1-c      TACTTCGTGATAAAGGATTGCAAAAAAAAAAAAAAAAAA
```

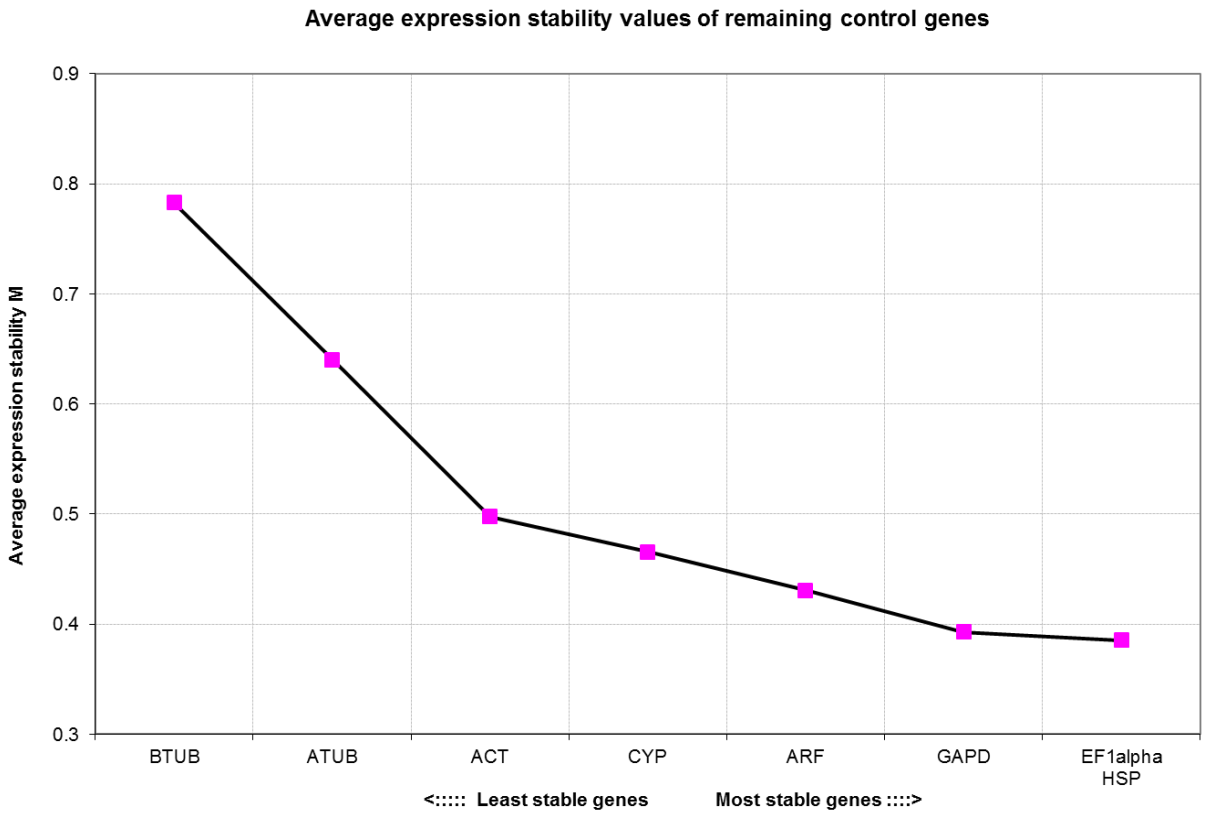
Appendix 4.9 Continued: Alignment of *LcGolS1* variants to indicate differences in 5' and 3' UTR lengths. Highlighted nucleotides indicate start and stop codon.



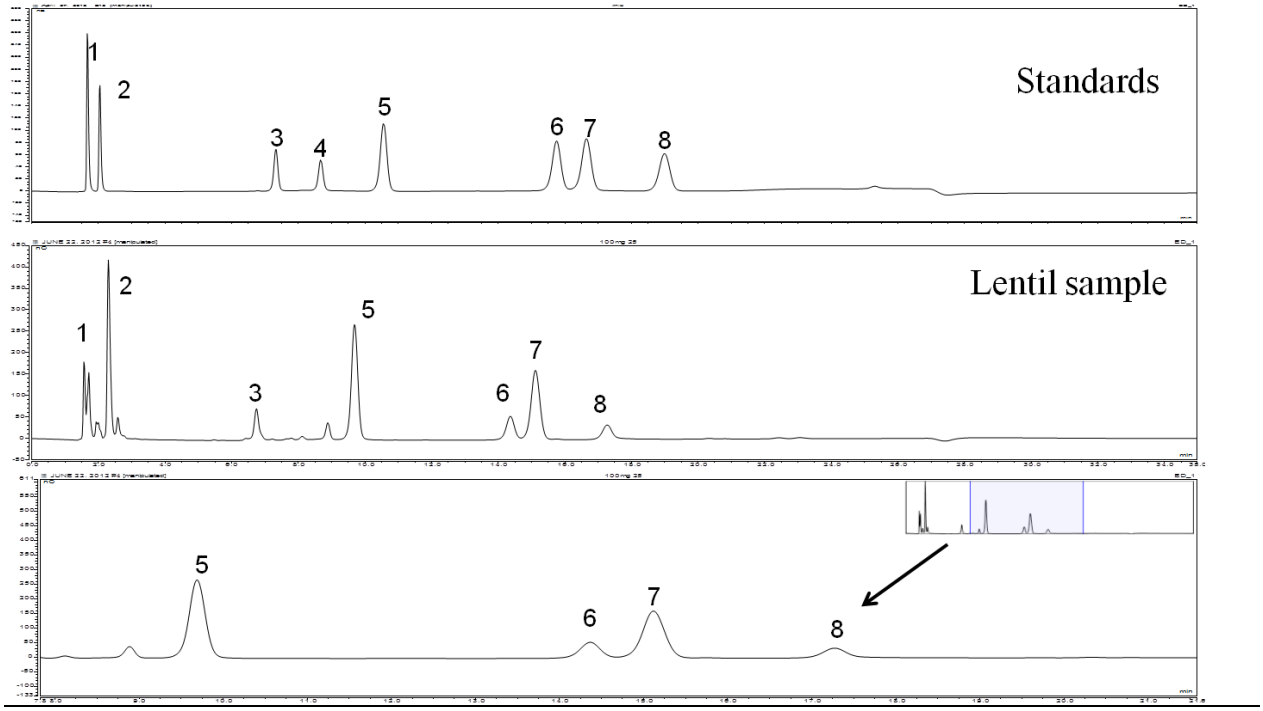
Appendix 4.10: Reference gene validation during seed development (ACT- Actin, ARF - Adenosyl ribosylation factor, ATUB - α - Tubulin, BTUB - β - Tubulin, CYP - Cyclophilin, EF1 - Elongation factor 1 α, GAPD - Glyceraldehyde - 3 - phosphate dehydrogenase, HSP - Heat shock protein, PPA - Phosphoprotein phosphatase, TF1- Transcription factor II A).



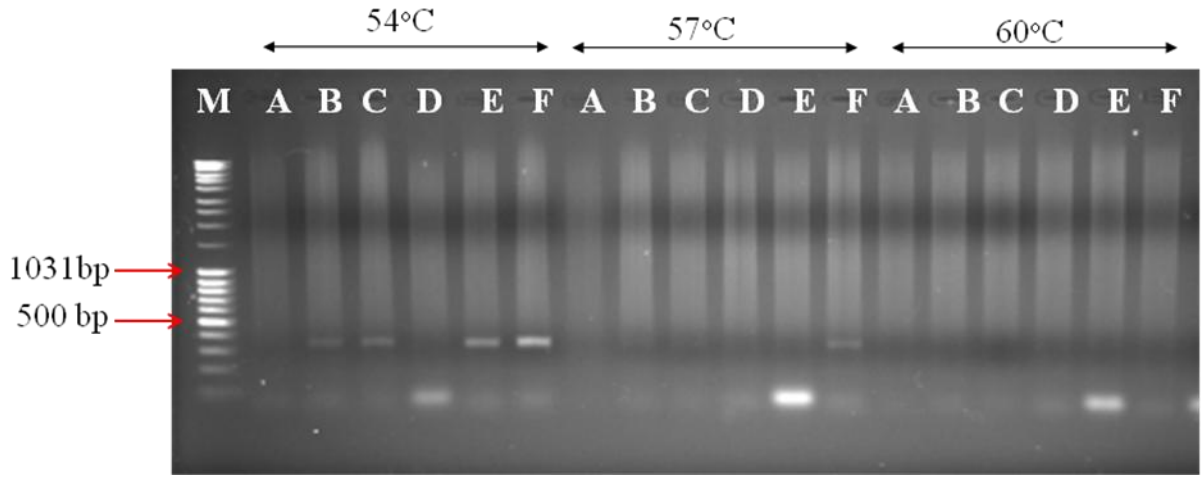
Appendix 4.11: Reference gene validation for leaf samples during circadian rhythm (ACT- Actin, ARF - Adenosyl ribosylation factor, ATUB - α - Tubulin, BTUB - β - Tubulin, CYP - Cyclophilin, EF1 - Elongation factor 1 α, GAPD - Glyceraldehyde - 3 - phosphate dehydrogenase, HSP - Heat shock protein, PPA - Phosphoprotein phosphatase, TF1- Transcription factor II A).



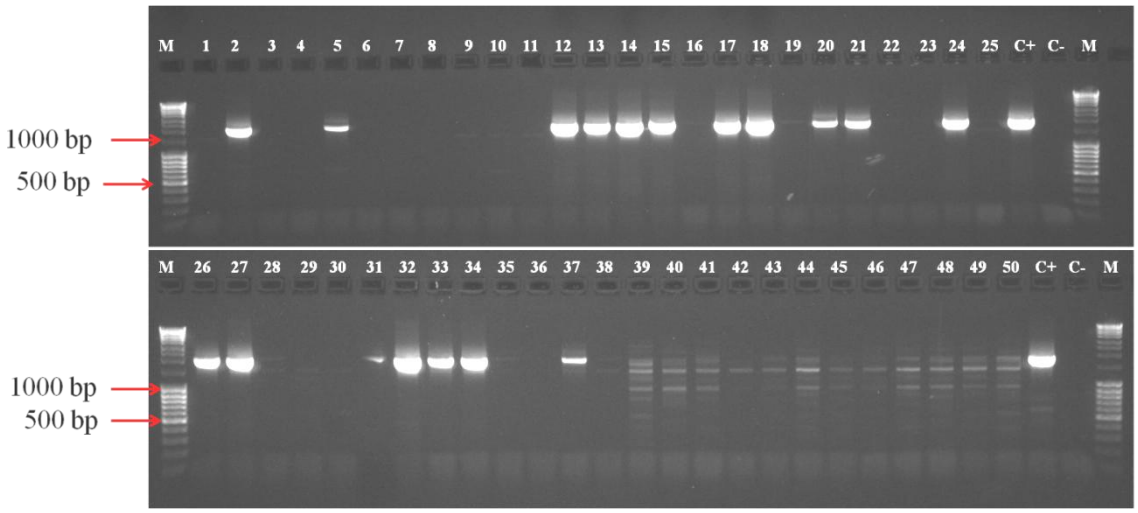
Appendix 4.12: Stability of reference gene for quantitative real time PCR analysis across various tissues- Leaf, roots and stem (ACT- Actin, ARF - Adenosyl ribosylation factor, ATUB - α - Tubulin, BTUB - β - Tubulin, CYP - Cyclophilin, EF1 - Elongation factor 1 α, GAPD - Glyceraldehyde - 3 - phosphate dehydrogenase, HSP - Heat shock protein, PPA - Phosphoprotein phosphatase, TF1- Transcription factor II A).



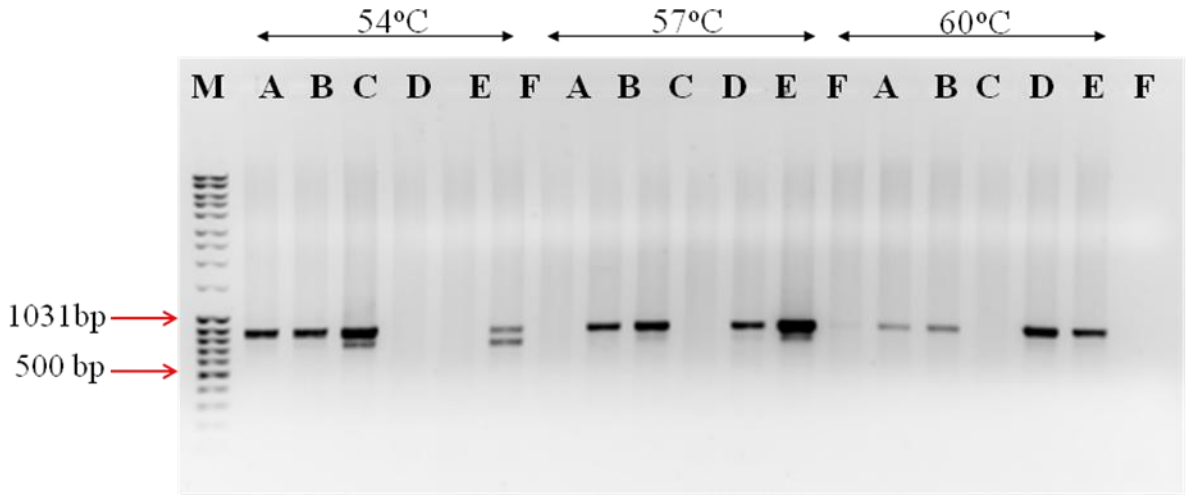
Appendix 4.13: Oligosaccharide separation obtained using HPAEC-PAD. Peak indicates compounds: 1 – *myo* inositol, 2 – galactinol, 3 – glucose, 4 – fructose, 5 – sucrose, 6 – raffinose, 7 – stachyose and 8 – verbascose



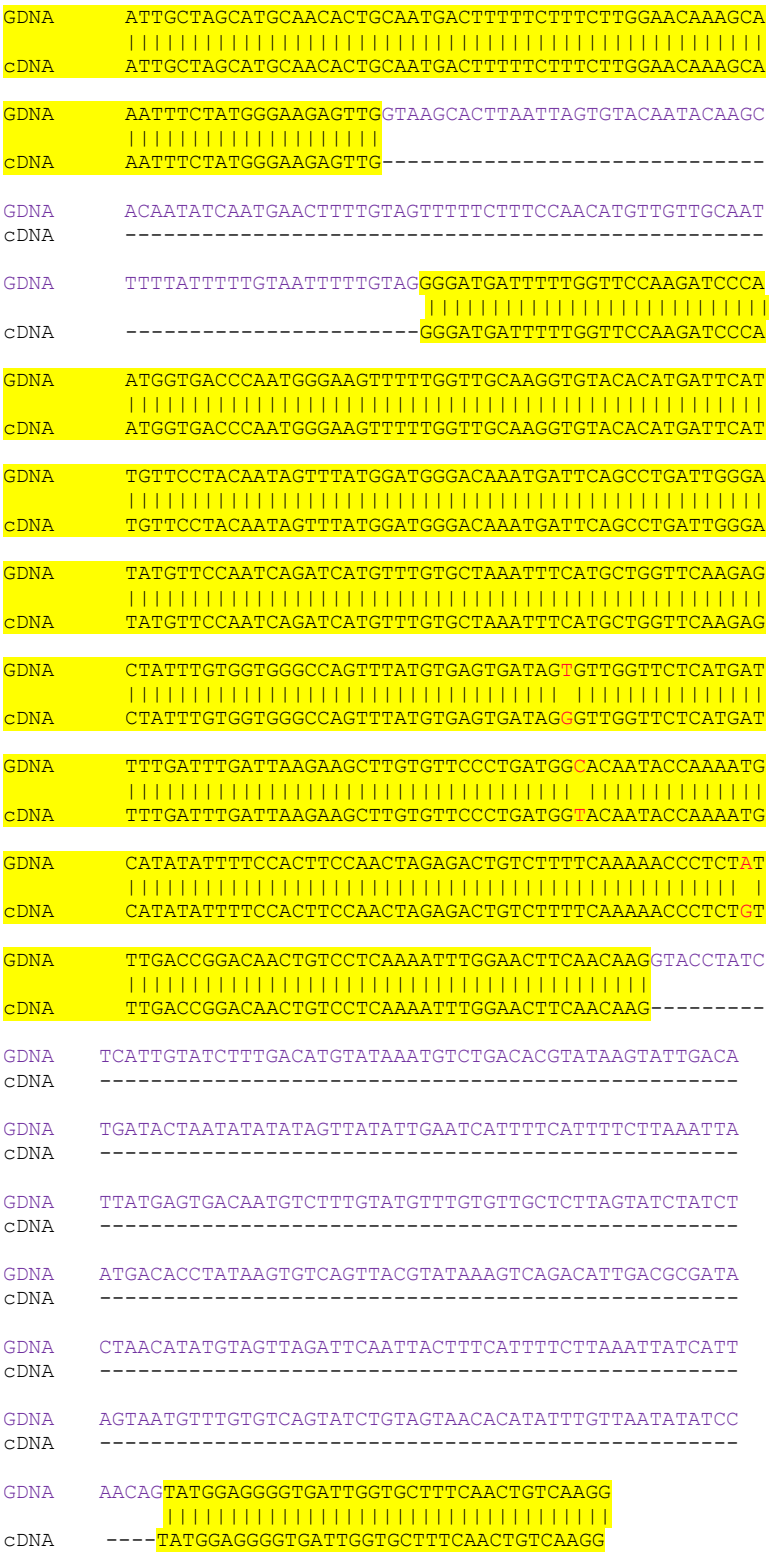
Appendix 5.1: PCR optimization for amplification of partial RS sequence using degenerate primers. Lane M indicates DNA marker. Different Mg^{2+} concentrations and primer concentrations indicate lanes A (Primer 0.2mM, Mg^{2+} 1.5mM), B (Primer 0.2mM, Mg^{2+} 1.5mM), C (Primer 0.2mM, Mg^{2+} 1.5mM), D (Primer 0.2mM, Mg^{2+} 1.5mM), E (Primer 0.2mM, Mg^{2+} 1.5mM), F (Primer 0.2mM, Mg^{2+} 1.5mM). Reactions A-F were performed at different melting temperatures 54 ^{o}C, 57 ^{o}C and 60 ^{o}C.



Appendix 5.2: Amplification of plaques obtained through primary screening for *LcRS* to check for presence of insert. Plaques were amplified using gene specific primers using plaques as a template for confirmation of the presence of insert. M indicates DNA ladder.



Appendix 6.1: PCR optimization for amplification of STS gene using degenerate primers. Lane M indicates DNA marker. Different Mg^{2+} concentrations and primer concentrations indicate lanes A (Primer 0.2mM, Mg^{2+} 1.5mM), B (Primer 0.2mM, Mg^{2+} 1.5mM), C (Primer 0.2mM, Mg^{2+} 1.5mM), D (Primer 0.2mM, Mg^{2+} 1.5mM), E (Primer 0.2mM, Mg^{2+} 1.5mM), F (Primer 0.2mM, Mg^{2+} 1.5mM). Reactions A - F were performed at different melting temperatures 54 ^{o}C, 57 ^{o}C and 60 ^{o}C. Band showing high intensity was cloned and DNA sequenced indicates *LcSTS1*.



Appendix 6.2: Sequence alignment of 884 bp *LcSTS* fragment amplified from genomic DNA and 1.3 kb *LcSTS* contig obtained from *Lens culinaris* transcriptome shotgun assembly. Highlighted region (Yellow) indicate exons and introns are indicated in blue. Nucleotides indicated in red indicates nucleotide differences from sequences obtained from the database.

Appendix 6.3: List of primers used in the study

Gene	Primer Name	Primer sequence (5' - 3')
Adenosyl ribosylation factor (*ARF*)	ARF_1F	AGTGGTTGAGGCAAGGGATGAGTT
	ARF_1R	TGGCGGAGTGAATGAAGACCAAGT
Elongation factor 1-α (*EF*)	EF_RT_F	AGCTGCCAACTTCACATCCCAA
	EF_RT_R	ACCATACCTGCATCGCCATTCT
Galactinol synthase 1 (*LcGolS1*)	GS1_1F	ATGGCACCGGAGATCGTTCC
	GS_1R	CACGGCACCTTCAGCAGCTTAA
	GS1_RT_F	AGAAGGTGCAATGGCCGAAAGA
	GS1_RT_R	GCTTGGCTCAAACAGAAACATGCC
Galactinol synthase 2 (*LcGolS2*)	GS2_RACE/R	GGATGACGCCACAACATAGCCAGCAC
	GS2_RACE/F	ATGGCTCCTGATCTCATAACTGC
	GS2_RT_F	TGATGGACTGTTTCTGTGAGGCCA
	GS2_RT_R	TGGCCATTGAACCTTGTCAGGA
Glyceraldehyde- 3 - phosphate dehydrogenase (*GAPD*)	GAPD_RT_F	ATGGACCATCAAGCAAGGACTGGA
	GAPD_RT_R	AGGTCAACAACGGAGACATCCACA
Granule bound starch synthase gene (*GbssI*)	GBSS_1F	AGGGTTCATTGGCAGGCTAGAAGA
	GBSS_1R	TGCTATCCCGATTGCTTTGCCA
Heat shock protein (*HSP*)	HSP_RT_F	TGGATTCAGCCTTGATGAGCCA
	HSP_RT_R	TGGCATGTCAGCATCAGCTTCA
Raffinose synthase (*LcRS*)	F/R 17	ACAAACGGGCACGAACTACAACAC
	R/R 1356	ACAAATTCCACACTAACCCCATCAAT
	RS_RACE/F1	AGGAAAGGGTGGCCATTTGCAAGGAA
	RS_RACE/F2	TGACTTTTGGGCTCGGTGGTGTTGAA
	RS_RT_F	TTTGGCATGCGCTTTGTGGGTA
	RS_RT_R	TTATCCACCGCCAAATCCTCCA
Stachyose synthase (*LcSTS*)	STS1RT_F	GTGGGCCAGTTTATGTGAGTGA
	STS1RT_R	TTGGTATTGTGCCATCAGGGA