Cystathionine γ-Synthase from *Cicer arietinum*, *Pisum sativum*, and *Lens culinaris* – Initial Genetic Characterization and Investigation of a Putative Second CGS Isoform in *Arabidopsis thaliana*

By

Jennifer Elaine Skanes

B.Sc. Honours, Carleton University, 2007

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Carleton University

Ottawa, Ontario, Canada

Your file Votre référence
ISBN: 978-0-494-87817-0

Our file Notre référence
ISBN: 978-0-494-87817-0

ABSTRACT

Methionine is an essential amino acid for humans and non-ruminant animals in that it cannot be synthesized *de novo* and therefore must be obtained in the diet. In plants, the committing step towards methionine biosynthesis is catalyzed by the enzyme cystathionine γ-synthase (CGS). This study represents the first characterization of CGS from the plant species *Cicer arietinum* (chickpea), *Pisum sativum* (field pea), and *Lens culinaris* (lentil), in which methionine is a limiting nutrient from a human perspective, and a CGS-like locus in the *Arabidopsis thaliana* genome (At1g33320). The CGS coding sequences from chickpea, field pea, and lentil are 70.0-71.0% identical to the characterized *A. thaliana* CGS sequence (At3g01120) at the amino acid level, contain the plant-specific conserved MTO1 regulatory region within the first exon, and complement the methionine auxotrophy of an *Escherichia coli* strain lacking the *metB* gene, encoding CGS. The expression pattern of CGS in these species is similar to that of the *A. thaliana* At3g01120 locus. Investigation of published plant genomes suggests that CGS is likely single copy in these diploid pulse species, in contrast to the two to three copies identified in the genomes of some *Brasicaceae* family members, including *A. thaliana*. Phenotypic assessment of *A. thaliana* T-DNA insertion lines demonstrated that while homozygous insertions within the coding region for the At1g33320 locus do not result in an altered phenotype, a heterozygous T-DNA line with an insertion in the At3g01120 locus demonstrated retarded growth, delayed flowering, and fewer siliques. It is proposed that the At1g33320 locus accounts for little to none of the total CGS activity in *A. thaliana*. A more detailed understanding of CGS and its regulation in pulses will assist in increasing the level of the limiting nutrient methionine in these crop species.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Susan Aitken, for providing me the opportunity to work in her lab on this project and others over the past several years. Dr. Aitken is a teacher and mentor of the highest calibre, and was a guiding factor in my decision to pursue teaching. I would also like to thank my committee members Dr. Shelley Hepworth, Dr. Brian Miki, and Dr. Iain Lambert for all their advice, feedback, and support, both for my research and career goals. Thank you to the members of the faculty and staff of the Biology Department for their advice and assistance, especially Ed Bruggink, Dr. Owen Rowland, Dr. Ashkan Golshani, and Dr. Andrew Simons.

My fellow students in the Aitken lab, past and present, became like a second family to me, and I would like to say thanks for all the friendship, fun times, and support through the years. A special acknowledgement to the Fabaceae Family members Ally Jaworski, Dominique Morneau, and honorary member Christine Lowe for protocol advice, plant care, research assistance, and general awesomeness. A big thank you to other Aitken lab members Emily Hopwood, Edgar Abou Assaf, and Colleen Woodhouse, as well as Matt Meier, William Hughes, and Nikolai Chepelev, graduate students from other research labs, for all the research help they've provided.

I would like to thank the members of the Let's Talk Science for making my three years as a coordinator so amazing, especially Dr. Rob Smith, Michelle Zenko, Anna Chkrebtii, and all of the staff at National Office. I must also extend a special thank you to Michelle and Anna for giving me the opportunity to participate in Science Travels.

This would not be complete without acknowledgements to all of my family and friends who have been so loving and supportive over all of these years. Thank you.

DEDICATION

TO:

ANDRÉ AND KATIE

ALISON AND AMY

MOM AND DAD

AND THE SKANES CLAN

With love and thanks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABRC	Arabidopsis Biological Resource Center
Acetyl-CoA	Acetyl-coenzymeA
aCGS	Arabidopsis CGS At3g01120
AHC	Adenosyl homocysteinase
AK	Aspartate kinase
ALS	Acetolactate synthase
APPA	$_{DL}$-E-2-amino-5-phosphono-3-pentanoic acid
APR	Adenosine 5'-phosphosulphate reductase
APS	ATP sulphurase
ASDH	Aspartate semialdehyde dehydrogenase
ATP	Adenosine triphosphate
BCAT	Branched chain aminotransferase
BgSd	Full-sized seed
BSA	Bovine serum albumin
CBL	Cystathionine β-lyase
cDNA	Complementary DNA
CGS	Cystathionine γ-synthase
CSC	Cysteine synthase complex
DAPAT	Diaminopimelate aminotransferase
DAPDC	Diaminopimelate decarboxylase
DAPE	Diaminopimelate epimerase
DHADH	Dihydroxy-acid dehydratase
DHDPR	Dihydropicolinate reductase
DHDPS	Dihydropicolinate synthase
DNA	Deoxyribonucleic acid
dNTPs	Deoxy nucleotide triphosphates
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EF1α	Elongation factor 1α
eFP	Electronic Fluoresce Pictograph
ESTs	Expressed sequence tags
F	Flower
gDNA	Genomic DNA
HEPES	4-(2-hydroxyethyl)-1-piperazineethanosulfonic acid
HMT	Homocysteine methyltransferase
HSD	Homoserine dehydrogenase
HSK	Homoserine kinase
HTS	Homoserine transsuccinase

IPTG	Isopropyl β-D-1-thiogalactopyryoside
Ka/Ks	Non-synonymous: synonymous mutation ratio
KARI	Ketol-acid reductoisomerase
KZ	Kozak sequence
LB	Luria broth
LUC	Luciferase
MGL	Methionine γ-lyase
ML	Mature leaf
MMT	Methionine methyltransferase
mRNA	Messenger RNA
MS	Methionine synthase
MTO	Methionine over accumulation
noRT	No reverse transcriptase
OAHS	*O*-acetylhomoserine
OAS	*O*-acetylserine
OAS-TL	*O*-acetylserine (thiol) lyase
O.D.	Optical density
OPHS	*O*-phosphohomoserine
OSHS	*O*-succinylhomoserine
PCR	Polymerase chain reaction
Pd	Seed pod
PLP	Pyridoxal-5'-phosphate
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
Rt	Root
SAH	*S*-adenosyl homocysteine
SAM	*S*-adenosylmethionine
SAMDC	SAM decarboxylase
SAMS	SAM synthase
SAT	Serine acetyltransferase
SDS	Sodium dodecyl sulphate
SIR	Sulphite reductase
SMM	*S*-methylmethionine
SmSd	Developing seed
St	Stem
SULTR	Sulphate uptake and transport
TAE	Tris acetic acid EDTA
TAIR	The Arabidopsis Information Resource
TB	Transformation buffer – for competent cell preparation
T-DNA	Transferred DNA

TD	Threonine deaminase
TIC	Translocation at the outer envelope membrane of the chloroplast
TOC	Translocation at the inner envelope membrane of the chloroplast
Tris	Tris(hydroxymethyl)aminomethane
TS	Threonine synthase
UTR	Untranslated region
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YL	Young leaf

1. INTRODUCTION

1.1 Importance of Amino Acid Balance in Food Crops

Humans and other non-ruminant animals cannot synthesize all of the 20 proteinogenic amino acids *de novo*, and those that must be obtained through diet are referred to as essential amino acids. While animal-based proteins generally contain an appropriate ratio of the essential amino acids, plant proteins may not contain an optimal balance of the required amino acids to satisfy human or livestock dietary requirements (Hesse *et al*, 2004). The amino acids methionine and lysine are of particular importance since they are nutritionally deficient in legumes and cereal grains, respectively, which are staples of human and animal diets throughout the world. Therefore, individuals who follow a vegetarian diet, or cannot easily obtain meat, and rely on pulses as staple foods, must be careful to balance food sources (*e.g.* cereals and pulses) to ensure that they receive sufficient amounts of all of the essential amino acids (Gepts *et al*, 2005).

1.1.1 Importance of Pulse Crops

The Food and Agriculture Organization (FAO) of the United Nations defines pulse crops as those legumes that are harvested primarily for dry grain, such as chickpeas (*Cicer arietinum*), lentils (*Lens culinaris*), vetches (*Vicia sp.*), lupins (*Lupinus sp.*), and some peas (*Pisum sp.*) and beans (*Phaseolus sp.*; Food and Agriculture Organization, 1994). These species are all members of the *Fabaceae* family of plants, which also includes the forage crops alfalfa (*Medicago sativa*), clover (*Medicago sp., Trifolium sp.*) and *Lotus japonicus*, the vegetable crops peanut (*Arachis hypogaea*), and other peas and beans, and the oilseed crop soybean (*Glycine max*). In Canada, pulse crops are not only

important as food and feed sources, but also from an economic perspective, as an important export commodity. Based on data from the 2009 growing season, Canada was the world's second largest producer of pulse crops, ranking first in the world for dry pea production, second for lentils, and fifth for chickpeas. Of the 5.2 million tonnes of pulse crops produced in that season, 75% was exported, to countries including India (25% of exports) and Bangladesh (12% of exports), representing a $2.1 billion dollar (CND) industry, and 40% of the world's trade in pulses (Agriculture and Agri-Food Canada, 2010).

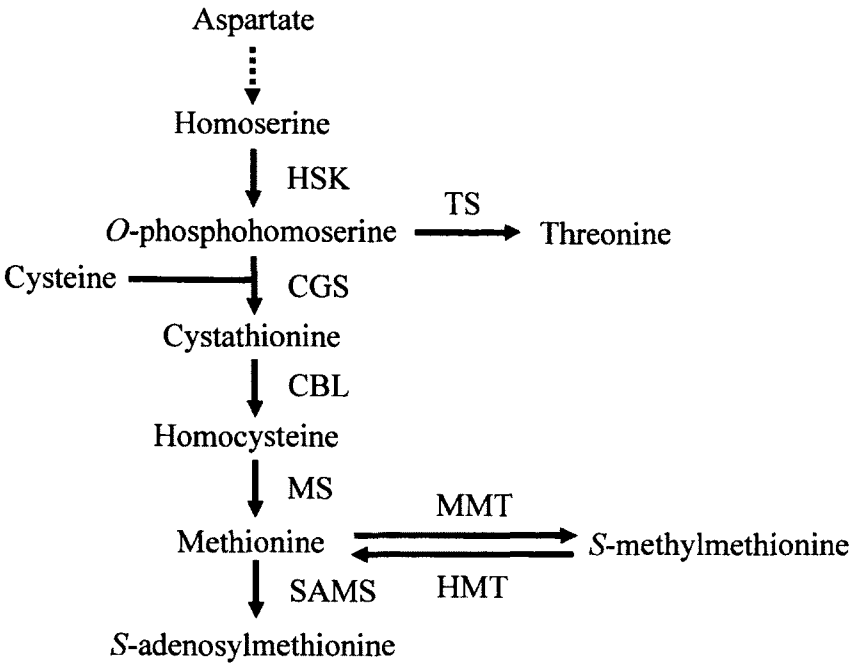
Grain legumes contain on average 20-40% protein by weight, and are a nutritionally significant source of all 15 essential minerals required in the human diet; however their deficiency in the sulphur containing amino acids, particularly methionine, requires that a diet reliant on pulses as a primary protein source be supplemented with methionine or with a methionine-rich protein source (Wang *et al*, 2003). Human nutritional deficiencies in methionine are manifested as potentially irreversible retardation of mental and physical development, and children under the age of four are particularly vulnerable (Muntz *et al*, 1998). Therefore, altering the nutritional composition of pulses to increase the level of methionine would be beneficial for human and animal nutrition. Production of such value-added crops could also boost the Canadian market for agricultural exports. To this end, understanding the methionine biosynthetic pathway in plants, particularly in the pulses chickpea, field pea, and lentil, is a critical first step.

1.2 Methionine Biosynthesis in Plants

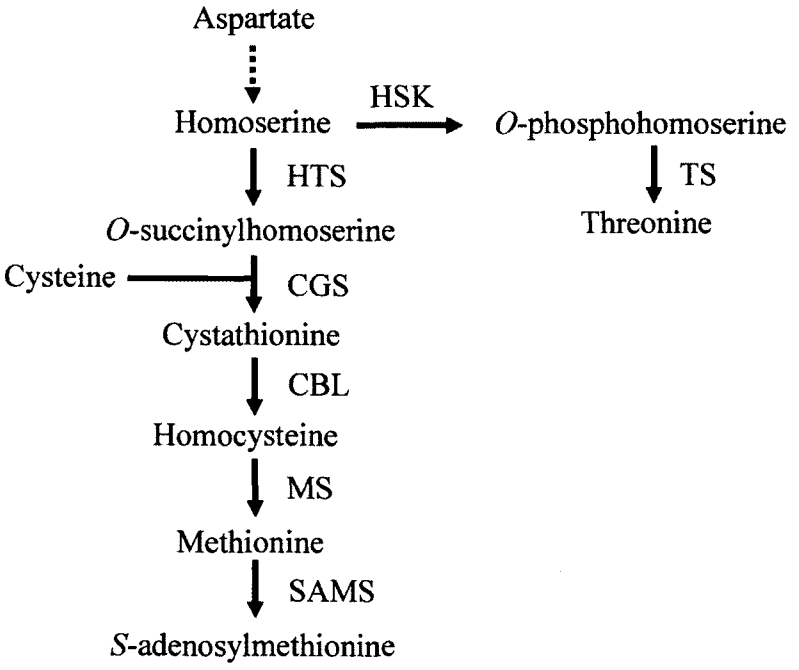
While animals and fungi synthesize cysteine from a serine backbone, and derive the sulphydryl moiety from methionine, bacteria and plants perform the opposite interconversion – synthesizing methionine by the transfer of the sulphide group from cysteine and a methyl group from methyl-tetrahydrofolate to an aspartate-derived backbone (Hacham *et al*, 2003). Cysteine is synthesized *de novo*, and subsequently converted to homocysteine, *via* the enzymes of the transsulphuration pathway (Figure 1). In bacteria the branch point metabolite between methionine and threonine synthesis is homoserine, whereas in plants the branch-point metabolite is *O*-phosphohomoserine (OPHS; Hacham *et al*, 2003). In the plant transsulphuration pathway cystathionine γ-synthase (CGS) catalyzes the condensation of OPHS and cysteine to form cystathionine, in the committing step in methionine synthesis (Datko *et al*, 1974). Cystathionine subsequently undergoes a β-elimination reaction, catalyzed by cystathionine β-lyase (CBL), to form homocysteine, which is methylated by methionine synthase (MS) to methionine (Giovanelli and Mudd, 1971; Eichel *et al*, 1995). Methionine and its derivative S-adenosylmethionine (SAM) are critical cellular metabolites; methionine is required for protein synthesis and SAM production, while SAM is a ubiquitous methyl donor and is a key biosynthetic precursor in the synthesis of lignin, biotin, polyamines, and the plant hormone ethylene. The majority (up to 80%) of the methionine synthesized in the cell proceeds to SAM or to *S*-methylmethionine (SMM) for methionine transport and storage (Jander and Joshi, 2010).

Figure 1 - Overview of the methionine biosynthetic pathway in: A) plants, and B) bacteria. Dashed arrows indicate multiple enzymatic conversions; solid arrows indicate single step enzymatic conversions. Enzyme abbreviations are: CBL – cystathionine β-lyase; CGS – cystathionine γ-synthase; HSK – homoserine kinase; HMT – homocysteine methyltransferase; HTS – homoserine transsuccinase; MMT – methionine methyltransferase; MS – methionine synthase; SAMS – *S*-adenosylmethionine synthase; TS – threonine synthase.

A


Aspartate
Homoserine
HSK
TS
O-phosphohomoserine
Threonine
Cysteine
CGS
Cystathionine
CBL
Homocysteine
MS
MMT
Methionine
S-methylmethionine
SAMS
HMT
S-adenosylmethionine

B


Aspartate
Homoserine
HSK
O-phosphohomoserine
HTS
TS
Threonine
O-succinylhomoserine
Cysteine
CGS
Cystathionine
CBL
Homocysteine
MS
Methionine
SAMS
S-adenosylmethionine

1.2.1 Cysteine Synthesis

Plants take up sulphate from the surrounding environment and transport it within the plant for use. Sulphur uptake and transport is provided by the high- and low- affinity sulphur transporter systems (SULTR). Once in the cell sulphate is adenylated by ATP-sulphurylase to adenosine 5'-phosphosulphate (APS; Figure 2). This is reduced to sulphite by APS reductase (APR) and further to sulphide by sulphite reductase (SIR). Sulphur metabolism in plants has been studied in great detail, and has been shown to be under complex regulation. Sulphur limitation demonstrates multiple levels of control over the sulphur metabolic pathway including through the action of transcription factors, microRNAs, and protein-protein interactions (reviewed in Takahashi *et al*, 2011).

Serine provides the backbone for cysteine synthesis, as sulphide replaces the acetate moiety of *O*-acetylserine (OAS), an activated form of serine, to yield cysteine (Saito, 2004). In plants, serine is condensed with acetyl-CoA by serine acetyltransferase (SAT) to form OAS. The diverse sub-cellular localizations of the various SAT isoforms have been implicated in the synthesis of cysteine for different end products, as required by each organelle, although primary synthesis occurs in the mitochondria (Figure 2; Noji *et al*, 1998). *O*-acetylserine is then combined with sulphide to form cysteine with the release of acetate by *O*-acetylserine (thiol) lyase (OAS-TL; Wirtz *et al*, 2004). Serine acetyltransferase and OAS-TL interact in plants to form the cysteine-synthase complex (CSC), and the regulation of cysteine synthesis is primarily dependant on the successful formation of an active complex. The OAS-TL substrate OAS has been demonstrated to destabilize the CSC, whereas sulphide has a stabilizing effect, thus controlling cysteine metabolism according to sulphur availability (Alvarez *et al*, 2010).

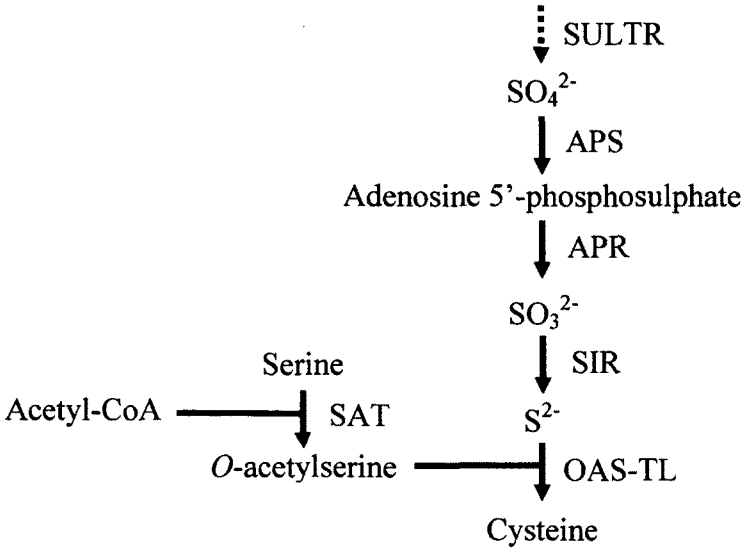


Figure 2 – Cysteine biosynthetic pathway in plants. Dashed arrows indicate multiple enzymatic conversions; solid arrows indicated single step enzymatic conversions. Enzyme abbreviations are: APR – adenosine 5'-phosphosulphate reductase; APS – ATP sulphurase; OAS-TL – *O*-acetylserine (thiol) lyase; SAT – serine acetyltransferase; SIR – sulphite reductase; SULTR – sulphate uptake and transport

1.2.2 Aspartate-Derived Backbone

The committing step towards lysine, threonine, isoleucine, and methionine synthesis is the phosphorylation of aspartate to aspartyl-β-phosphate by aspartate kinase (AK; Figure 3). This kinase is present in the chloroplast as mono- and bi-functional enzymes, each with unique catalytic and regulatory features. The mono-functional enzymes possess only AK activity and are subject to feedback inhibition by lysine, and in one report also by SAM (Rognes *et al*, 1980; Tang *et al*, 1997). The bi-functional enzymes encode both AK and homoserine dehydrogenase (HSD) activities and are subject to feedback inhibition by threonine (Paris *et al*, 2002a). The different regulation of each class of enzymes can be explained by the downstream effects of each. In both cases, the product is aspartyl-β-phosphate which is acted on by aspartate semialdehyde dehydrogenase (ASDH) to form aspartate-4-semialdehyde (Paris *et al*, 2002b). From the perspective of mono-functional enzymes there is no further conversion, leaving the product available to dihydrodipicolinate synthase (DHDPS) – the committing step towards lysine (Mazelis *et al*, 1977). In the case of the bi-functional enzymes, aspartyl-β-phosphate is converted to aspartate-4-semialdehyde by ASDH, and is then transformed to homoserine, the committing step towards methionine and threonine synthesis, through the homoserine dehydrogenase (HSD) action of the bi-functional enzyme (Paris *et al*, 2002a). Increased levels of threonine feedback inhibit the bi-functional enzyme, leaving the aspartyl-β-phosphate substrate for lysine synthesis. Feedback inhibition of AK has been exploited in agriculture as the combination of threonine and lysine is an effective herbicide due to the complete shut down of the aspartate-derived amino acid pathways at both variants of AK, resulting in methionine starvation (Tewari-Singh *et al*, 2004).

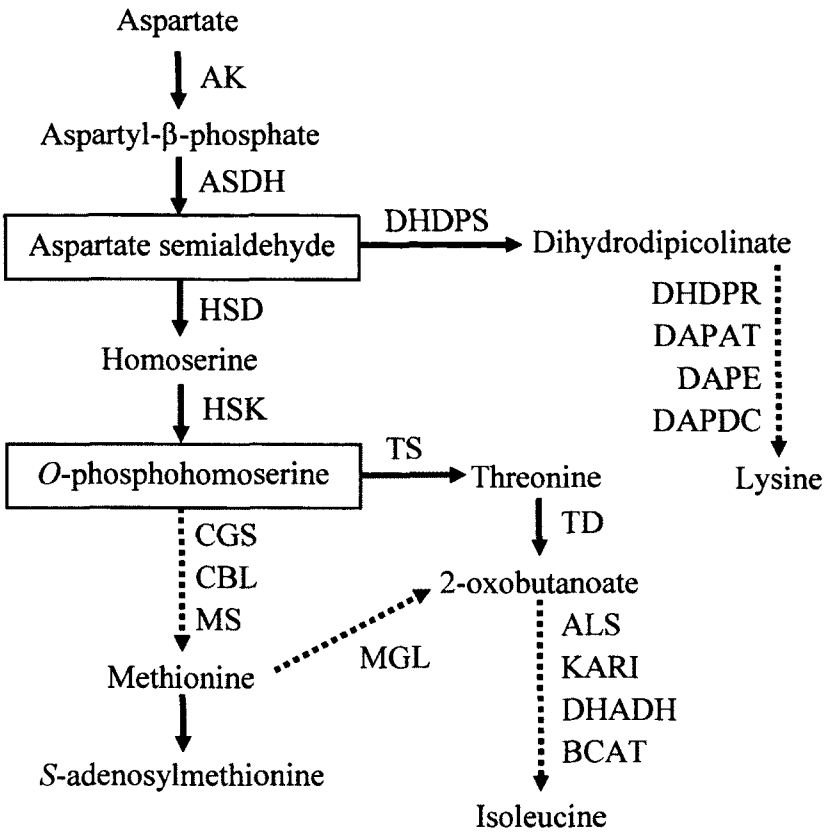


Figure 3 – Biosynthetic pathways for the aspartate-derived amino acids lysine, methionine, threonine, and isoleucine. Boxed metabolites are located at branch points in the pathway. Dashed arrows indicate multiple enzymatic conversions; solid arrows indicated single step enzymatic conversions. Enzyme abbreviations are: AK – aspartate kinase; ALS – acetolactate synthase; ASDH – aspartate semialdehyde dehydrogenase; BCAT – branched chain amino acid aminotransferase; CBL – cystathionine β-lyase; CGS – cystathionine γ-synthase; DAPAT – diaminopimelate aminotransferase; DAPDC – diaminopimelate decarboxylase; DAPE – diaminopimelate epimerase; DHADH – dihydroxy-acid dehydratase; DHDPR – dihydropicolinate reductase; DHDPS – dihydropicolinate synthase; HSD – homoserine dehydrogenase; HSK – homoserine kinase; KARI – ketol-acid reductoisomerase; MGL – methionine γ-lyase; MS – methionine synthase; TD – threonine deaminase; TS – threonine synthase

Homoserine kinase (HSK) catalyzes the phosphorylation of homoserine to *O*-phosphohomoserine (OPHS), the branch point metabolite between methionine and threonine/isoleucine biosynthesis (Figure 3; Lee and Leustek, 1999). In plants HSK is reported to be localized to the chloroplast (Wallsgrove *et al*, 1983). Some reports indicate that plant HSK is sensitive to feedback from threonine, isoleucine, valine, and SAM (radish (Baum *et al*, 1983), and pea (Thoen *et al*, 1978)); however, recombinant HSK from *A. thaliana* is reported not to be inhibited by these compounds (Lee and Leustek, 1999). Other regulation, if present, appears relaxed – over expression of HSK in *A. thaliana* did not yield any increase in the downstream metabolites (*e.g.* methionine or threonine) suggesting that the homoserine substrate was limiting, and therefore the point of control of this pathway is with this upstream metabolite. Supporting this, when supplemented with homoserine, these over expressing lines produced higher levels of both threonine and methionine (Lee *et al*, 2005).

1.2.3 Branch Points of the Aspartate-Backbone Pathway

Aspartate-4-semialdehyde is the branch point metabolite between lysine and the other aspartate-derived amino acids: methionine, threonine, and isoleucine. The committing step towards lysine – the reaction of aspartate-4-semialdehyde with pyruvate to form dihydrodipicolinic acid – is catalyzed by plastid-targeted dihydropicolinate synthase (DHDPS; Figure 3; Mazelis *et al*, 1977). DHDPS has been shown to be feedback inhibited by lysine, which, when combined with AK over-expression, resulted in increased levels of threonine but not lysine (Lee *et al*, 2005). Thus, lysine controls its own synthesis primarily through inhibition of the activity of DHDPS, coupled with

control of the mono-functional AKs. There are few reports of the remainder of the lysine metabolic pathway aside from the identification of the associated genes: dihydrodipicolinate reductase (DHDPR), diaminopimelate aminotransferase (DAPAT), diaminopimelate epimerase (DAPE), and diaminopimelate decarboxylase (DAPDC; recently reviewed in Jander and Joshi, 2010). Little additional information is available on the activity of these enzymes other than suggestions that they do not occupy regulatory positions in the lysine metabolic pathway.

O-phosphohomoserine occupies the branch point between the methionine and threonine/isoleucine metabolic pathways (Figure 3). Plastid targeted threonine synthase (TS) converts OPHS to threonine with the release of inorganic phosphate. Research focused on understanding the regulation of both methionine and threonine synthesis is discussed below (Section 1.2.6). The threonine levels in the cell appear to be primarily regulated by threonine feedback of the bifunctional AK-HSD to limit accumulation of homoserine, the substrate for threonine and methionine synthesis, balanced with the increased TS activity stimulated by SAM in the case of adequate cellular methionine status. Threonine accumulation may also be attenuated by alteration in threonine catabolic enzyme activities and diversion to isoleucine synthesis.

Isoleucine may be synthesized from 2-oxobutanoate which can be a product of either the degradation of threonine by threonine deaminase (TD) or methionine by methionine γ-lyase (MGL; Jander and Joshi, 2010; Figure 3). Conversion of 2-oxobutanoate to isoleucine is performed by the actions of acetolactate synthase (ALS), ketol-acid reductoisomerase (KARI), dihydroxy-acid dehydratase (DHADH), and branched chain aminotransferase (BCAT). Isoleucine regulates its own metabolism

through allosteric feedback inhibition of TD and ALS. Feedback regulation of MGL by isoleucine has not been investigated to date.

1.2.4 Methionine Biosynthesis

Homocysteine, synthesized by the action of the enzymes of the transsulphuration pathway, is converted to methionine by the enzyme methionine synthase (Figure 1). The committing step in methionine biosynthesis by plants, catalyzed by cystathionine γ-synthase (CGS), is the condensation of cysteine and OPHS to produce cystathionine and free phosphate. Cystathionine β-lyase (CBL) catalyzes the cleavage of cystathionine to homocysteine and pyruvate, thereby completing the transfer of the sulphur moiety of cysteine to the C4 backbone of homoserine. Both CGS and CBL have plastid targeting sequences and have been purified from chloroplasts (Wallsgrove *et al*, 1983). Reports of CBL regulation are limited and since the availability of the CBL substrate, cystathionine, is controlled by the activity of CGS, which regulates the flux of metabolites through the methionine synthetic pathway, CBL is likely not as strictly regulated as CGS (Hesse *et al*, 2004).

Methionine synthase (MS) catalyzes the transfer of a methyl group from N^5-methyltetrahydrofolate to homocysteine to yield methionine. This enzyme is present in all organisms, as it is required to regenerate methionine from homocysteine, which is produced from *S*-adenosylhomocysteine following donation of the methyl group of SAM (see Section 1.2.5 below). Mammalian MS is cobalamin dependant, while cobalamin-independent enzymes have been identified in plants (Eichel *et al*, 1995). The three plant

isoforms of MS have different sub cellular localizations, and are targeted to the cytosol, chloroplast, and mitochondria (Ravanel *et al*, 2004).

1.2.5 Metabolic Fate of Methionine

Despite the cellular requirement of methionine for protein synthesis, it is estimated that up to 80% of methionine is transformed to SAM through the action of SAM synthase (SAMS). This cytosolic enzyme family condenses methionine with adenosine triphosphate (ATP) to release SAM and free phosphate. Due to the high proportion of methionine converted to SAM, and the multiple pathways which begin with this metabolite, SAM is considered the end-point of the methionine biosynthetic pathway (Figure 4). The remainder of the methionine in the cell is used for protein synthesis or transport and storage (Amir, 2010).

Until recently it was unclear if methionine was synthesized in plants in individual tissues as required or synthesized primarily in specific locations and transported around the plant. However, research in *A. thaliana* has revealed two protein families which support the latter theory through the synthesis of *S*-methylmethionine (SMM) from methionine and SAM, and the subsequent methyl transfer from SMM to homocysteine to release two methionine molecules (Figure 4). These reactions are catalyzed by methionine methyltransferases (MMTs) and homocysteine methyltransferases (HTMs), respectively. Research suggests that this cycle is used both for methionine transport and storage as well as to balance methionine and SAM levels in the cell – essentially preventing over-production of SAM which would lead to depletion of free methionine pools (Mudd and Datko, 1990).

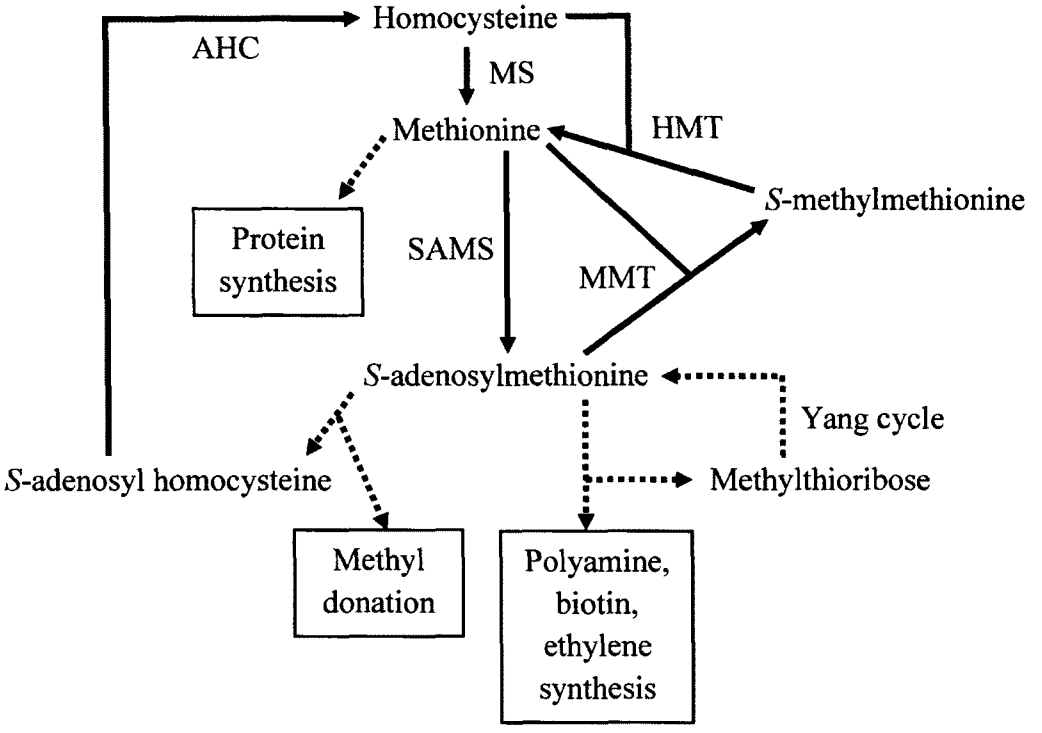


Figure 4 – Metabolic fates of methionine and *S*-adenosylmethionine in plants. Dashed arrows indicate multiple enzymatic conversions; solid arrows indicated single step enzymatic conversions. Enzyme abbreviations are: AHC – adenosyl homocysteinase; HMT – homocysteine methyltransferase; MMT – methionine methyltransferase; MS – methionine synthase; SAMS – *S*-adenosylmethionine synthase

S-adenosylmethionine is also a primary methyl group donor, making it essential for normal cell function. The action of any number of methyl transferases removes the methyl group, releasing *S*-adenosyl homocysteine (SAH), which is hydrolyzed by adenosyl homocysteinase (AHC) to release homocysteine, which re-enters the cycle as a substrate for MS. *S*-adenosylmethionine is processed though the action of various enzymes and is a precursor for polyamines, catalyzed by SAM-decarboxylase (SAMDC), or ethylene and biotin synthesis, catalyzed by ACC synthase. Both of these biosynthetic pathways release methylthioribose which is recycled back to SAM though the Yang cycle (recently reviewed in Amir, 2010).

Although it was previously thought to be absent from plants, the identification of a gene in *A. thaliana* encoding methionine γ-lyase (MGL) activity presents another metabolic fate of methionine. Methionine is degraded by MGL, releasing methanethiol, α-ketobuterate, and ammonia (Rebeille *et al*, 2006). The exact role for this degradation pathway in plants is as yet unclear, although suggestions include a role for methanethiol in the production of volatile sulphur-containing compounds, or possibly a pathway to regenerate cysteine, although there are currently no data to support this (Goyer *et al*, 2007). Recent proposals have also indicated a potential link between increased methionine and isoleucine levels in plants though the release of α-ketobuterate, by MGL, which can be converted to 2-oxobutanoate. Accordingly, as the level of methionine rises there is a corresponding increase in degradation, thereby releasing more α-ketobuterate, which flows into the isoleucine biosynthetic pathway, bypassing the main point of regulatory control at TD (Joshi and Jander, 2009).

1.2.6 Key Regulatory Points in the Methionine Biosynthetic Pathway

There are four proposed points of methionine-specific regulation in the aspartate-derived amino acid pathway (Figure 5):

1) The synergistic action of SAM to enhance lysine inhibitory feedback of the monofunctional AKs. Although the specific mechanism of this regulation remains unknown, it has been suggested that inhibition of the mono-functional AKs would result in an increased shunting of metabolites towards threonine synthesis through the bi-functional AKs when the levels of lysine and methionine (and therefore SAM) surpass demand in the cell (Curien *et al*, 2007).
2) Recently it was proposed that increased levels of lysine may influence methionine content by regulating the expression of SAMS. Research suggests that this mechanism would be most relevant in early plant development when the demand for amino acids for protein synthesis is high; resulting in an increase in lysine and methionine without a corresponding decrease in threonine (Hacham *et al*, 2007).
3) *S*-adenosylmethionine has been implicated in the feedback regulation of CGS protein levels by inhibiting CGS translation. The precise mechanism of this action, involving the MTO1 region, a highly conserved 11 amino acid stretch within the first exon of plant CGS, or an adjacent putative domain, remains under debate, although the involvement of SAM has been confirmed (Ominato *et al*, 2002, Chiba *et al*, 2003; Hacham *et al*, 2006; Onoue *et al*, 2011).

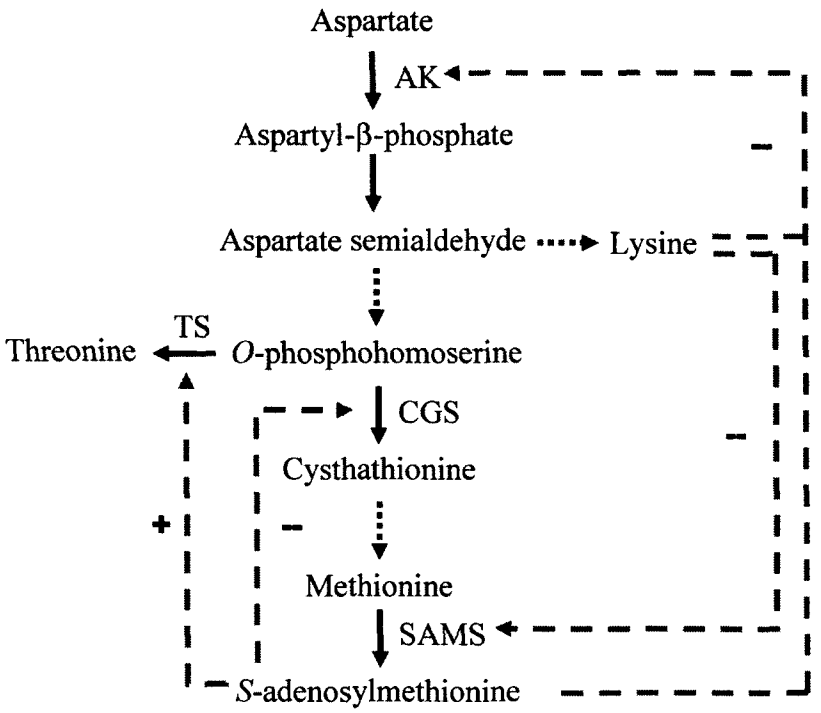


Figure 5 – Methionine-relevant regulation of the aspartate-derived amino acid pathway. Dashed grey lines indicate regulation where '+' represents positive feedback and '—' represents negative feedback. Dotted black arrows indicate multiple enzymatic conversions; solid black arrows indicated single step enzymatic conversions. Enzyme abbreviations are: AK – aspartate kinase; CGS – cystathionine γ-synthase; SAMS – *S*-adenosylmethionine synthase; TS – threonine synthase

4) *S*-adenosyl methionine has also been shown to be a potent allosteric activator of TS – increasing activity by up to 20-fold. This, coupled with the much lower Km^{OPHS} of SAM-activated TS (*e.g.* 30 μM *vs* 120 μM if not activated, in *A. thaliana*) compared to CGS (2500 μM) would increase the flux of OPHS towards threonine synthesis and away from methionine when SAM levels are sufficient (Curien *et al*, 1996; Laber *et al*, 1999).

The first two of these regulatory mechanisms have received little research attention in the context of controlling methionine levels; however, regulation of methionine homeostasis by the latter two mechanisms are strongly supported by examination of a series of methionine over-accumulating (MTO) *A. thaliana* mutant lines, each with mutations in a location which affect the functioning of either a regulatory factor or enzyme activity. Three independent types of MTO mutations have been identified in the genes encoding CGS, TS, and SAMS.

It has been demonstrated, both *in vitro* and *in vivo,* that SAM mediates the negative regulation of CGS expression via modulation of enzyme translation. In methionine fed plants or protoplasts CGS expression was down-regulated to the point of being almost undetectable (Giovanelli *et al*, 1985; Chiba *et al*, 1999). Later *in vitro* transcription/translation experiments found the methionine down-stream metabolite SAM to be the effector of this regulation (Chiba, 2003). Plant CGS sequences have an N-terminal extension (encoded by the first exon), not present in bacterial CGS sequences, which is proposed to contain both a plastid targeting sequence, and a regulatory region (Chiba *et al*, 1999; Hacham *et al*, 2002). The identification of this region was based

primarily on the MTO1 series of mutants, which are a collection of seven *A. thaliana* lines with distinct point mutations clustered in a conserved region in this N-terminal extension (Inaba *et al*, 1994; Ominato *et al*, 2002). The MTO1 mutations abolish the SAM-mediated feedback inhibition of CGS translation, thereby leading to the over accumulation of methionine and its metabolites (Ominato *et al*, 2002).

Threonine synthase competes with CGS for their common substrate OPHS, and provides another key regulatory mechanism for maintenance of the cellular methionine pool in plants. Unlike bacterial TS (*thrC*), the *A. thaliana* TS enzyme is allosterically activated by SAM, resulting in an up to 20-fold increase in k_{cat} from 0.04 to 0.86 s^{-1} (Laber *et al*, 1999). The N-terminal region of TS from plant sources, not present in bacterial TS, has been linked to SAM binding, as truncation of this region in the *A. thaliana* enzyme results loss of SAM-responsiveness (Curien *et al*, 1996). The *mto2-1* mutant of the gene encoding TS in *A. thaliana* accumulates up to 22-fold more soluble methionine than wild-type plants, and is the result of a single base pair point mutation (K204 → R) altering the ability of the enzyme to bind its PLP cofactor, thereby compromising its activity (Bartlem *et al*, 2000). The result is a decrease in the soluble threonine concentration in these plants, to as low as 6% of wild type levels. The increased levels of methionine in *mto2-1* plants were attributed to both the lack of competition for the OPHS substrate, and inability of the SAM feedback inhibition mechanism to sufficiently reduce CGS expression. The *mto2-1* plants also demonstrated severe growth abnormalities, as evidenced by poor root growth and lack of development past initial germination (Bartlem *et al*, 2000).

A third type of *A. thaliana mto* (*mto3-1* and *mto3-2*) variant, carrying mutations in the SAMS3 gene, accumulates approximately 20 – 200 fold more soluble methionine in the rosette leaves than wild type plants. The methionine content of the *mto3-1* line is increased up to 200-fold, and yet the plants demonstrate an almost wild-type growth pattern except for a slight development delay (flowering time delayed approximately 5 days), and a 21.7% decrease in lignin content. It is proposed that the mutation of the alanine at position 120 of SAMS3 to a threonine alters the ATP binding domain, such that the enzyme is no longer functional, although the exact mechanism of this loss of function is unknown (Shen *et al*, 2002). In the *mto3*-2 mutant the aspartic acid residue at position 167 is replaced with asparagine. This mutation is proposed to disrupt the substrate binding pocket, preventing the coordination of methionine (Goto *et al*, 2002). Methionine accumulation in these plants is the result of the mutations abolishing SAMS3 function, permitting the build up of methionine through the reduced conversion of methionine to SAM (Goto *et al*, 2002; Shen *et al*, 2002).

1.2.7 Enzyme and Metabolite Transport in the Methionine Biosynthetic Pathway

The majority of the methionine metabolic enzymes are chloroplast localized – from the synthesis of the upstream metabolites aspartate and serine, through the transsulphuration pathway of CGS and CBL and final step of MS (Ho *et al*, 1999; Wilkie *et al*, 1995; Wallsgrove *et al*, 1983; Ravanel *et al*, 2004). Only the subsequent conversions of methionine to SAM or SMM are localized in the cytosol. This necessitates transport of these chloroplast enzymes from the cytosolic site of synthesis into the targeted organelle location. Identification of the sub cellular localization of these

enzymes was performed through the analysis of the catalytic activities present in different sub cellular fractions. In many cases the plastid targeting sequences have been identified through Edman degradation of purified proteins, for example, the N-terminal residue of mature CGS from *Arabidopsis thaliana* is the valine at position 68 in the coding sequence (Ravanel *et al*, 1998).

Proteins can be imported into the chloroplast through a series of mechanisms, the most common of which is through the translocation at the outer envelope membrane of the chloroplast (TOC) and translocation at the inner envelope member of the chloroplast (TIC) complexes in conjunction with molecular chaperones (reviewed in Jarvis, 2008). In this pathway, proteins carrying an N-terminal plastid targeting sequence interact with chaperones and the TOC complex at the surface of the chloroplast and are fed into the inter-membrane space. If this is the final destination for this protein it is released, if not, it the TIC complex and additional chaperones direct transport further into the chloroplast. The plastid targeting sequence is cleaved immediately upon arrival at the final sub cellular location through the action of a stromal processing peptidase. The exact players and fine details of this mechanism have not yet been resolved, and further research will provide clarification on the mechanisms of intra-chloroplast targeting and differentiation between chloroplast and other organelle targeting mechanisms (Jarvis, 2008).

Like the enzymes of the pathway, transport of the pathway metabolites is critical, and yet is poorly understood. Methionine metabolism has been suggested to exist as a source-sink relationship in plants, with the majority of the synthesis occurring in the leaves followed by transport around the plant (Inaba *et al*, 1994). The initial substrates for methionine metabolism – serine and asparate – are synthesized in the chloroplast from

photosynthetically derived compounds (Ho *et al*, 1999; Wilkie *et al*, 1995). The differing sub cellular localizations of MS in plants suggest that methionine can be directly synthesized in the chloroplast, as well as from recycling pathways in the cytosol and mitochondria. This would imply that methionine transport out of the chloroplast would be required, although no transporters have been identified to date. Also, methionine uptake by plants and protoplasts support inter-cellular transport capabilities as well (Chiba *et al*, 1999).

The methionine downstream metabolites SAM and SMM are synthesized exclusively in the cytosol, thus necessitating their transport into sub cellular compartments (*e.g.* the nucleus for DNA methylation) and around the plant (*e.g.* SMM from source leaf tissues to sink seed tissues; Amir, 2010). To date a SAM transporter has been characterized from *Arabidopsis thaliana* (AtSAMT1), however no SMM transporters have been identified, despite substantial evidence for SMM transport within the plant (Bouvier *et al*, 2006; Amir, 2010). Future clarification of the transport of these methionine metabolites will provide more in depth understanding of the methionine biosynthetic pathway as a whole, and will be required for future attempts at increase methionine content in plants.

1.3 Increasing Methionine Content in Crops

Altering the flux through the biosynthetic pathways of amino acids is an attractive strategy for increasing the levels essential amino acids in plant derived foods. However, due to the stringent regulatory control of these pathways, metabolic engineering attempts have generally produced either minor increases in amino acid levels, which are either not sufficient from a human or animal dietary standpoint, or resulted in plants with detrimental or lethal phenotypes (Ufaz and Galili, 2008). An additional factor that may complicate metabolic engineering attempts is that increasing cellular concentrations of a target metabolite can trigger degradation pathways, thereby limiting accumulation (Muntz *et al*, 1998). As such, the use of strong, constitutive promoters to drive expression of modified genes in the methionine metabolic pathway may be an ineffective method towards increasing methionine. Rather, targeted expression to increase metabolites slightly without overwhelming the pathway might prove a more successful option.

Metabolic engineering of the biosynthetic pathways of amino acids derived from aspartate, including methionine and threonine, has been attempted in several plant species. Expression of an antisense construct of TS in potato resulted in a 60-200-fold increase in methionine in leaf tissue and a corresponding 45-69% reduction in threonine (Zeh *et al*, 2001). The *mto1* variants of *A. thaliana* exhibit 10 – 40-fold increases of soluble methionine, but also display developmental abnormalities, exemplifying the challenges inherent in metabolic engineering attempts to alter the amino acid profile of plants (Galili and Hofgen, 2002). The seed-specific expression of a feedback-insensitive isoform of aspartate kinase caused an increase in both free threonine (17 fold) and

methionine (3 fold) in tobacco seeds (Tabe and Higgins, 1998). The results of these experiments illustrate the challenges involved in metabolic engineering and demonstrate that the development of an increased understanding of the regulation of methionine biosynthesis is an essential precursor to the successful engineering of pulses to increase methionine concentration in a nutritionally-significant manner. In the context of pulses, where methionine is limiting, understanding the committing step towards methionine biosynthesis at CGS would be a logical first step.

1.4 Cystathionine γ-Synthase in Plants

1.4.1 Genetic Characterization

The cDNA sequences of CGS from numerous plant species have been reported and the corresponding genomic sequences are available for *Arabidopsis thaliana*, *Oryza sativa*, *Nicotiana tabaccum*, *Ostreococcus lucimarinus* and *Chlamydomonas reihardtii*. There are between 60 – 99 and 29 – 33% amino acid sequence identity among plant CGS sequences and between plant and bacterial CGS enzymes, respectively (Hesse *et al*, 2004). Initial studies of plant CGS had predicted this gene to exist as single copy per genome, supported by Southern blot analysis of CGS in an assortment of species (Table 1). However, sequencing of the *A. thaliana* genome revealed a putative second copy of CGS on chromosome 1 and, unlike the characterized CGS loci, situated on chromosome 3 of *A. thaliana*, the gene on chromosome 1 has received little attention.

Table 1 – CGS copy number in assorted plant species as reported in literature.

Species	**Ploidy**	**CGS Copy Number**	**Methodology**	**Reference**
Arabidopsis	Diploid	2	Genome Sequence	Arabidopsis Genome Initiative, 2000[a]
Arabidopsis	Diploid	1	Southern blot	Kim and Leustek, 1996
Soybean	Tetraploid[b]	2	Genome sequence	Schmutz *et al*, 2008[a]
Soybean	Tetraploid[b]	2	Southern blot	Hughes *et al*, 1999
Potato	Tetraploid	2	Genome sequence	Xu *et al*, 2011[a]
Potato	Tetraploid	Low copy	Southern blot	Riedel *et al*, 1999
Tomato	Diploid	1	Southern blot	Katz *et al*, 2006
Strawberry	Diploid	1	Southern blot	Marty *et al*, 2000
Apricot	Diploid	1	Southern blot	Marty *et al*, 2000
Peach	Diploid	1	Southern blot	Marty *et al*, 2000
Commercial strawberry	Octoploid	At least 4	Southern blot	Marty *et al*, 2000

[a] Reference provided for publication of the genome sequence
[b] Soybean is reported as a partially diplodized tetraploid

The characterized CGS (At3g01120) sequence was identified *via* its ability to complement the methionine auxotrophy of a CGS-deficient *E. coli* strain and the corresponding enzyme was subsequently characterized (Kim and Leustek, 1996; Ravanel *et al*, 1998). This is also the locus of the *mto1* mutants (Inaba *et al*, 1994). Like other reported plant CGS sequences, when aligned with bacterial CGS, the At3g01120 protein has an N-terminal extension, comprising exon 1 of the gene, which is proposed to contain a plastid targeting sequence and the MTO1 region. Removal of the entire first exon of this gene does not affect the ability of the encoded enzyme to complement a CGS-deficient *E. coli* strain (Hacham *et al*, 2006). This is the locus referred to by researchers when describing CGS from *A. thaliana*.

The second locus (At1g33320) is uncharacterized and appears in genome notation and published articles as "similar to" the characterized At3g01120 gene. The At1g33320 locus lacks the equivalent of the first exon of the At3g01120 locus, and is, therefore, similar to bacterial CGS. This locus has been proposed to be a CGS pseudogene due to the lack of expressed sequence tags (ESTs). However, this is balanced by its appearance in micro-array data sets (Winter *et al*, 2007). In one micro-array data set this locus was proposed to be significantly down-regulated in plants with reduced cysteine synthesis through quadruple T-DNA insertion mutants of serine acetyltransferase isoforms in *A. thaliana* (Watanabe *et al*, 2010).

1.4.2 Enzyme Characterization

The CGS enzyme is a homo-tetramer with one covalently bound pyridoxal 5'-phosphate (PLP) cofactor per subunit and an subunit molecular weight of 35.4 – 53 kDa, depending on the species (Hesse *et al*, 2004). The *A. thaliana* CGS (aCGS) At3g01120 cDNA comprises an open reading frame of 1689 nucleotides, encoding a polypeptide with a molecular mass of 59 kDa (Ravanel *et al*, 1998). Within the N-terminal sequence of the aCGS protein, typical of CGS from plant species, are two distinct components, a plastid (chloroplast stroma) targeting sequence and the MTO1 regulatory region, comprising amino acids 1 – 68 and 76 – 88, respectively (Ravanel *et al*, 1998; Ominato *et al*, 2002). Following transport to the chloroplast the targeting sequence is cleaved, resulting in a mature aCGS protein of 52.9-kDa subunits, with an N-terminal valine, corresponding to position 68 of the unprocessed protein (Ravanel *et* al, 1998).

In plants, CGS catalyzes the condensation of OPHS with cysteine to release cystathionine and phosphate. Arabidopsis At3g01120 has been shown to use the three homoserine derivatives (OPHS, OSHS, and OAHS) in *E. coli* complementation studies, whereas the bacterial CGSs can only use OSHS and/or OAHS, depending on the strain (Figure 1; Hacham *et al*, 2003). Kinetic parameters for this enzyme are available from several species, purified either from source or a recombinant host. The K_m of CGS for OPHS is 250 – 500 fold greater than that of SAM-activated TS (2500 μM compared to 30 μM for *A. thaliana*), demonstrating the primary role of allosteric regulation in modulation of the flux of this branch-point metabolite between the competing methionine and threonine biosynthetic pathways. There are conflicting reports on the ability of plant CGSs to perform direct sulphydrylation – the condensation of a substituted homoserine

with sulphide to form homocysteine – and whether this alternative pathway would represent a physiologically relevant route to methionine in plants. It has been reported that direct sulphydrylase activity accounts for ~3% of methionine synthesis in plants (Macnicol *et al*, 1981; Negrutiu *et al*, 1985; Hesse *et al*, 2004).

1.4.3 Regulation

The regulation of CGS in plants is post-transcriptional, with no evidence of allosteric modulation, as exemplified by the SAM-mediated activation of TS. While there is a general consensus that CGS expression in *A. thaliana* is tightly regulated, primarily *via* a post-transcriptional mechanism involving exon 1 and feedback by SAM, differing theories for this regulation have been suggested (Ominato *et al*, 2002; Kreft *et al*, 2003; Hacham *et al*, 2006; Onoue *et al*, 2011). Two models have been proposed, each involving a short sequence within the first exon of the gene encoding *A. thaliana* CGS (At3g01120):

1) The research presented by Naito *et al* identifies a highly conserved 11 amino acid stretch in exon1 as the site for feedback inhibition by SAM, through interruption of translation, referred to as the MTO1 region for the methionine over-accumulating phenotype observed in mutants of this region. The researchers propose that the methionine metabolite SAM interacts with the nascent polypeptide as it exits the ribosome during protein synthesis, inhibiting translation and leading to mRNA degradation. It is proposed that past a critical level of SAM, CGS translation is reduced and TS activity is increased, *via* allosteric activation, to shunt the OPHS substrate towards threonine synthesis (Ominato *et al* 2002; Onoue *et al*, 2011).

2) The research groups of Hesse and Amir have proposed an alternative regulatory mechanism based on their observations that not all plant CGSs appear to demonstrate the same feedback to CGS levels in the presence of SAM as those seen in aCGS (At3g01120), despite having identical MTO1 regions (*i.e.* potato; Kreft *et al*, 2003). Instead they propose that *A. thaliana* is an exception rather than an appropriate model. Western blots for *A. thaliana* CGS consistently demonstrate two bands ~3 kDa apart. The appearance of this band cannot be explained by the second CGS locus, as the predicted molecular mass of the putative gene product would be too small. Rather, the researchers have identified shortened mRNA species in some Arabidopsis tissues with internal 87 or 90 nucleotide deletions just downstream of the MTO1 region – the gene product of which could explain the shorter band observed in the Western blots. Translation of this shortened mRNA does not appear to be sensitive to methionine or SAM feedback regulation. They propose that the deleted region forms a stable hair-pin structure which is responsible for ribosomal stalling the presence of SAM, and that the removal of this sequence from the mRNA alleviates the SAM-mediated feedback regulation (Hacham *et al*, 2006).

The research group led by Naito has proposed that the MTO1 region of exon 1 of aCGS is solely responsible for the observed SAM-dependent inhibition of translation, which results in mRNA cleavage (Ominato *et al*, 2002; Chiba *et al*, 2003; Onouchi *et al*, 2005; Onoue *et al*, 2011). Recent research has confirmed that SAM binds to the nascent aCGS polypeptide as it is translated, such that the ribosome stalls at the codon corresponding to

the serine at position 94. If SAM is present at sufficient levels (1 mM in these *in vitro* studies), then the ribosome stalls and an RNase, which has not yet been identified, cleaves the mRNA upstream of the ribosome (Onouchi *et al*, 2005). Similar CGS mRNA instability has been observed for *Lemna paucicostata*; however, research in potato has shown that at least one of the two potato CGS isoforms is insensitive to SAM-induced mRNA degradation, despite the presence of a highly-conserved MTO1 region (Giovanelli *et al*, 1985; Kreft *et al*, 2003). This suggests that the MTO1 region alone is not sufficient for CGS regulation, and that other regulatory mechanisms must also exist.

In contrast, Amir and Hoefgen, maintain that other regulatory mechanism(s) exist and have suggested a regulatory role for a hairpin structure (87-90 nucleotides), which is immediately 3' of the MTO1 region in the aCGS transcript (nucleotides 296 – 386; Hacham *et al*, 2006). The effect of deleting this hairpin section remains unclear as Hacham *et al.* (2006) state that its removal abolished the SAM-mediated destabilization of the aCGS transcript, while Ominato *et al* (2002), reported no change in responsiveness to SAM. It should be noted that the different methods employed by these research groups makes the direct comparison of their results difficult – the research performed by Hacham *et al* deleted only the proposed hair-pin coding sequence, whereas Ominato *et al* did not consider the hair-pin, but found SAM mediated regulation in any construct containing the MTO1 region, and no regulation in its absence. Interestingly, the predicted hairpin may not be present to the same extent, if at all, in the transcripts of CGS from plant species other than *A. thaliana.* In contrast, the MTO1 region is conserved in all plant CGS sequences available, with the sole exception of a single amino acid difference in one of the two potato CGS isoforms (*StCGS2*) (Hesse *et al*, 2004). Investigation of *StCGS1*, which possesses a completely conserved MTO1

sequence, calls into question the primacy of this region, as this isoform is reported to be insensitive to SAM-mediated feedback (Kreft *et al*, 2003).

Regardless of the specific regulatory model, the *mto1* mutant lines show increased levels of soluble methionine coupled with slight phenotypic alterations such as reduced growth rate, delayed bolting, and decreased fresh weight at bolting. Methionine levels peak at 25 days after transplantation to soil, when they are reported to be as high as 40-fold that of wild-type rosette leaves, and then decline to approximately double that of wild type in rosette leaves (Inaba *et al*, 1994). Interestingly, the timing of the methionine decrease in these plants corresponds to the switch between the vegetative and reproductive stages of plant development, while methionine levels in wild-type plants remain fairly consistent throughout development. Additionally, the expression of CGS was not reduced by the exogenous application of methionine, as it was shown to do in the wild type (Chiba *et al*, 1999). Over-expression of CGS in *A. thaliana* under the strong constitutive promoter CaMV 35S has proven challenging, due to frequent and unstable co-suppression of transgenic and native CGS. In the absence of co-suppression, CGS over-expressing plants appear phenotypically normal; however do accumulate increased levels of methionine (8 – 15-fold) and the transport and storage molecule SMM (6 – 11-fold), but no significant increase in SAM (Gakiere *et al*, 2002).

Two independent lines with anti-sense CGS have been reported. Both lines used the At3g01120 locus as their template for the anti-sense construct and no analysis of the impact on the expression of the putative CGS at locus At1g33320 was performed. Neither line eliminated CGS expression as both possessed CGS activity, albeit at 9-20 fold reduced levels. These lines presented moderate to severe phenotypes including stunted growth, loss of apical dominance, chlorotic leaves with red/brown

discolourations, deformed flowers and siliques, lack of flowering, and failure to develop beyond cotyledon opening following germination. Interestingly, these phenotypic are also characteristic of SAM deficiency *e.g.* loss of apical dominance due to lack of DNA methylation and abnormal silique development, which is linked to reduced capacity for spermidine synthesis. Both lines reported at least partial phenotypic rescue by feeding downstream metabolites such as homocysteine, methionine, or SAM to the plants, but not by feeding upstream metabolites such as cysteine, or homoserine (Gakiere *et al*, 2000; Kim and Leustek, 2000).

1.5 Thesis Rationale – Hypotheses and Objectives

Research into methionine biosynthesis in *A. thaliana* has enhanced our understanding of this pathway and its regulation (as reviewed in Hesse *et al*, 2004; Jander and Joshi, 2010). However, despite the situation of OPHS, the substrate of CGS, at a key metabolic branch-point in the biosynthesis of methionine, information on this enzyme, in plants other than *A. thaliana*, tobacco, and potato, is very limited, and there are no reports concerning CGS from common pulses, which are deficient in methionine, from the perspective of human nutrition.

The goal of this project is to perform the initial steps in the characterization of CGS from the common pulses chickpea, field pea and lentil; crops which provide approximately $2 billion (CDN) in export revenue annually in Canada. As well, an initial investigation of the second putative CGS from *A. thaliana* (At1g33320) will be performed. This characterization will comprise:

1) Sequencing of the cDNA sequence of CGS from the target pulse species. It is hypothesized that the coding sequence of CGS from these species will resemble CGS sequences obtained from other plant species such as *A. thaliana* and *Glycine max*. The sequences will be analyzed to identify regions of interest, such as the conserved MTO1 region in the first exon and structurally and catalytically important residues, by comparison with the crystal structure of CGS from *N. tabacum*.
2) Complementation of an *E. coli* methionine auxotroph, deficient in CGS activity, to demonstrate the CGS activity of enzyme encoded by the cDNA. It is hypothesized that any clone obtained containing the above mentioned characteristics will be able to complement a methionine auxotrophy, as demonstrated by previous studies (Kim and Leustek, 1996; Hacham *et al*, 2002).

3) Investigation of CGS copy number in plant species through bioinformatics techniques to explore whether the second putative copy of CGS observed in the *A. thaliana* genome is common among plant species, or an exception specific to this species. It is hypothesized that in general CGS in plants will be found to be single copy per genome, *i.e.* that there will be one copy of CGS in diploid species, two in tetraploids, with the exception of the *Brasicaceae* family of plants.
4) Phenotypic analysis of T-DNA insertion lines in *A. thaliana* at loci At3g01120 and At1g33320 to investigate the *in vivo* role of the putative CGS at locus At1g33320. It is hypothesized that homozygous T-DNA insertion lines for the At1g33320 locus will not demonstrate a phenotype discernable from wild-type, whereas homozygous insertion lines in At3g01120 are hypothesized to be lethal, based on the severely affected phenotypes observed in CGS knockdown experiments (Gakiere *et al*, 2000; Kim and Leustek, 2000).
5) Determination of the expression pattern of CGS through the stages of plant development in the targeted pulse species through real time PCR. It is hypothesized that the expression pattern of CGS in these plants will resemble the expression pattern of the *A. thaliana* locus At3g01120 (Winter *et al*, 2007).
6) Investigation of SAM-mediated CGS regulation through performance of *in vitro* transcription and translation of exon 1::reporter gene fusions through the methods described by Naito *et al* (Chiba *et al*, 2003; Onouchi *et al*, 2005; Onoue *et al*, 2011). It is hypothesized that the exon 1 of the targeted pulse species, specifically the MTO1 region, will demonstrate sensitivity to SAM-mediated feedback when assayed in this system.

2. METHODS

2.1 Materials

Unless otherwise specified, all chemicals were from Fisher Scientific (Ottawa, Canada) or Sigma-Aldrich (Oakville, Canada). X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) was from Research Organics (Cleveland, USA). *E. coli* strain DH10B, 1kb DNA ladder, and dNTPs were purchased from Invitrogen (Carlsbad, USA). Restriction endonucleases, MLV-Reverse Transcriptase, and T4 DNA Ligase, were obtained from New England Biolabs (Ipswich, USA) and used with the appropriate manufacturer-supplied buffers. The *taq* and *pfu* polymerases were purified from expression vectors generously donated by Drs. Myron Smith and Ashkan Golshani (Carleton University, Ottawa, Canada), respectively. The *pfu* DNA polymerase reaction buffer was purchased from Stratagene (Cedar Creek, USA) and RNAse-inhibitor (SUPERas-inhibitor) was from Ambion (Streetsville, Canada). Oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, USA). Wizard Plus SV MIDI- and MINI-Prep Plasmid DNA Isolation Systems, pGEM-T Easy Cloning Vector, TNT Coupled Wheat Germ Extract System, and the pSP64-ployA vector were from Promega (Madison, USA). Glass Milk was purchased from MP Biomedicals (Solon, USA). The pGreen vector was a kind gift from Dr. Shelley Hepworth (Carleton University; Ottawa, Canada). The DNeasy Plant Maxi kit, RNeasy Plant MINI kit and on-column DNAse treatment were Qiagen products (Mississauga, Canada). The Genome Walker kit was obtained from Clonetech (Moutainview, USA) and the Illustra GFX PCR DNA and Gel Band Purification kits were from GE Healthcare (Baie d'Urfe, Canada). DNA sequencing was provided by DNA Landmarks (Sait-Jean-sur-Richelieu, Canada) and BioBasic

(Markham, Canada).

Seeds of chickpea (*Cicer arietinum* var. Vanguard), field pea (*Pisum sativum* var. Striker) and lentil (*Lens culinaris* var. Rosetown), varieties currently under cultivation in Canada, were kindly provided by Dr. T. Warkentin (U. Saskatchewan; Saskatchewan, Canada). *Arabidopsis thaliana* Columbia-0 ecotype wild type seeds were a generous gift from Dr. Shelley Hepworth (Carleton University; Ottawa, Canada). T-DNA insertion line seed stocks were purchased from the Arabidopsis Biological Resource Center (Columbus, USA). Plants were potted in the soil-vermiculite mixture BM1 Growing Mix (Berger Horticulture; Saint-Modeste, Canada). Pots were purchased from Ritchie's Feed and Seed (Ottawa, Canada) and the 20-20-20 fertilizer was purchased from Plant-Prod (Brampton, Canada).

2.2 Growth of Plants

Pulses: Seeds were surface sterilized by soaking in 95% ethanol for 5 minutes followed by distilled water for an additional 5 minutes before being planted at a density of one plant per pot and a depth of 1 cm in plastic pots (10 cm x 10 cm x 12 cm), containing a commercial soil mixture (BM1 Growing Mix, Berger Horticulture), which had been autoclaved prior to potting, and placed in an AT40 controlled environment chamber (Conviron; Winnipeg, Canada) set to photoperiod and temperature conditions of 16/8 hours (600 $\mu mol/m^2/s$ light/dark) and 22/16 °C (day/night), respectively. Plants were supplemented with 500 mL of Hoagland's nutrient solution per tray of 10 pots (Table 2) every 14 days (Hoagland and Arnon, 1950), and otherwise watered 3 times per week with 500 mL of tap water.

Table 2 - Components of Hoagland's nutrient solution and AT media

Nutrient	**Hoagland's Nutrient (μmol/L)**	**AT Media (μmol/L)**
$Ca(NO_3)_2$	10,000	2000
KNO_3	10,000	5000
KH_2PO_4	4000	2500
$MgSO_4$	2000	2000
H_3BO_3	92.5	70
$MnCl_2$	18.3	14
Fe-EDTA	18.0	49.5
$CuSO_4$	0.6	0.5
$ZnSO_4$	1.5	1.0
$NaMoO_4$	0.2	0.2
NaCl	---	10
$CoCl_2$	---	0.01

***Arabidopsis thaliana*:** Seeds were surface sterilized, as described for the pulses, before being vernalized at 4°C in the dark for 3 days on plates containing AT nutrients (Table 2) and 0.7% agar (Haughn and Somerville, 1986). These plates were subsequently transferred to an AT60 controlled environment chamber (Conviron; Winnipeg, Canada), set to continuous light at 200 μmol/m^2/s and 22°C, for 7 days. Seedlings were then transplanted to plastic pots (4 cm x 6 cm x 6 cm), at a density of one seedling per pot, containing a commercial soil mixture (BM1 Growing Mix, Berger Horticulture) which had been autoclaved before potting, fertilized with 500 mL of 1 g/L 20/20/20 fertilizer per 48-pot tray (Plant-Prod), and covered with clear plastic wrap for 7 days. The plants were fertilized again as described above when uncovered, and subsequently watered twice per week with 500 mL per tray of tap water.

2.3 RNA Isolation and cDNA Synthesis

Plant tissue (~100 mg) was harvested, placed in a 1.5-mL microcentrifuge tube and frozen immediately on dry ice. A stainless steel ball bearing and appropriate Qiagen RNA isolation buffer (RTL or RLC, depending on type of plant tissue, Table 3) were added and the tissue was homogenized for 5 minutes at room temperature, at a frequency of 1/30 sec in a tissue homogenizer (Retsch, model MM301 mixer mill). RNA was extracted using the RNase Plant Mini kit (Qiagen) according to the manufacturer's protocol, including the on-column DNAse treatment. The concentration and purity of RNA samples were determined as the absorbance at 260 nm and the 260/280 nm ratio, using a Nanodrop ND-1000 spectrophotometer (Fischer Scientific; Ottawa, Canada).

Table 3 – Type and age (after planting) of pulse species tissues and corresponding RNA extraction buffers used for RNA isolation with the Qiagen RNeasy Plant Mini kit for first strand cDNA synthesis.

Tissue	Age of Plant at Harvest	RNA Buffer
Young leaf	3 weeks, selected within the top 3 stem nodes	RLT
Mature leaf	6 weeks, selected within the 3^{rd} – 5^{th} stem nodes, prior to plant flowering	RLT
Root	3 weeks, selected the bottom 1cm of roots after removing soil and rinsing in tap water	RLT
Stem	3-6 weeks, selected within the 2^{nd} – 4^{th} stem nodes, node tissue avoided	RLT
Flower	8-9 weeks (chickpea and field pea) or 9-10 weeks (lentil), selected entire floral organ	RLT
Seed pod	9-10 weeks (chickpea and field pea) or 10-11 weeks (lentil), selected entire pod organ after flower senescence but prior to seed development	RLC
Immature seed	10-12 weeks, opened the pods and selected only small developing seeds ~3 – 6 mm (chickpea), ~4 – 6 mm (field pea), ~ 2 – 4 mm (lentil)	RLC
Full sized seed	12-13 weeks, opened the pods and selected only full grown seeds prior to onset of senescence	RLC

In preparation for the synthesis of first-strand cDNA, 1 μg total RNA was combined with 5 μM d(18)-T oligonucleotide primer and sterile, deionised water, in a total volume of 10 μL, and heated to 70°C for 5 minutes, followed by immediate incubation on ice for 2 minutes. RNAse Inhibitor (20 units; Applied Biosystems), 0.2 mM dNTPs, and reverse transcriptase reaction buffer (New England Biolabs) were added and the reaction mixture was incubated at 37°C for 5 minutes prior to the addition of 0.2 units of M-MuLV reverse transcriptase (New England BioLabs) and subsequent incubation at 42°C for 1 hour. The reverse transcriptase was heat inactivated by incubation at 70°C for 10 minutes and the resulting cDNA stored at -20°C.

2.4 *E. coli* Transformation and Colony PCR

The protocol of Inoue *et al* (1990) was modified for preparation of chemically competent *E. coli* cells. A 10 mL culture of *E. coli* strain DH10B (Invitrogen) in Luria-Bertani broth (LB) media (1% tryptone (w/v), 1% NaCl (w/v), 0.5% yeast extract (w/v)) was incubated overnight (~16 hours) at 30°C in a shaking incubator set to 200 rpm. A 2-mL aliquot of this culture was used to inoculate 1 L of LB media, which was subsequently incubated at 18°C and 200 rpm, until the $O.D._{600}$ reached 0.5 (approximately 20 hours). The 1 L culture was incubated on ice for 10 minutes prior to centrifugation in sterile bottles at 5000 rpm at 4°C for 10 minutes. The cell pellets were resuspended in 100 mL ice-cold, filter-sterilized transformation buffer (TB; 10 mM HEPES (pH 6.7), 15 mM $CaCl_2$, 250 mM KCl) and re-centrifuged as described above. This wash was repeated and the cell pellets were resuspended to a final volume of 20 mL in TB buffer containing 7% v/v glycerol and incubated on ice for 10 minutes before being

aliquoted and stored at -80°C until use. This method was also employed for the preparation of competent cells of the *E. coli* strain ER1821*CGS:aadA*, with the alteration of including 50 μg/mL spectinomycin in the liquid LB cultures (Lodha, 2011).

All *E. coli* transformations were performed *via* the heat shock method (Mandel and Higa, 1970). A 50 μL aliquot of competent cells were thawed on ice, mixed with no more than 100 ng of DNA ligation product or no more than 20 ng purified plasmid, and incubated on ice for 30 minutes. The cells were then heat shocked in a water bath at 42°C for 45 seconds and incubated on ice for 2 minutes before being combined with 1 mL LB media and incubated at 37°C, with shaking at 200 rpm, for 1 hour. Aliquots of this culture were plated on LB media plates containing 1.5% agar and antibiotic, as appropriate for the *E. coli* strain and plasmid, and incubated at 37°C for approximately 16 hours. Transformants were screened using colony PCR.

A colony of interest was picked from the plate with a sterile toothpick and mixed with sufficient sterile, deionised water for a single 50 μL PCR reaction. Cells from the same colony were also streaked on a second plate, which was incubated for 8-16 hours at 37°C. PCR reagents were added to the 200-μL, flat-cap, microcentrifuge tube containing the resuspended cells, to final concentrations of: 50 mM KCl, 10 mM Tris-HCl (pH8.5), 0.15% (v/v) Triton-X, 5% (v/v) glycerol, 2.5 mM $MgCl_2$, 240 μM dNTPS, *Taq* polymerase, and 200 nM of each primer (specific to the vector; Appendix 1). The thermocycler reaction conditions of these reactions were: an initial denaturation at 95°C for 10 minutes followed by 30 cycles of 95°C for 1 minute, 55°C for 1 minute, and 72°C for 3 minutes and a final extension at 72°C for 10 minutes. Products were visualized, *via* exposure to UV light, by electrophoresis at 140V on a 1X TAE gel (40 mM Tris-HCl, 20

mM acetic acid, 1 mM EDTA (pH 8.0) of 1% agarose (w/v) containing 0.00003% ethidium bromide (v/v)).

2.5 pGEM-T Easy Cloning

Amplicons obtained by PCR and required solely for sequencing were inserted between the T-overhangs of the pGEM-T Easy vector (Promega). This procedure exploits the natural tendency of *Taq* polymerase to extend PCR amplicons with a 3' adenine nucleotide. The pGEM-T vector has been digested with the blunt cutting restriction enzyme *EcoRV*, and subsequently treated to add thymine nucleotides to the 3' ends of the resulting linearized vector (Zhou and Gomez-Sanchez, 2000). The reaction mixture for ligation of a PCR product into the pGEM-T vector comprised: rapid ligation buffer (60 mM Tris-HCl (pH 7.8), 20 mM $MgCl_2$, 20 mM DTT, 2 mM ATP, 10% polyethylene glycol), 50 ng pGEM-T Easy vector, ~90 ng purified PCR product, and 3 units of T4 DNA ligase in a final volume of 10 μL. The reaction was incubated at room temperature for 1 hour and a 5 μL aliquot (~40-70 ng/μL) was employed for transformation of 50 μL the *E. coli* strain DH10B *via* the heat shock method, as described above. Cells were plated on LB 1.5% agar plates containing 100 μg/mL ampicillin, 80 μg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal), and 0.5 mM IPTG and incubated at 37°C for approximately 18 hours. The colour of the resulting colonies indicates either the closing of the vector with no insert in the β-galactosidase gene (blue), or the presence of an insert disrupting the β-galactosidase gene (white). At least 4 white colonies per transformation were further screened for the presence of an insertion of the expected size by colony PCR using the primers pGEM-SEQ-f and pGEM-SEQ-r

(Appendix 1), which flank the insertion site in the β-galactosidase gene. Colonies carrying plasmids with inserts of the expected size were selected for inoculation of 10 mL cultures of LB, containing 100 μg/mL ampicillin, and grown overnight (~16 hours) in a shaking incubator at 37°C and 200 rpm. Plasmids were isolated using the Promega Wizard Plus SV Miniprep kit according to the manufacturer's protocol, and sent for sequencing (DNA Landmarks or BioBasic).

2.6 Sequence Analysis

Settings for all bioinformatics tools used in this and subsequent sections are available in Appendix 2. Sequences were screened to remove vector segments using VecScreen (National Center for Biotechnology Information, 2010). Sequences obtained in the reverse orientation (*i.e.* sequencing with a reverse primer) were converted to their reverse complement using the "Reverse Complement" online tool available at The Sequence Manipulation Suite (Stothard, 2000).

When two or more primers were employed, to ensure complete coverage of insert, the resulting sequences were compiled into a single contiguous sequence, "contig", using the CAP3 Sequence Assembly program (Huang and Madan, 1999). Individual sequences were compared to the contig using the ClustalW Multiple Sequence Alignment Tool (Larkin *et al*, 2007). Any mis-matches between an individual sequence and the contig were checked manually by examination of the sequencing trace using the "Analyzed Trace View" of the Applied Biosystems "Sequence Scanner Software v1.0" (Applied Biosystems, 2005). In the case were the trace was found to be correct, the traces of any other sequences covering the base in question were also examined manually, and

adjustments to the contig made as required. Translations of the sequences were generated using the "Six-Frame Translation" tool available on Zbio.net (Zbio.net, 2006).

2.7Amplification and Sequencing of Segments of CGS from Pulse Crops

Degenerate oligonucleotide primers for the targeted coding sequences were designed based on multiple sequence alignment of a selection of available plant CGS cDNA sequences (*Arabidopsis thaliana* GenBank accession no. AF039206; *Glycine max* AF141602; *Fragaria vesca* FVAJ1451) generated using ClustalW (Larkin *et al*, 2007; Appendix 2). First strand cDNA derived from RNA isolated from young leaf tissue from each target pulse species was used as a template in combination with degenerate primers (Appendix 1) for PCR amplification of segments of the CGS coding region. The PCR components and cycling conditions were as described in Section 2.4, replacing the primers for the degenerate oligonucleotides (Appendix 1), the template with 5 μL of cDNA, a final volume of 100 μL, and increasing the extension time in the cycling from 3 to 3.5 minutes. The amplified cDNA fragments were visualized on a 1% agarose gel and fragments of expected lengths were ligated to the pGEM-T Easy vector (Promega), transformed into chemically competent DH10B cells (Invitrogen), and screened for insertion, prior to plasmid isolation and sequencing, as described above.

The sequences obtained were used to design species-specific primers for amplification of the 3' and 5' CGS cDNA segments from chickpea, lentil and field pea (Appendix 1), *via* adapter-tagged cDNA and Genome Walking, respectively. Amplification of the 3' end of the target pulse CGS cDNAs was performed using an adapter-poly-dT primer during cDNA synthesis and subsequent PCR with a

corresponding adapter PCR primer paired with a species specific primer (Figure 6; Appendix 1), with the same reaction mixture and cycling conditions as described above. PCR products were ligated to the pGEM-T Easy vector, transformed into the *E. coli* strain DH10B *via* the heat shock method, screened, and the plasmids isolated and sequenced, as described above.

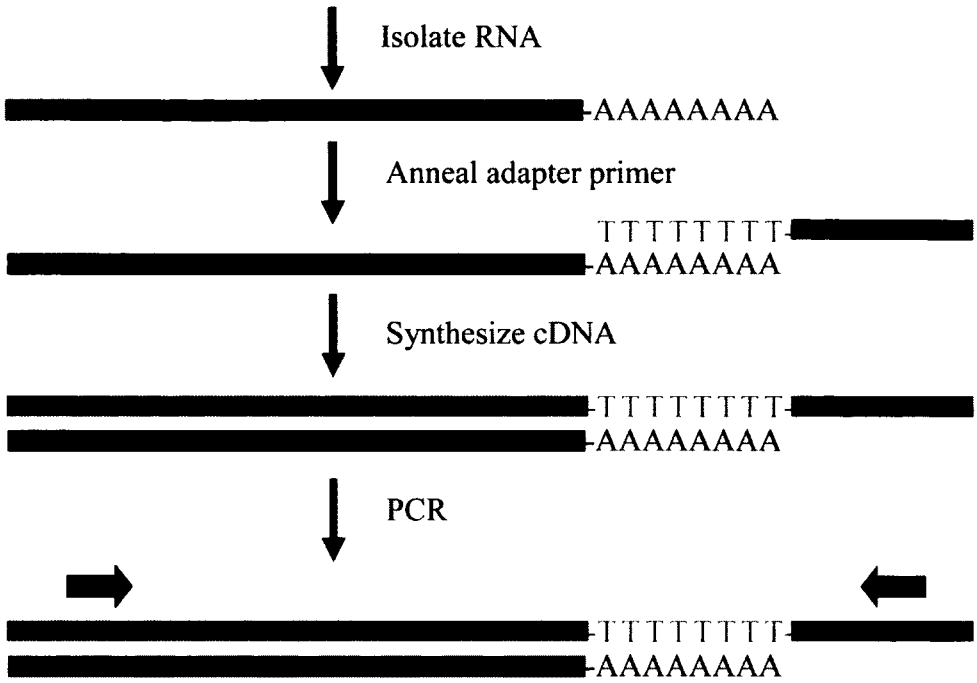


Figure 6 - Creation of adapter-tagged cDNA for amplification of the 3' end of mRNA. The oligonucleotide (red bar) primer has the sequence 5' - GTCGGATCCACCTGGAGATCGT$_{(18)}$ - 3' as the 5' portion of the adaptor-poly-dT oligonucleotide employed for first-strand cDNA synthesis. This primer is annealed to isolated RNA (black bar) in the place of the standard oligo-dT$_{(18)}$ primer. The resulting cDNA (green bar) is used as template for PCR amplification employing one primer designed specifically for the adapter sequence (red arrow) paired with a primer designed within the region of interest (black arrow). The resulting amplicon is ligated into an expression vector for sequencing.

Amplification of the 5' end of the target pulse CGS genes was performed by employing the GenomeWalker kit (Clonetech). Genomic DNA (gDNA) from each of the three pulse species was extracted from plant leaf tissue. The leaf tissue was homogenized over dry ice with a mortar and pestle. The gDNA was isolated using the DNeasy Plant Maxi kit (Qiagen), according to the manufacturer's protocol, and eluted with 1.8 mL buffer AE (10 mM Tris-HCl, 0.5 mM EDTA (pH 9.0)). The isolated genomic DNA was stored at 4°C until use.

Four Genome Walker libraries were constructed from the gDNA isolated for each of the three target pulse species, for a total of 12 libraries (Figure 7). Approximately 500 ng of gDNA from each species was digested at 37°C for 16 hours with 160 units each of four restriction endonucleases, *DraI* (TTT^AAA), *EcoRV* (GAT^ATC), *HpaI* (GTT^AAC), *ScaI* (AGT^ACT), in the appropriate manufacturer supplied buffer, with the addition of 80 μg BSA if required (New England BioLabs). These restriction endonucleases were selected for their bias towards A/T content, as plant genomes are A/T rich (Yu *et al*, 2002), and for the production of blunt ends which permit the subsequent ligation of the Genome Walker adapter segments. A sample of the digested genomic DNA was analyzed by gel electrophoresis on a 0.5% agarose (w/v) gel, stained with ethidium bromide and visualized as described above, to confirm digestion of the gDNA. Purified, digested gDNA was ligated with the Genome Walker adapter (6 μM) at 16°C overnight. Following inactivation of the ligase by incubation at 70°C for 5 minutes, the ligation product was brought to a final volume of 80 μL.

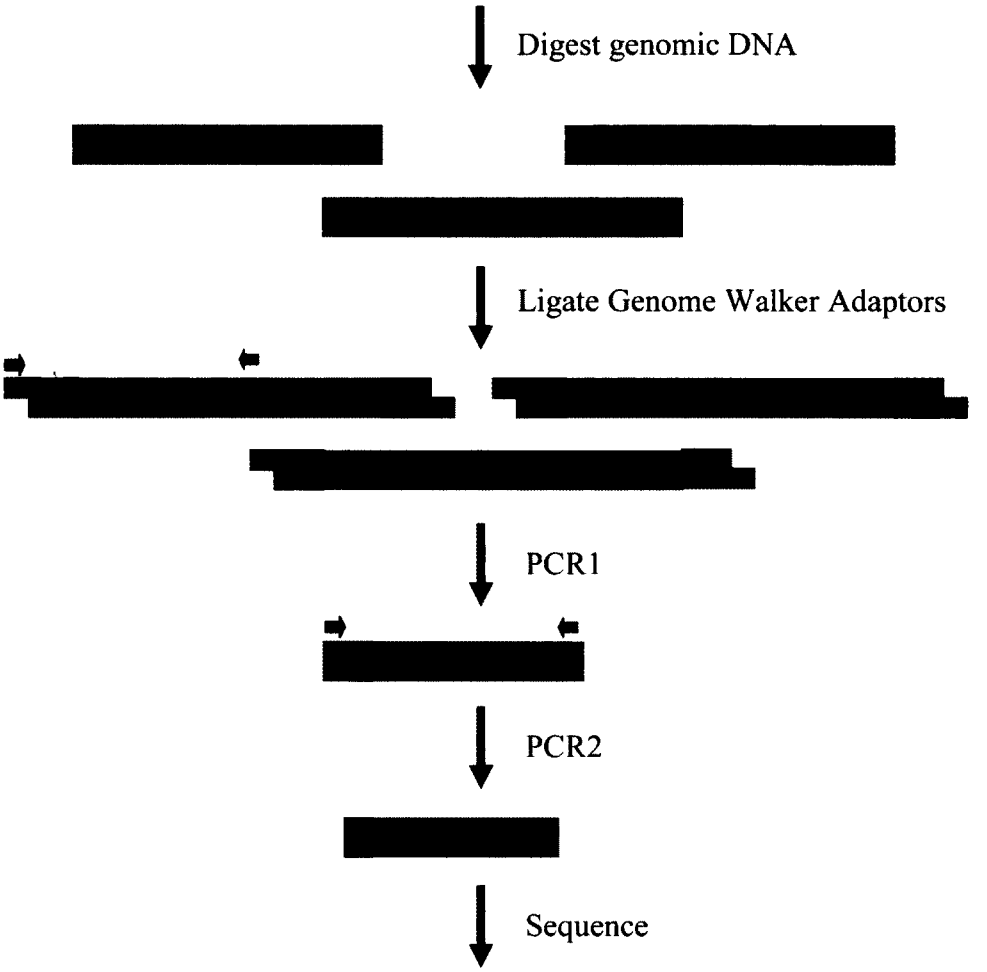


Figure 7 – Schematic of the GenomeWalking protocol (Clontech). Genomic DNA is isolated, digested into fragments (blue, green, and red bars) and Genome Walker Adapters (black boxes) are ligated to the resulting blunt ends. The first and second PCR reactions use nested oligonucleotide primers specific to the Genome Walker Adapters (black arrows) and the target sequence (blue arrows). The resulting amplicon is sequenced.

The Genome Walker libraries were used as templates for nested PCR amplifications to obtain sequence 5' of the previously amplified central coding segments for CGS from the three pulse species (Figure 7). A primary PCR was performed with the Genome Walker adapter primer 1 paired with a CGS species-specific primer within the 5' end of the known sequence. Each 50 μL PCR reaction comprised: Advantage 2 PCR buffer (40 mM Tricine-KOH (pH 8,7), 15 mM KCH_3CO_2, 3.5 mM $Mg(CH_3CO_2)_2$, 3.75 μg/mL BSA, 0.0005% Tween-20, 0.0005% Nonidet-P40), 0.2 mM dNTPs, 0.2 nM adapter primer 1 (Appendix 1), 0.2 nM species specific primer (Appendix 1), Advantage 2 Polymerase Mix and 1 μL of the library. The primary PCR reaction was carried out under the cycling conditions: initial denaturation at 94°C for 2 minutes, followed by 6 cycles of 94°C for 25 seconds and 72°C for 3 minutes, 31 cycles of 94°C for 25 seconds and 67°C for 3 minutes, and completed with a final extension at 67°C for 7 minutes.

The products of the first PCR reaction were used as a template for a secondary PCR amplification pair Genome Walker adapter primer 2 paired with a second CGS-specific primer 5' of the first for increased specificity (Figure 7). The template for the second PCR reaction was 1 μL of a 1/50 dilution of the primary PCR product. Cycling conditions for the secondary PCR: initial denaturation at 94°C for 2 minutes, followed by 4 cycles of 94°C for 25 seconds and 72°C for 3 minutes, 19 cycles of 94°C for 25 seconds and 67°C for 3 minutes, and completed with a final extension at 67°C for 7 minutes. The resulting amplicons were visualized on a 1% (w/v) agarose gel, as described above, and those greater than 500 base pairs in length were excised and purified using the GFX kit (GE Healthcare), according to the manufacturer's protocol. The purified DNA was

ligated to the pGEM-T Easy vector, transformed into the *E. coli* strain DH10B, and plasmids from successful transformants were isolated and sequenced as described above.

2.8 Cloning of Full-Length cDNAs Encoding CGS from Pulse Crops

The sequences obtained for each pulse species, from amplifications with degenerate primer or adapter-tagged cDNA and the Genome Walker cloning strategies, were aligned using the program CAP3 (Huang and Madan, 1999). Oligonucleotide primers were designed, based on these alignments, to amplify the complete coding sequence of CGS from each of the target pulse species. Primers were also designed to truncate the sequence at its 5' end to remove the sequence corresponding to the plastid targeting sequence in *Arabidopsis thaliana*, as described by Ravanel *et al* (1998). The "start" and "truncation" primers were designed to contain an *NdeI* (CA^TATG) restriction site, whereas the "stop" primers contained a *SacI* (GAGCT^C) restriction site for insertion into the expression vector pTrc-99a2 (Farsi *et al*, 2009; Appendix 1).

RNA was isolated from young leave tissue for chickpea, field pea, and lentil, and first-strand cDNA synthesis was performed as described above. The cDNA was used as a template in combination with species-specific "start" and "stop" primers to amplify the CGS sequence for cloning. The PCR reaction contained *pfu* DNA polymerase reaction buffer (20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 0.1 mg/mL BSA; Stratagene), 240 μM dNTPS, *pfu* polymerase, 200 nM of each species specific "start" and "stop" primer (Appendix 1), and 5 μL of cDNA, in a total volume of 100 μL. The cycling conditions of these reactions were: an initial denaturation at 95°C for 3 minutes, followed by 30 cycles of 95°C for 1 minute,

annealing temperature (based on primer combination) for 1 minute, and 72°C for 4 minutes, and completed with a final extension at 72°C for 10 minutes. The amplicons were visualized by electrophoresis on a 1% agarose (w/v) gel and purified using the Illustra GFX DNA PCR and Gel Band Purification kit (GE Healthcare), according to the manufacturer's protocol.

The pTrc-99a2 vector and PCR amplicons were digested with the restriction endonucleases *NdeI*/*SacI* (40 units of each endonuclease per 50 uL digest; New England BioLabs) in the manufacturer supplied buffer, at 37°C for 2 hours, followed by purification using the Illustra GFX PCR and Gel Band Purification kit (GE Healthcare). The purified, digested vector was treated with 10 units Antarctic Phosphatase (New England BioLabs) per 50 μL reaction, in the manufacturer supplied buffer at 37°C for 2 hours. The phosphatase treated vector and digested PCR amplicons were separated on a 1% agarose (w/v) gel, excised and purified using the Illustra GFX PCR and Gel Band Purification kit (GE Healthcare), according to the manufacturer's protocol.

Ligation of the digested amplicons with the linearized vector was accomplished by combining ~100 ng of purified amplicon with sufficient vector DNA to obtain a 3:1 molar ratio of insert:vector and 400 units T4 DNA ligase (New England BioLabs) in the manufacturer supplied buffer, in a total volume of 25 μL. The ligation reaction was incubated at room temperature for 4 hours before being transformed into the *E. coli* strain ER1821*CGS::aadA* as described above. Cells were plated on LB media containing 1.5% agar (w/v), 100 μg/mL ampicillin, and 50 μg/mL spectinomycin and incubated overnight at 37°C. Transformants were screened by colony PCR using the pTrc-99a specific primers pTrc-SEQ-f and pTrc-SEQ-r, which flank the multiple cloning site of the vector

(Appendix 1). Plasmids from successful transformants were isolated and sequenced as described above, using primers pTrc-SEQ-f and pTrc-SEQ-r as well as species specific internal sequencing primers (Appendix 1).

Truncations of the full-length CGS coding sequences were constructed to remove the putative plastid targeting sequence by PCR amplification. The PCR reactions contained *pfu* polymerase, the "truncation" and "stop" primers (Appendix 1) for each plant species, and 50 ng of the appropriate full-length clone plasmid as a template, as described above. The amplicons were subcloned into the pTrc-99a2 vector, the resulting transformants were screened, and the plasmids sequenced as described above.

2.9 Complementation of Methionine Auxotrophy of *E. coli* Strain ER1821*CGS::aadA*

The full length CGS coding sequence from *Arabidopsis thaliana* was amplified from the ABRC plasmid C104884 (TAIR accession no. 1005372321). Primers for the amplification of the coding region were designed such that the "start" primer contained an *NdeI* (CA^TATG) restriction site at the start codon, whereas the "stop" primer contained a *BamHI* restriction site (G^GATCC) immediately downstream of the stop codon (Appendix 1). Amplification of the coding region, ligation into the pTrc-99a2 expression vector, creation of truncations to remove the plastid targeting sequence, and sequencing of both constructs were performed as described above for the pulse species sequences (Section 2.8) with the replacement of the *SacI* restriction enzyme with *BamHI* and the cDNA PCR template with 50 ng of the C104884 plasmid.

10 mL LB media cultures containing 100 μg/mL ampicillin and 50 μg/mL spectinomycin were inoculated with *E. coli* strain ER1821*CGS::aadA* containing an empty pTrc-99a2 plasmid, pTrc-99a2 containing the *E. coli metB* gene (Farsi *et al*, 2009; Lodha, 2011), or truncated CGS from *Arabidopsis thaliana*, chickpea, field pea, or lentil. An additional 10 mL LB media culture containing 50 μg/mL spectinomycin was inoculated with the ER1821*CGS::aadA* strain containing no plasmid. These cultures were incubated overnight (~16 hours) at 30°C in a shaking incubator at 200 rpm and subsequently pelleted by centrifugation at 5000 rpm, 4°C, for 10 minutes, washed three times with 5 mL sterile 0.85% saline, and resuspended in 1 mL sterile 0.85% saline. The cultures were plated in triplicate on a series of M9 minimal media plates containing 1.5% agar (w/v), 47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% dextrose (w/v), 1 μg/mL thiamine, 100 μg/mL ampicillin, 50 μg/mL spectinomycin and either: 1) no additives, 2) IPTG (0.2 mM), 3) methionine (37.5 μg/mL). Plates were incubated at 30°C for 24-48 hours.

2.10 Bioinformatics for CGS Copy Number in Sequenced Plant Genomes

Sequenced, annotated plant genomes were queried for the presence and copy number of putative CGS sequences (Table 4). In all cases the amino acid sequences of the *Arabidopsis thaliana* characterized CGS (At3g01120) and putative CGS (At1g33320) were independently employed for searches in each genome database using the BLASTP function (Appendix 2).

Table 4 – Database information for genomes queried with the At3g01120 and At1g33320 protein sequences.

Species	Common Name	Database	Version	Website	Reference
Arabidopsis thaliana	Arabidopsis	Phytozome v7.0	TAIR10	www.phytozome.net	Arabidopsis Genome Initiative, 2000
Arabidopsis lyrata		Phytozome v7.0	v1.0	www.phytozome.net	Hu *et al*, 2011
Aquilegia coerulea	Columbine	Phytozome v7.0	Initial	www.phytozome.net	Pre-release[a]
Brachypodium distachyon		Phytozome v7.0	v1.0	www.phytozome.net	The International Brachypodium Initiative, 2010
Brassica rapa	Turnip	BrassicaDB	v1.0	http://brassica.bbsrc.ac.uk/BrassicaDB/	The Brassica rapa Genome Sequencing Project Consortium *et al*, 2011
Carica papaya	Papaya	Phytozome v7.0	Initial	www.phytozome.net	Ming *et al*, 2008
Chlamydomonas reinhardtii	Green algae	Phytozome v7.0	v4.3	www.phytozome.net	Merchant *et al*, 2007
Citrus clementina	Clementine	Phytozome v7.0	v0.9	www.phytozome.net	Pre-release[a]
Citrus sinensis	Orange	Phytozome v7.0	v1.0	www.phytozome.net	Pre-release[a]
Cucumis sativa	Cucumber	Phytozome v7.0	Initial	www.phytozome.net	Pre-release[a]
Eucalyptus grandis	Eucalyptus	Phytozome v7.0	Initial	www.phytozome.net	Pre-release[a]
Glycine max	Soybean	Phytozome v7.0	Glyma1.0	www.phytozome.net	Schmutz *et al*, 2010
Lotus japonicus		PlantGDB	Initial	http://www.plantgdb.org/	Cannon *et al*, 2007
Manihot esculenta	Cassava	Phytozome v7.0	Cassava4.1	www.phytozome.net	Pre-release[a]
Medicago trunculata	Barrel medic	Phytozome v7.0	Mt3.0	www.phytozome.net	Cannon *et al*, 2007

Species	Common Name	Database	Version	Website	Reference
Mimulus guttatus	Monkey flower	Phytozome v7.0	v1.1	www.phytozome.net	Pre-release[a]
Orzya sativa	Rice	Phytozome v7.0	MSU 6.0	www.phytozome.net	Yu *et al*, 2002
Physcomitriella patens	Moss	Phytozome v7.0	v1.6	www.phytozome.net	Rensing *et al*, 2007
Populus trichocarpa	Poplar	Phytozome v7.0	v2.0	www.phytozome.net	Tuskan *et al*, 2006
Prunus persica	Peach	Phytozome v7.0	v1.0	www.phytozome.net	Pre-release[a]
Ricinus communis	Ricin	Phytozome v7.0	v0.1	www.phytozome.net	Chan *et al*, 2010
Selaginella moellendoffii	Spikemoss	Phytozome v7.0	v1.0	www.phytozome.net	Banks *et al*, 2011
Setaria italica	Foxtail millet	Phytozome v7.0	v2.1	www.phytozome.net	Pre-release[a]
Solanum lycopersicum	Tomato	PlantGDB	Initial	http://www.plantgdb.org/	Pre-release[a]
Solanum tuberosum	Potato	PotatoGDB	Initial	www.potatogenome.net	Xu *et al*, 2011
Sorghum bicolour	Sorghum	Phytozome v7.0	Sbi1.4	www.phytozome.net	Paterson *et al*, 2009
Thellungiella halophila		Phytozome v7.0	Initial	www.phytozome.net	Pre-release[a]
Vitis vinifera	Grape	Phytozome v7.0	Initial	www.phytozome.net	Jaillon *et al*, 2007
Volvox certeri	Green algae	Phytozome v7.0	v1.0.3	www.phytozome.net	Prochnik *et al*, 2010
Zea mays	Maize	Phytozome v7.0	5a.59	www.phytozome.net	Pre-release[a]

[a] Genome sequence released ahead of publication

Results returned from the database queries were compared to known *A. thaliana* protein sequences using the BLASTP tool available at http://blast.ncbi.nlm.nih.gov/ (Appendix 2). The sequence was then assigned a putative identity based on the *A. thaliana* sequence with the best score (*i.e.* E value < e^{-100}) and in this manner it was possible to assign most sequences as similar to CGS, CBL, or MGL. However, some monocot sequences did not reliably (*i.e.* best E value score > e^{-100}) match any known *A. thaliana* sequence and, as such, were assigned as "unknown". The GenomeViewer tool, associated with each plant genome database, was used to inspect the genomic structure of each result identified as a putative CGS for the presence and number of introns. An unrooted phylogenetic tree of all of the CGS and unknown sequences was created using Mobyle Archaeopteryx (Han and Zmasek, 2009; Appendix 2)

2.11 T-DNA Insertion Lines for At1g33320 and At3g01120

The Salk Institute Genomic Analysis Laboratory T-DNA Express: Arabidopsis Gene Mapping Tool was queried for T-DNA insertion lines in the genes at loci At1g33320 and At3g01120 (Alonso *et al*, 2003). Lines with reported insertions, in or near the targeted loci (*i.e.* within 100 base pairs up stream of the translational start site through the last exon), which could be expected to interfere with expression, were ordered (Table 5; Figure 8).

Table 5 - T-DNA insertion lines ordered from The Salk Institute Genomic Analysis Laboratory, the approximate locations of the insertion in each line, and the line zygosity as reported by the supplier (Alonso *et al*, 2003).

Locus	Line	Approximate Location	Reported Zygosity
At1g33320	SALK_058208	~300 base pairs upstream of start codon	heterozygous
	SALK_128836C	Within intron 1	homozygous
At3g01120	SALK_001032	~1000 base pairs upstream of start codon	heterozygous
	SALK_045913C	Within exon 3	homozygous
	SALK_124839	Within exon 6	heterozygous
	SALK_049643	Within intron 8	heterozygous
	SALK_064727	Within exon 11	heterozygous

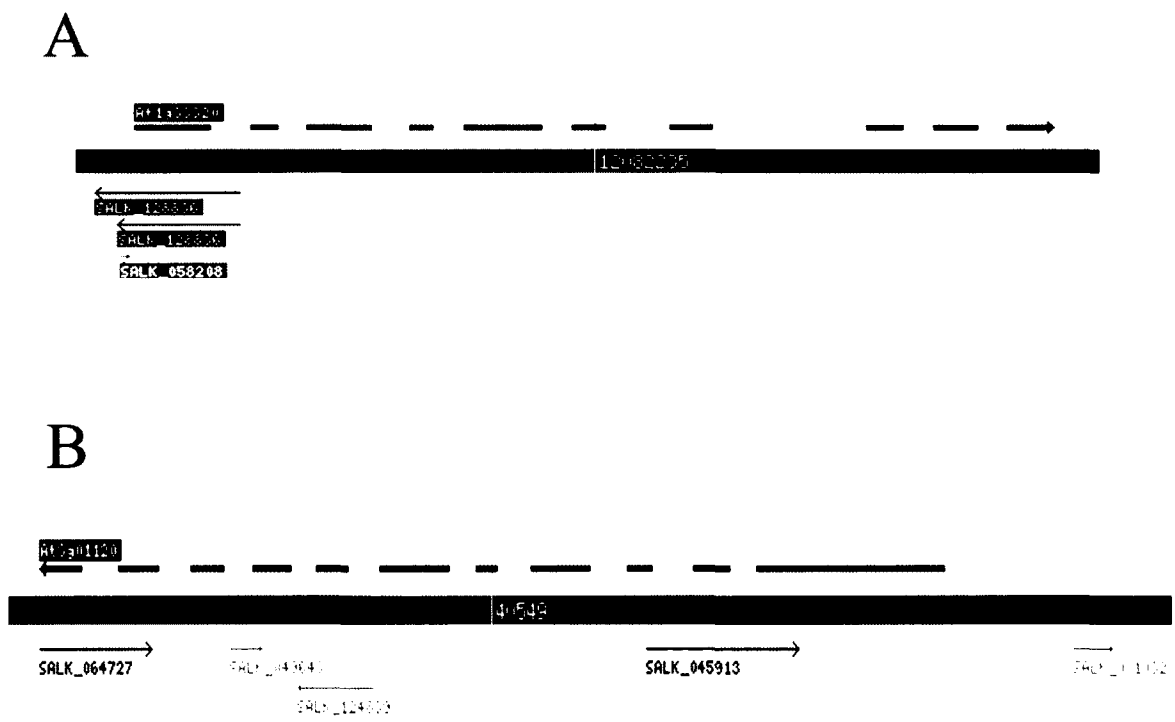


Figure 8 - Schematic of T-DNA insertion locations for *Arabidopsis thaliana* loci At1g33320 (A) and At3g01120 (B). Blue bars represent genomic DNA. Green bars above represent coding regions for the respective loci. Pink and green arrows below the genomic DNA represent the T-DNA insertions. Generated from the T-DNA Express: Arabidopsis Gene Mapping Tool (Alonso *et al*, 2003).

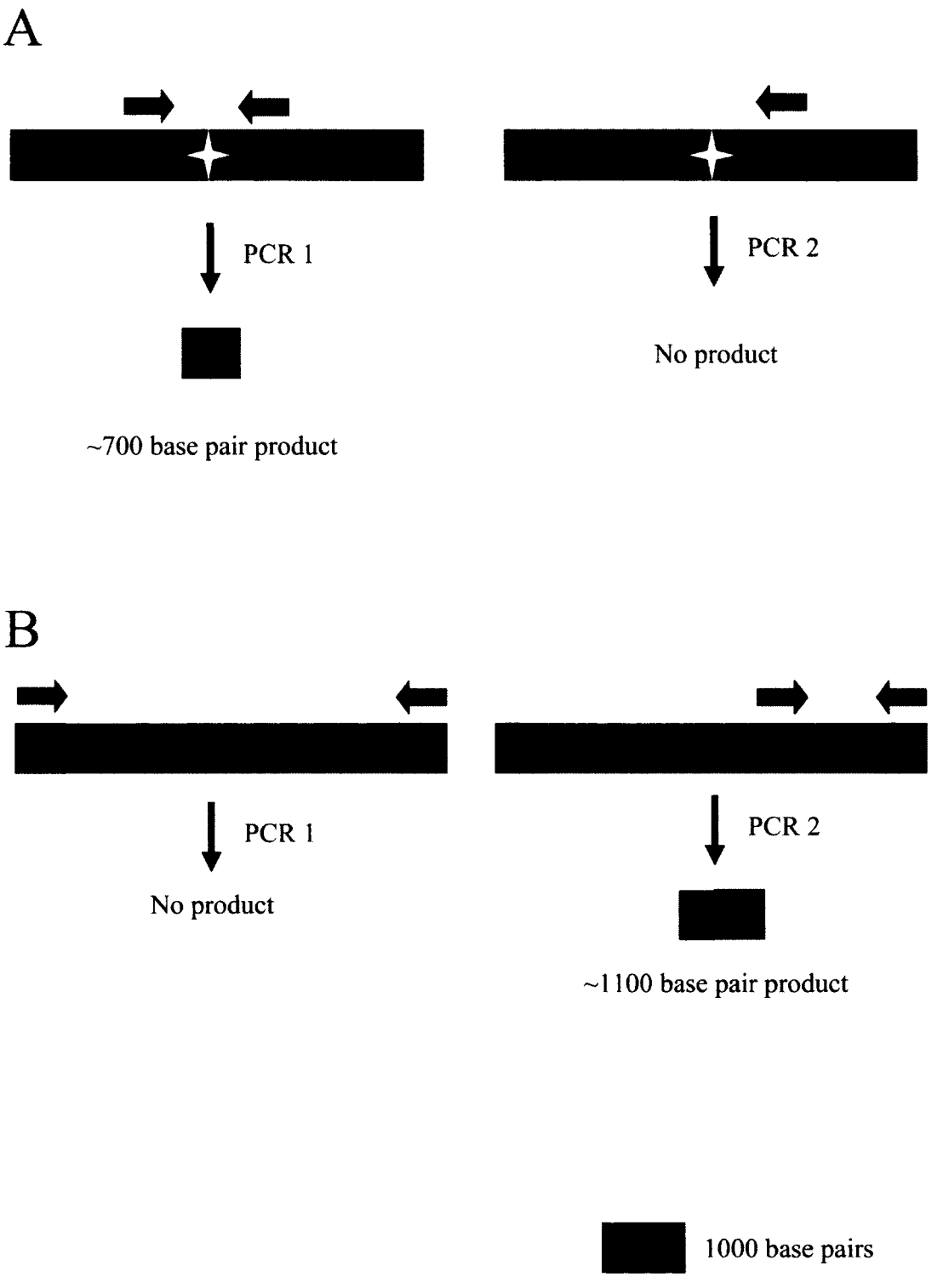
Seeds from these lines and Col-0 wild-type seeds were grown as described above. Leaves were harvested approximately four weeks after transplanting to soil, immediately frozen on dry ice in 1.5 mL microcentrifuge tubes, and stored at -80°C until genomic DNA extraction. The leaves were ground over dry ice using a microfuge tube pestle. The frozen tissue powder was combined with 400 μL of extraction buffer (200 mM Tris-HCl (pH 7.5), 250 mM NaCl, 25 mM EDTA (pH 8.0), 0.5% SDS), vortexed for 5 seconds, and centrifuged at 12,000 rpm in a micro-centrifuge for 10 minutes. The supernatant (300 μL) was transferred to a new tube, combined with 300 μL of isopropanol, and incubated at room temperature for 10 minutes, followed by centrifugation at 12,000 rpm for 10 minutes. The supernatant was discarded and the pellet washed twice with 75% ethanol:water (v/v), followed by centrifugation at 12,000 rpm for 5 minutes. The supernatant was removed and the pellet air dried before being resuspended in 100 μL sterile, deionised water. The gDNA was stored at -20°C until use.

Each plant was genotyped by analysis of PCR-generated amplicons (Figure 9). Genotyping of each individual required a pair of PCR reactions: 1) primers flanking the insertion site (L and R primers); and 2) R primer paired with an insertion-specific primer (B). Each of these amplifications was performed in the individual of interest as well as a wild-type control. The L and R primers were designed such that, in the absence of a T-DNA insertion, they should yield a product ~700 base pairs in length (depending on the specific primer pair), whereas the R and B primer combination should yield no product. In contrast, if an insertion is present, the L and R primer combination will fail to yield a band due to the in ability to fully amplify across the >4000 base pair insertion, whereas the R and B primers should yield a product of ~1100 base pairs (depending on the

specific primers). In the case of a heterozygous individual the ~700 base pair product for the L and R primer reaction and the ~1100 base pair product for the R and B primer reaction should both be observed.

Each genotyping PCR reaction contained: 50 mM KCl, 10 mM Tris-HCl (pH 8.5), 0.15% (v/v) Triton-X, 5 mM $MgCl_2$, 200 μM dNTPS, *Taq* polymerase, 500 nM of each primer (Appendix 1), 2 μL of gDNA isolation, and made to a final volume of 20 μL with sterile, deionised water. The cycling conditions of these reactions were: an initial denaturation at 95°C for 10 minutes, 35 cycles of 95°C for 1 minute, annealing (temperature based on primer combination) for 1 minute, and 72°C for 3.5 minutes, followed by a final extension at 72°C for 10 minutes. The PCR products were visualized by electrophoresis on a 1% agarose (w/v) TAE gel as described above.

Figure 9 - Schematic of the relative location of oligonucleotide primers employed for genotyping PCR reactions 1 and 2 to screen for the presence of T-DNA insertions in plants grown from seed obtained from the Salk Institute Genomic Analysis Laboratory. Brown bars represent wild-type genomic DNA, blue bars represent insertion DNA, yellow stars indicate putative insertion site, and arrows represent the L (red), R (green), and B (purple) primers. The expected results of genotyping amplifications 1 and 2 on (A) wild-type DNA (*i.e.* no insertion) and (B) genomic DNA containing a T-DNA insert. Scale bar represents 1000 base pairs.


A
PCR 1
PCR 2
No product
~700 base pair product
B
PCR 1
PCR 2
No product
~1100 base pair product
1000 base pairs

2.12 Expression Analysis of CGS in Pulse Species by Real Time PCR

Primers for real time PCR expression analysis of CGS and Elongation Factor 1α (EF1α) from chickpea, field pea and lentil were designed using the Integrated DNA Technologies' web tool "Primer Quest" and the sequences obtained from amplification of segments of the coding sequence with degenerate oligonucleotide primers (Rozen and Skaletsky, 1996-1998; Appendix 1, 2).

Total RNA was isolated from a set of at least three distinct plants for each of the tissues sampled (young leaf, mature leaf, root, stem, flower, immature pod, immature seed and mature seed; Table 3) for each of the three pulse species, the RNA samples were pooled, and first strand cDNA synthesis was performed as described above. Pooled RNA was also used to create the "no reverse transcriptase" (noRT) control, for which cDNA synthesis is performed replacing the reverse transcriptase enzyme with water. Each real time PCR experiment contained a standard curve comprising four dilutions of a (10 ng/μL – 0.01 ng/μL of the *E. coli metB* gene in the pTrc99a vector), experimental reactions with cDNA template and gene-specific primers, and control reactions (Table 6). Each experiment was performed in triplicate. The efficiency of amplification was also determined for each gene of interest with cDNA derived from a young leaf source for each species in a run including the same dilution series of the standard vector, four ten-fold dilutions of the cDNA template, and negative control reactions (Table 6).

Table 6 - Experimental set up for real time PCR expression analysis of CGS in the target pulse species chickpea, field pea, and lentil.

Experiment	Primers	Reaction	Replicates	Template
Experimental	CGS or EF1α	Experimental	Quadruplicate	1:2 diluted cDNA in water
		Negative control	Triplicate	1:2 diluted noRT control in water
		Negative control	Triplicate	Water
	Standard curve	Standard curve x4	Quadruplicate	Diluted vector
		Negative control	Triplicate	Water
Efficiency	CGS or EF1α	Experimental x4	Quadruplicate	Diluted young leaf cDNA
		Negative control	Triplicate	1:2 diluted noRT control
		Negative control	Triplicate	Water
	Standard curve	Standard curve x4	Quadruplicate	Diluted vector
		Negative control	Triplicate	Water

Each 25 µL reaction was comprised of: 50 mM KCl, 10 mM Tris-HCl (pH8.5), 0.15% (v/v) Triton-X, 2.5 mM $MgCl_2$, 200 µM dNTPS, *Taq* polymerase, 304 nM of each primer (Appendix 1), SYBR GreenI (1/50,000 final dilution from supplier stock; Sigma-Alrich), and 5 µL template (Table 6). Thermal cycling conditions for the reactions were: initial denaturation at 95°C for 10 minutes, followed by 45 cycles at 95°C 20 seconds, 60°C for 20 seconds, 72°C for 30 seconds, then 80°C for 15 seconds. A fluorescence measurement was made after each amplification cycle, following the incubation at 80°C. Reaction specificity was confirmed by performing a melt curve analysis following the final cycle during which the temperature was increased in 1°C increments from 72°C – 95°C, with a 15 second incubation and a fluorescence reading following each increment. The progress of the reactions and melt curve were monitored with a Rotor-Gene 6000 thermocycler (Corbett BioScience; Montreal, Canada).

Data was analyzed using the Rotor-Gene 6000 software 1.7 (Build 87). The melt curve was analyzed to produce a graph of dF/dT *vs* temperature (°C), and manually inspected for peaks representing desired amplicons. These peaks were expected in reactions containing cDNA or vector, but not in reactions containing noRT cDNA (indicative of genomic DNA contamination) or reactions with no template (indicative of reaction contamination). The vector dilution series was used to automatically calculate the critical threshold between 0 and 0.1 relative fluorescence. The critical threshold values for each of the experimental quadruplicate reactions were averaged. The efficiency of the amplifications were calculated automatically from the efficiency experiments with the threshold determined as described above.

The critical threshold levels of CGS from each tissue was normalized to EF1a, and set relative to the expression in the young leaf tissue using Equation [1]:

$$\text{Ratio} = \frac{(E_{CGS})^{\Delta CP(CGS)(YL-\text{tissue})}}{(E_{EF1\alpha})^{\Delta CP(EF1\alpha)(YL-\text{tissue})}} \quad [1]$$

where E represents the efficiency of the amplification (as determined by the efficiency experiment), as specified for either CGS or EF1α, and ΔCP represents the difference in averaged critical threshold values between the young leaf and the tissue of interest for CGS or EF1α (Pfaffl, 2001). The relative expression ratio was calculated for each of the triplicate experimental runs for each tissue, averaged and the standard deviation calculated. The JMP9 statistical analysis software package from SAS was used to determine statistical significance between tissues, by ANOVA.

2.13 Amplification of the Exon 1/2 Boundary of CGS from Pulse Species

Genomic DNA was isolated from leaf tissue from each of the targeted pulse species using the Qiagen DNA Mini Plant kit (Qiagen) as described above. Species specific primers (Appendix 1) were used to amplify over the predicted exon1/2 boundary. The PCR reactions contained: 50 mM KCl, 10 mM Tris-HCl (pH8.5), 0.15% (v/v) Triton-X, 2.5 mM $MgCl_2$, 240 μM dNTPS, *Taq* polymerase, 200 nM of each of a species specific forward and reverse primer (Appendix 1), and 1 μL of gDNA. The cycling conditions of these reactions were: an initial denaturation at 95°C for 3 minutes, 30 cycles of 95°C for 1 minute, annealing temperature (based on primer combination) for 1 minute, and 72°C for 4 minutes, followed by a final extension at 72°C for 10 minutes. The amplified cDNA fragments were visualized by agarose gel electrophoresis, as described

above. The PCR products were inserted between the T-overhangs of the pGEM-T Easy vector, screened, and the plasmids isolated and sequenced.

2.14 Constructs for Luciferase Assay for Post-Transcriptional SAM Regulation

Three luciferase-containing constructs were required: 1) luciferase with an upstream Kozak sequence; 2) luciferase with an upstream Kozak sequence and the amino acid sequence MAVS fused in-frame with the start codon to provide a control containing the same translational start sequence as the plant constructs; and 3) luciferase with no upstream modifications to be later fused in frame with the plant CGS exon 1 sequences of interest. Primers for amplification of the luciferase gene from the pGreen plasmid were designed to contain restriction endonuclease sites for sub-cloning into the pSP64 Poly(A) vector (Promega; Appendix 1). Primers were designed such that the "start" primer for the Kozak-luciferase and Kozak-MAVS-luciferase constructs contained an *XbaI* site (T^CTAGA), an *XmaI* site (C^CCGGG) for the luciferase construct to be fused to plant exon 1 sequences, and the luciferase "stop" primer contained a *SacI* site (GAGCT^C).

An LB plate containing 1.5% agar and 100 μg/mL ampicillin was streaked with *E. coli* harbouring the pGreen vector, from a glycerol stock received from Dr. Shelley Hepworth (Carleton University), and incubated at 37°C for approximately 16 hours. A single colony was selected from this plate and used to inoculate a 10 mL LB liquid culture containing 100 μg/mL ampicillin, which was incubated overnight at 37°C with shaking at 200 rpm. The plasmid was isolated using the Wizard MINIPrep plasmid isolation kit, following the manufacturer's protocol. The luciferase coding sequence was

amplified from the isolated pGreen vector by PCR using *pfu* polymerase and the appropriate oligonucleotide primer pair for each construct (Appendix 1; Section 2.9).

Amplicons of predicted sizes were purified by glass milk according to the manufacturer's protocol (MP Biomedicals). The PCR product was combined with an 600 µL of 6 M NaI. Following the addition of 5 µL of glass milk, the mixture was vortexed for 30 seconds and incubated on ice for 15 minutes. The mixture was centrifuged at 12,000 rpm for 15 minutes, and the supernatant discarded. The pellet was washed twice in ice-cold 500 µL of New Wash solution (50 mM NaCl, 10 mM Tris-HCl (pH 7.6), 25 mM EDTA), before elution in 50 µL of sterile, deionised water at 55°C for 3 minutes. The solution containing the DNA was separated from the glass milk by centrifugation at 12,000 rpm in a micro-centrifuge and the supernatant was transferred to a new tube. Purified DNA was digested with restriction endonucleases as designated by the sites engineered in the oligonucleotide primers, in the appropriate manufacturer supplied buffer, at 37°C for 2 hours. The digestion products were separated on a 1% agarose (w/v) TAE gel, excised, melted at 60°C in 600 µL 6 M NaI, and purified using glass milk, as described above.

The pSP64 Poly(A) vector (Promega) was transformed into the *E. coli* strain ER1821*CGS::aadA* and plated on LB plates containing 1.5% agar and 100 µg/mL ampicillin, followed by incubation overnight at 37°C. A single colony was selected and used to inoculate a 100 mL LB liquid culture containing 100 µg/mL ampicillin, followed by incubation overnight at 37°C with shaking at 200 rpm. The cells were pelleted and the plasmid isolated by the Wizard MIDIPrep Plasmid Isolation System (Promega). The vector was digested with *XbaI/SacI* or *XmaI/SacI* (New England BioLabs) in

manufacturer supplied buffers at 37°C for 2 hours. The digestion products were purified by the glass milk method described above, and then treated with Antarctic Phosphatase (New England BioLabs) in the manufacturer supplied buffer for 2 hours at 37°C. The phosphatase- treated product was purified from a 1% agarose (w/v) TAE gel using glass milk. Insertion of the luciferase inserts into the pSP64 vector linearized was performed, followed by transformation into the *E. coli* strain ER1821*CGS::aadA*. Cells were plated on LB media containing 1.5% agar (w/v), 100 μg/mL ampicillin, and 50 μg/mL spectinomycin and incubated overnight at 37°C. Plasmids from successful transformants were isolated and sequenced using the pSP64-SEQ and M13- primers (Appendix 1).

A 100 mL LB liquid culture containing 100 μg/mL ampicillin, and 50 μg/mL spectinomycin was inoculated with the ER1821*CGS::aadA* cells harbouring the construct containing the luciferase gene with the 5' *XmaI* site from a glycerol stock. The culture was incubated overnight at 37°C with shaking at 200 rpm and the plasmid was isolated using the Wizard MIDIPrep Plasmid Isolation System (Promega). The vector was digested with *XbaI/XmaI*, purified by glass milk, treated with Antarctic Phosphatase and purified from a 1% agarose (w/v) TAE gel, as described above.

Exon 1 of each plant CGS was amplified, using *pfu* polymerase, from plasmid stocks using species specific primers (Appendix 1), engineered to contain the recognition sequences for the restriction endonucleases *XbaI* (T^CTAGA) and *XmaI* (C^CCGGG) for the KZ-Exon1-F and Exon1-R primers respectively, with 50 ng of the appropriate plasmid as a template. Amplicons were purified by the glass milk method, digested with *XbaI* and *XmaI* restriction endonucleases (New England BioLabs) and ligated into the digested and phosphatased pSP64-luciferase vector, followed by transformation into

ER1821*CGS::aadA*. Transformants were selected and the plasmids isolated and sequenced with primers pSP64-SEQ and pLUC-SEQ as described above.

Overlap-extension site-directed mutagenesis was employed to engineer the equivalent of the MTO1-1 mutation into each of the exon1:luciferase fusion constructs. The primer pair for mutagenesis were designed over the MTO1 region altering the GGT coding for glycine to AGT for serine for each of the *Arabidopsis thaliana* and pulse CGS sequences (Appendix 1). In separate reactions, using the sequenced wild-type exon1-luciferase as a template, the species-specific KZ-Exon1-F primer was paired with the appropriate MTO1-R primer and the species specific MTO1-F primer with the appropriate Exon1-R primer for amplification by *pfu* polymerase. Amplicons were purified by glass milk (MP Biomedicals) and used as template for the reassembly PCR with the KZ-Exon1-F and Exon1-R primers and 3 μL of each purified mutagenesis PCR product as template. The cycling conditions of these reactions were: an initial denaturation at 95°C for 5 minutes, annealing at 50°C for 1 minute, extension at 72°C for 10 minutes, followed by 30 cycles of 95°C for 1 minute, annealing temperature (based on primer combination) for 1 minute, and 72°C for 4 minutes, and a final extension at 72°C for 10 minutes. Amplicons of the expected size were excised from a 1% agarose gel, purified using glass milk, digested with the restriction endonucleases *XbaI* and *XmaI* (New England BioLabs) and ligated into the digested, purified, pSP64-luciferase vector, followed by transformation into the ER1821*CGS::aadA* cell line. Plasmids isolated and sequenced with primers pSP64-SEQ and pLUC-SEQ as described above.

The luciferase control and plant exon1::luciferase fusion vectors, as well as the positive control vector from the kit were transcribed and translated *in vitro* using the TNT

Quick Coupled Transcription/Translation System (Promega) in half the volume recommended by the manufacturer's protocol. Each 25 μL reaction contained 12.5 μL wheat germ extract, 1 μL reaction buffer, 0.5 μL polymerase, 0.5 μL RNAse inhibitor, 0.25 μL each amino acid mix minus methionine and minus leucine, and 0.5 μg purified plasmid DNA. The reaction was incubated at 30°C for 90 minutes. A 2.5 μL aliquot of the reaction was combined with 50 μL of luciferase assay reagent in a black 96-well plate and the fluorescence read using a FLUOstar Optima fluorometer (BMG Labtech).

3. RESULTS

3.1 Sequencing of CGS from Chickpea, Field Pea, and Lentil

Amplification of segments of CGS from chickpea, field pea, and lentil was performed using degenerate primers located in areas of plant CGS nucleotide sequences that are highly conserved between species (Appendix 1). Sequencing of the resulting two overlapping amplicons demonstrated coverage of 87% of the coding sequence of CGS, based on soybean sequence (Figure 10). The percent identity of the nucleotide sequences were determined using the online tool ClustalW2 – Multiple Sequence Alignment Tool (Appendix 2; Larkin *et al*, 2007) in comparison to the *A. thaliana* locus At3g01120, tobacco, and soybean isoform 1 CGS sequences available from GenBank (Table 7; accession numbers U83500, AF097180 and AB035300, and AF141602 respectively). These plant sequences were selected for comparison because the CGS enzyme from *A. thaliana* has been characterized (Ravanel *et al*, 1998; Kusano *et al*, 2010), there is a crystal structure available for the tobacco CGS enzyme (Steegborn *et al*, 1999), and the close phylogenetic relationship between soybean and the target pulse species (*i.e.* all are members of the *Fabaceae* family). The first reported tobacco mRNA sequence (AF097180) has since been supplemented with additional 5' sequence (AB035300; Ominato *et al*, 2002). Soybean is an alloploid, and as such has two copies of CGS. The sequence referred to here as soybean isoform 1 is the first identified CGS sequence from soybean cDNA. A second was identified later during genome sequencing. The high nucleotide sequence identity (63.0 - 85.0%) between the *A. thaliana*, tobacco and soybean sequences and the sequences obtained from degenerate primer amplification suggest that these amplicons correspond to CGS coding sequences.

Figure 10 – Fragments (A and B) of CGS amplified from chickpea, field pea, and lentil by degenerate primer PCR (orange bars) are compared to the full-length soybean CGS sequence (GenBank accession AF141602; purple bar). Start and stop codons and the location of the MTO1 regulatory region are indicated on the soybean sequence by green, red, and blue arrows respectively. Scale bar (black) is 250 base pairs.

Table 7 – Nucleotide sequence identity of fragments of the CGS coding sequences amplified from chickpea, field pea, and lentil by degenerate primer PCR with the corresponding nucleotide sequences from *A. thaliana*, tobacco, and soybean.

Fragment ID	***A. thaliana*[a] (%)**	**Tobacco[b] (%)**	**Soybean[c] (%)**
Chickpea A	75.0	70.0	85.0
Chickpea B	69.0	69.0	80.0
Field pea A	77.0	71.0	85.0
Field pea B	63.0	65.0	74.0
Lentil A	75.0	71.0	84.0
Lentil B	69.0	70.0	78.0

[a] GenBank accession no. U83500
[b] GenBank accession nos. AF097180 and AB035300
[c] GenBank accession no. AF141602

Adapter-tagged cDNA was used as a template for amplification of the 3' ends of the CGS mRNA using the adapter primer in combination with species-specific primers for chickpea, field pea, and lentil (Figure 6). Each amplicon overlapped the coding region sequence obtained by degenerate primer cloning by at least 58 base pairs, thereby allowing for alignment with the previously obtained sequences (Table 8). The amplicons also spanned the 3' untranslated region (3'-UTR) of the cDNA, revealing this segment to be 143 – 193 base pairs in length, depending on the species (Table 8).

Genome Walking was used to obtain species-specific sequences for the 5' end of the CGS coding regions, including the start codon, for the three target species (Figure 7). The sequences obtained ranged from 627 – 754 base pairs in length and overlapped the coding sequence segments by at least 75 base pairs. A putative start codon was identified for each species, and the presence of the MTO1 regulatory region was identified from translation of the nucleotide sequence of each amplicon (Table 9).

Table 8 – Characteristics of amplicons obtained from PCR amplification of the 3' ends of chickpea, field pea, and lentil CGS using species specific internal primers combined with the adapter primer on adapter-tagged cDNA template.

Species	Total Length (base pairs)	Overlap with Known Sequence (base pairs)	Length Between Stop Codon and Poly-A Tail (base pairs)
Chickpea	226	58	143
Field Pea	277	65	193
Lentil	670	458	189

Table 9 – Sequences obtained with GenomeWalker kit (Clontech; Figure 7) for 5' sequence of CGS from chickpea, field pea, and lentil.

Species	Library	Length (base pairs)	Overlap with Known Sequence (base pairs)	Putative Start Codon[a]	Location of MTO1 coding sequence[b]
Chickpea	*DraI*	734	75	459-461	651-686
	ScaI	688	128	357-359	549-584
Field Pea	*DraI*	754	153	371-373	567-604
Lentil	*HpaI*	627	256	165-167	363-398

[a] Putative base pair location within fragment, based on translational open reading frame in frame with sequences obtained through degenerate primer amplification and sequencing and comparison to soybean isoform 1 (GenBank accession no. AF141602)

[b] Putative base pair location within fragment, based on comparison of translated sequence beginning at the putative start codon to MTO1 regulatory sequence as identified from *A. thaliana* (Ominato *et al*, 2002)

Expression vectors containing full length CGS from the chickpea, field pea, and lentil were constructed by amplification of the coding sequence from cDNA and insertion into the pTrc-99a2 vector (Farsi *et al*, 2009). The full-length CGS sequences (Appendix 3) are 74.0 – 85.0% and 70.0 – 85.0% identical in nucleotide and amino acid sequence to CGS from *Arabidopsis thaliana*, tobacco, and soybean (accession numbers U83500, AF097180 and AB035300, and AF141602 respectively; Table 10) when compared using the online tool ClustalW2 – Multiple Sequence Alignment Tool (Larkin *et al*, 2007). The sequences demonstrate non-synonymous: synonymous mutation (Ka/Ks) ratios of 0.2, as calculated by the Ka/Ks Calculation Tool (Bergen Center for Computational Science, 2011), indicating that there is selective pressure to conserve the sequence. This ratio compares well with other large scale comparisons, for example, Ka/Ks rations between human and rodent genes average ~0.2 (Siltberg and Liberles, 2002). A multiple sequence alignment of the chickpea, field pea, lentil, *A. thaliana*, and tobacco amino acid sequences was analyzed using the web-based tool BOXSHADE to identify regions of similarity (Figure 11). The pulse species contain conserved regions of interest, particularly around *A. thaliana* plastid targeting sequencing truncation site, MTO1 region, exon 1/2 boundary, and active site lysine.

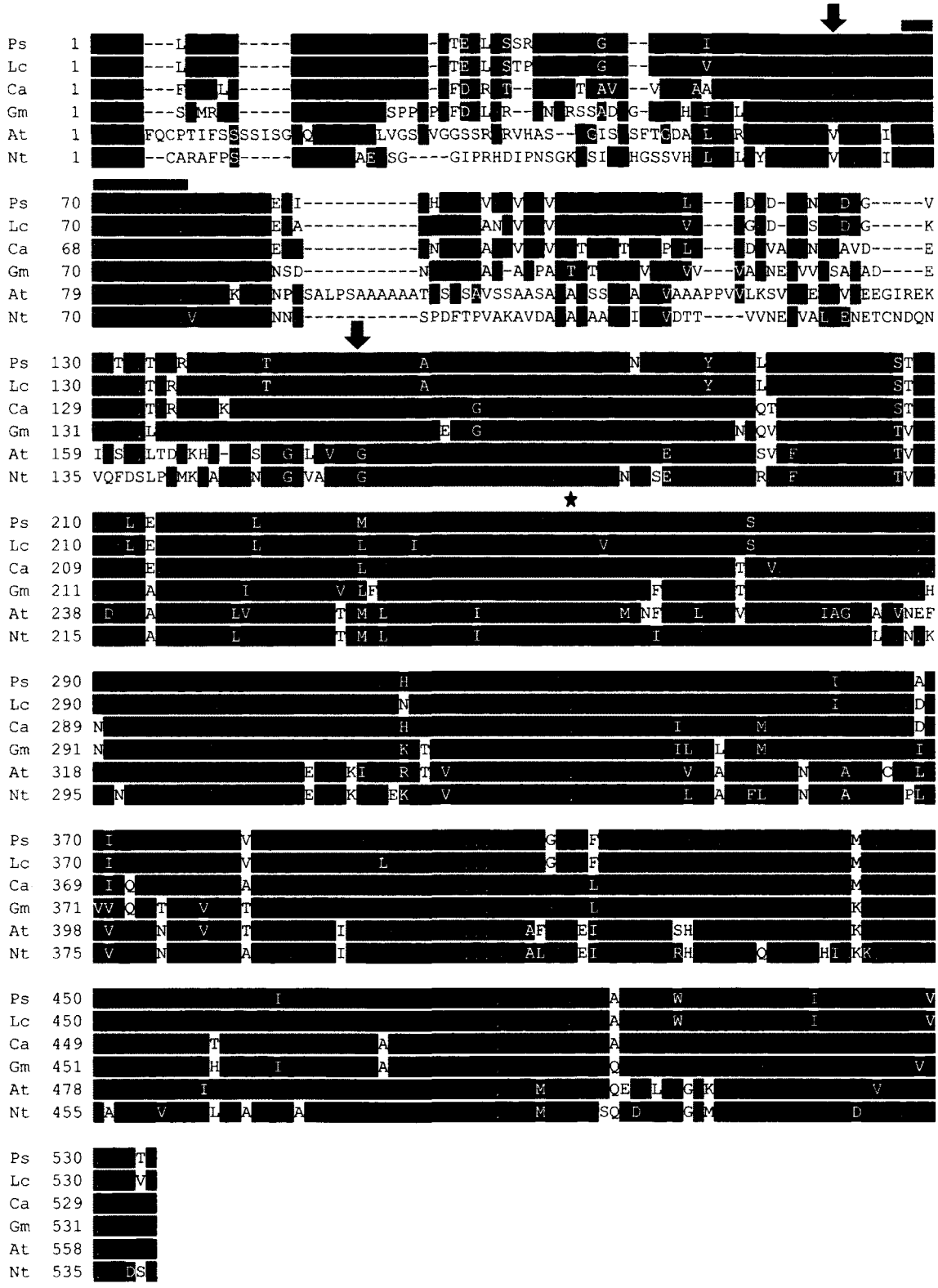
The truncation site for the plastid-targeting sequence in *A. thaliana* is identified in Figure 11 (Ravanel *et al*, 1998). Oligonucleotide primers were designed to truncate the chickpea, field pea, lentil, and *A. thaliana* CGS sequences and the resulting amplicons were inserted into the pTrc-99a2 expression vector. Sequencing of these constructs produced identical sequences to the full-length constructs with the exception of the introduced start codon at the truncation site.

Table 10 - Comparison of nucleotide and amino acid sequence identity between the CGS sequences of chickpea, field pea, and lentil, with the corresponding sequences of *Arabidopsis thaliana*.

Species	Sequence	*A. thaliana*[a] (%)	Tobacco[b] (%)	Soybean[c] (%)
Chickpea	Nucleotide	75.0	74.0	85.0
	Amino acid	71.0	70.0	85.0
Field Pea	Nucleotide	76.0	74.0	84.0
	Amino acid	70.0	69.0	81.0
Lentil	Nucleotide	74.0	73.0	84.0
	Amino acid	70.0	68.0	81.0

[a] GenBank accession no. U83500
[b] GenBank accession nos. AF097180 and AB035300
[c] GenBank accession no. AF141602

Figure 11 – Deduced amino acid sequence alignment, highlighted with BoxShade, of CGS from field pea (Ps), lentil (Lc), chickpea (Ca), soybean isoform 1 (Gm; GenBank accession number AF141602), Arabidopsis locus At3g01120 (At; GenBank accession number U83500) and tobacco (Nt; GenBank accession number AF097180 and AB035300). Green arrow indicates the cleavage location for the plastid targeting sequence in *A. thaliana*; blue bar indicates MTO1 region; red arrow indicates the plant exon 1/2 boundary; purple star indicates the active site lysine bound to the PLP cofactor.


Ps 1
Lc 1
Ca 1
Gm 1
At 1
Nt 1
Ps 70
Lc 70
Ca 68
Gm 70
At 79
Nt 70
Ps 130
Lc 130
Ca 129
Gm 131
At 159
Nt 135
Ps 210
Lc 210
Ca 209
Gm 211
At 238
Nt 215
Ps 290
Lc 290
Ca 289
Gm 291
At 318
Nt 295
Ps 370
Lc 370
Ca 369
Gm 371
At 398
Nt 375
Ps 450
Lc 450
Ca 449
Gm 451
At 478
Nt 455
Ps 530
Lc 530
Ca 529
Gm 531
At 558
Nt 535

3.2 Complementation of Methionine Auxotrophy of the *E. coli* Strain ER1821*CGS::aadA*

To confirm that the activity conferred by the coding sequences obtained from chickpea, field pea, and lentil cDNA is that of CGS, the truncated variants, in the expression vector pTrc-99a2, were assayed for their ability to complement the methionine auxotrophy of the *E. coli* strain ER1821*CGS::aadA*, in which *metB* has been replaced with a gene encoding *adenylyltransferase* (*aadA*), conferring resistance to spectinomycin. The methionine auxotrophy of this strain is such that growth on minimal media will only be supported in the presence of a transgene conferring CGS activity (Lodha, 2011). Constructs containing the chickpea, field pea, and lentil truncated coding sequences, as well as control constructs for truncated *A. thaliana* CGS, lacking the plastid targeting sequence, *E. coli metB*, and an empty expression vector, were individually transformed into the ER1821*CGS::aadA* strain. Transformants were replica plated on M9 minimal media with 1) no additives, 2) 0.2 mM IPTG or 3) 37.5 μg/mL methionine.

The "cells only" control cultures (*i.e.* containing no vector) demonstrated no growth on any plates after 48 hours, due to a lack of resistance to ampicillin, conferred by the expression vector pTrc-99a2 (Farsi *et al*, 2009; Figure 12, Table 11). All other cultures demonstrated growth on the methionine containing plates within 24 hours of incubation. Cells containing the empty expression plasmid pTrc-99a2 could not complement the methionine auxotrophy, regardless of the presence of IPTG, as demonstrated by the lack of growth even after 48 hours incubation. This is expected as this construct contains no coding sequence conferring CGS activity. The *E. coli metB* construct demonstrated growth on the M9 minimal media in the absence of IPTG within

24 hours, but not in the presence of IPTG after 48 hours, as previously observed (Farsi *et al*, 2009; Lodha, 2011). The CGS constructs corresponding to the four plant species tested all complemented the methionine auxotrophy of the *metB*-deficient strain to allow growth within 48 hours when plated on M9 media containing IPTG (Figure 12, Table 11). This was expected for the *A. thaliana* construct, as its ability to rescue *metB*-deficient strains has been demonstrated (Kim and Leustek, 1996; Hacham *et al*, 2002). Successful complementation of the methionine auxotrophy of the *E. coli* strain ER1821*CGS::aadA* by the truncated chickpea, field pea, and lentil constructs confirms that the enzymes encoded by these sequences possess CGS activity.

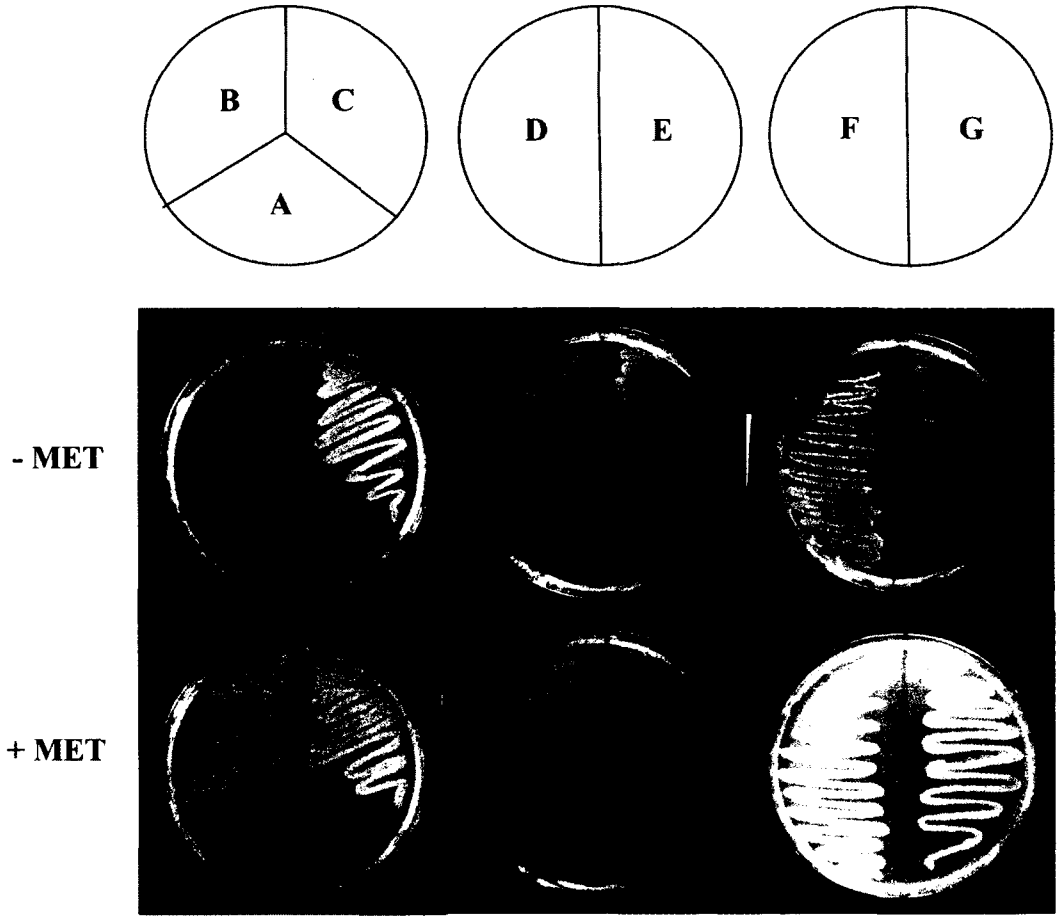


Figure 12 – Complementation of the *E. coli* methionine auxotrophic strain ER1821*CGS::aadA* on M9 minimal media by A – cells without plasmid control; B – empty plasmid control; and with plasmid containing C – *E. coli metB*; D – truncated CGS from *A. thaliana*; E – truncated CGS from chickpea; F – truncated CGS from field pea; or G – truncated CGS from lentil. The top row of plates contain no methionine, the bottom row contains 37.5 μg/mL methionine. Plates were incubated at 30°C for 24-48 hours (Table 11).

Table 11 – Incubation times for complementation of the methionine auxotrophic *E. coli* strain ER1821*CGS::aad*.

Plasmid Insert	M9	M9 + 0.2 mM IPTG	M9 + 37.5 µg/mL MET
No plasmid	48 hours – no observable complementation	48 hours – no observable complementation	48 hours – no observable complementation
No insert	48 hours – no observable complementation	48 hours – no observable complementation	24 hours – complete complementation[a]
E. coli metB	24 hours – complete complementation	48 hours – limited complementation[b]	24 hours – complete complementation
A. thaliana CGS1[c]	48 hours – limited complementation[b]	48 hours – complete complementation	24 hours – complete complementation
Chickpea CGS	48 hours – limited complementation[b]	48 hours – complete complementation	24 hours – complete complementation
Field pea CGS	48 hours – limited complementation[b]	48 hours – complete complementation	24 hours – complete complementation
Lentil CGS	48 hours – limited complementation[b]	48 hours – complete complementation	24 hours – complete complementation

[a] Complete complementation indicates that colony growth was dense

[b] Limited complementation indicates that some colony growth was visible following 48 hour incubation, however, this growth was minimal, with only a few small, scattered colonies.

[c] *A. thaliana* CGS1 sequence from the ABRC plasmid C104884

3.3 Bioinformatics for CGS Copy Number in Sequenced Plant Genomes

The 30 annotated plant genomes available as publicly accessible databases were queried with the translated sequences of the *A. thaliana* loci At3g01120 and At1g33320 coding regions (Table 4; Appendix 2). All identified sequences of interest were subsequently compared to the *A. thaliana* proteome to assign putative function as CGS, CBL, MGL, or unknown. In general, each plant species contained single copy each per genome of CGS and CBL, and up to four copies per genome MGL, with the exception of three members of the *Brassicaceae* family: *A. thaliana, T. halophila* and *B. rapa.* The diploid species *A. thaliana* and *T. halophila* each contain two CGS-like sequences, while the paleoalloploid *B. rapa* possesses three copies (Table 12). All sequences assigned as putative CGS were submitted to the online tool Moybyle Archaeopteryx (Appendix 2; Han and Zmasek, 2009) for the creation of an unrooted phylogenetic tree (Figure 13).

Most plants contained a CGS with an MTO1 coding region with the exception of the algae species *V. certeri* and *C. reinhardtii*, which do not contain this sequence (green branch, Figure 13). Some sequences, specifically the second *Z. mays*, second *S. bicolour*, *R. communis*, and *S. moellendoffii* do not contain an MTO1 region as reported, although translation based on a translation initiation in the mRNA sequence 5' of the reported initiation site, this region would be present (Table 12; red stars, Figure 13). The potato sequence containing some unresolved sequence in the region where the MTO1 sequence would be expected (green star, Figure 13). Some of the *Brassicacaea* family members contained an additional CGS copy per genome (At1g33320 locus in *A. thaliana*, one of the two identified loci in *T. halophila,* and one of the three identified loci in *B. rapa*) which did not contain MTO1 regions (orange box in Figure 13).

The monocot species also contain sequences, at single to multi-copy per genome, which could not be assigned a putative classification as CGS, CBL, or MGL by comparison to the *A. thaliana* proteome. These sequences cluster distinctly from the putative CGS sequences (yellow branch, Figure 13), do not contain the MTO1 regulatory region sequence, and possess a distinct gene structure (Figure 14). With the exception of the sequences from algae species, which are organized as a single exon, the coding sequence of putative plant CGS genes is spread over 11 exons (Figure 14; Table 12). The unidentified monocot genes all include 2 – 4 exons (Figure 14).

Chickpea, field pea, and lentil are members of the *Fabaceae* family of plants. Although complete genome sequences are unavailable for these species, complete genome sequences have been reported for three other species of the *Fabaceae* family: soybean (*Glycine max*), medic (*Medicago trunculata*), and *Lotus japonicus*. All three of these species appear to have a single CGS copy per genome, the standard 11 exon genomic sequence organization, and the MTO1 regulatory region within exon 1 (light blue branch in Figure 13, Table 12).

Table 12 – Putative plant CGS sequences from genome database mining.[a]

Species	Ploidy	# CGS	Genome Identifier[b]	Figure 13 Identifier[c]	MTO1 Region?	Genome Structure
Arabidopsis thaliana	Diploid	2	At1g33320	Athal1	No	Standard[d]
			A3g01120	Athal3	Yes	Standard
Arabidopsis lyrata	Diploid	1	477417	Alyra	Yes	Standard
Aquilegia coerulea	Diploid	1	AcoGoldSmith_v1.002452m.g	Acoer	Yes	Standard
Brachypodium distachyon	Diploid	2	Bradi1g61260	Bdist1	Yes	Standard
			Bradi1g69730	Bdist2	No	Monocot
Brassica rapa	Diploid[e]	3	Bra001020	Brapa1	Yes	Standard
			Bra039144	Brapa2	Yes	Standard
			Bra008703	Brapa3	No	Standard[d]
Carica papaya	Diploid	1	evm.TU.contig_28082.1	Cpapa	Yes	Standard
Chlamydomonas reinhardtii	Diploid	1	Cre60.g792150	Crein	No	Algae
Citrus clementina	Diploid	1	clementine0.9_006855m.g	Cclem	Yes	Standard
Citrus sinensis	Diploid	1	orange1.1g009523m	Csine	Yes	Standard
Cucumis sativa	Diploid	1	Cucsa.338600	Csati	Yes	Standard
Eucalyptus grandis	Diploid	1	Eucgr.I01959	Egran	Yes	Standard
Glycine max	Tetraploid[f]	2	Glyma09g37020	Gmax1	Yes	Standard
			Glyma18g49660	Gmax2	Yes	Standard
Lotus japonicus	Diploid	1	chr1.CM0141.340.nc	Ljapo	Yes	Standard
Manihot esculenta	Alloploid	2	cassava4.1_005058m.g	Mescu1	Yes	Standard
			cassava4.1_004490m.g	Mescu2	Yes	Standard

Species	Ploidy	# CGS	Genome Identifier[b]	Figure 13 Identifier[c]	MTO1 Region?	Genome Structure
Medicago trunculata	Diploid	1	Medtr7g010560	Mtrun	Yes	Standard
Mimulus guttatus	Tetraploid	2	mgv1a002982m.g	Mgutt1	Yes	Standard
			mgv1a004543m.g	Mgutt2	Yes	Standard
Orzya sativa	Diploid	4	LOC_Os03g25940	Osati1	Yes	Standard
			LOC_Os10g25950	Osati2	No	Monocot
			LOC_Os10g26010	Osati3	No	Monocot
			LOC_Os06g13450	Osati4	No	Monocot
Physcomitirella patens	Tetraploid	2	Pp1s49_246V6	Ppate1	Yes	Standard
			Pp1s152_35V6	Ppate2	Yes	Standard
Populus trichocarpa	Diploid	1	POPTR_0017s12240	Ptric	Yes	Standard
Prunus persica	Diploid	1	ppa004232m.g	Ppers	Yes	Standard
Ricinus communis	Diploid	1	29635.t000001	Rcomm	Yes[g]	Standard
Selaginella moellendoffii	Diploid	1	143363	Smoel	Yes[g]	Standard
Setaria italica	Diploid	2	Si035253m.g	Sital1	Yes	Standard
			Si039843m.g	Sital2	No	Monocot
Solanum lycopersicum	Diploid	1	40806078	Slyco	Yes	Standard
Solanum tuberosum	Tetraploid	2	PGSC0003DMT400064705	Stube1	N/A[h]	Standard
			PGSC0003DMT400021841	Stube2	Yes	Standard
Sorghum bicolour	Diploid	2	Sb01g023050	Sbico1	No	Monocot
			Sb01g034130	Sbico2	Yes[g]	Standard
Thellungiella halophila	Diploid	2	Thhalv10020394m.g	Thalo1	Yes	Standard
			Thhalv10013649m.g	Thalo2	No	Standard[d]
Vitis vinifera	Diploid	1	GSVIVG01011450001	Vvini	Yes	Standard

Species	Ploidy	# CGS	Genome Identifier[b]	Figure 13 Identifier[c]	MTO1 Region?	Genome Structure
Volvox certeri	Diploid	1	75793	Vcert	No	Algae
Zea mays	Tetraploid	4	GRMZM2G113873	Zmays1	Yes	Standard
			GRMZM2G075153	Zmays2	Yes[g]	Standard
			GRMZM2G320085	Zmays3	No	Monocot
			GRMZM2G056418	Zmays4	No	Monocot

[a] Sequences were identified by BLASTP of both *A. thaliana* CGS loci (At3g01120 and At1g33320) against the available annotated plant proteomes.

[b] Genome Identifier – the unique identification assigned to each sequence in the specific database. Searching the database specified in Table 4 with this number will provide the complete record of the locus including sequence information.

[c] Figure 13 Identifier – the unique identification assigned to each sequence in reference to Figure 13.

[d] The genome structure of these sequences resembles the Standard 11 exon, 10 intron structure, however the equivalent of exon 1 is absent from the sequence, resulting in a 10 exon, 9 intron structure.

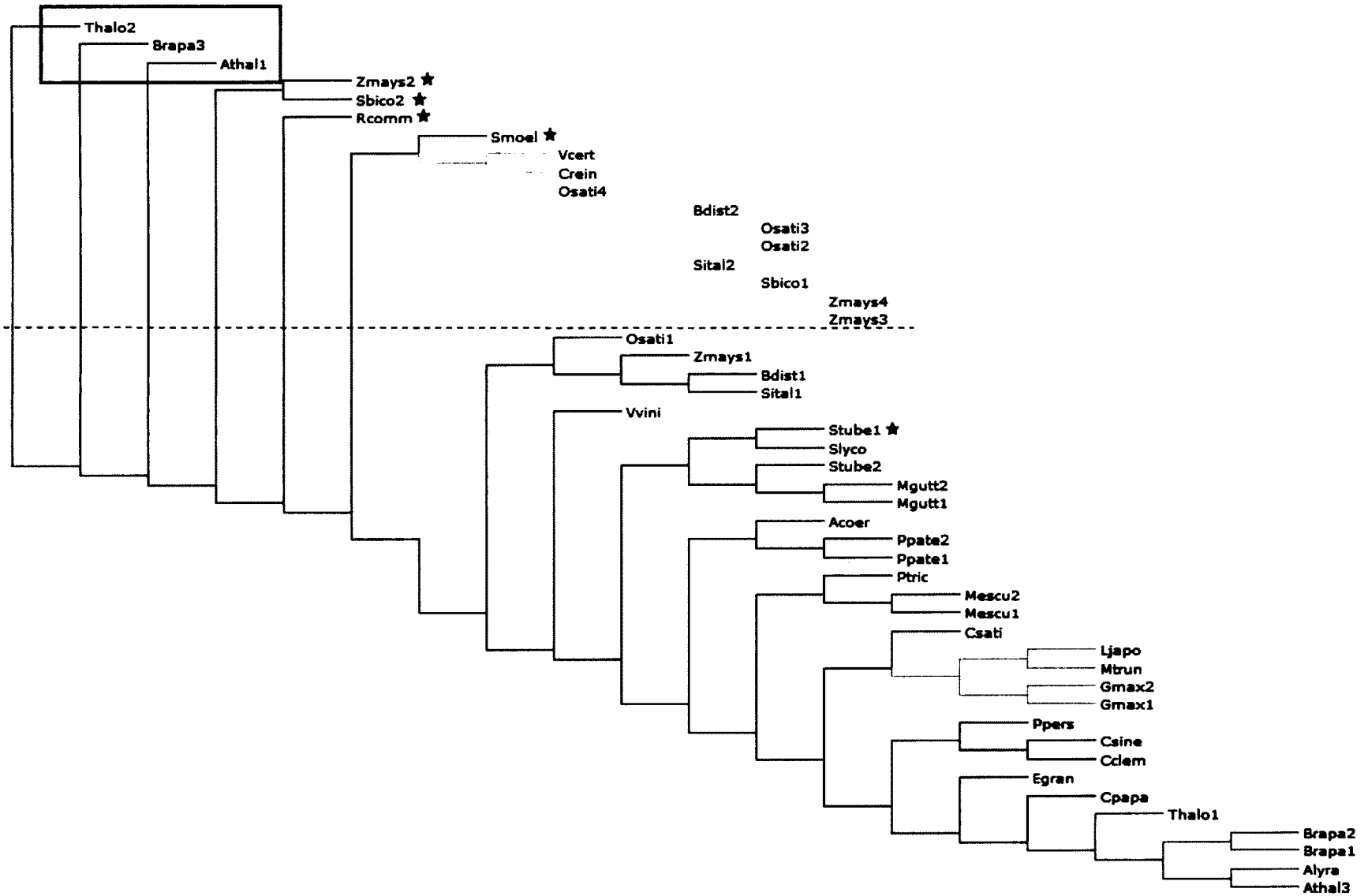
[e] The genome of *B. rapa* is reported as a diploid, but recognized as being a paleoalloploid.

[f] The genome of *G. max* is reported as a partially diplodized tetraploid.

[g] The protein sequence reported appears truncated. When the complete mRNA is translated, an additional upstream translation initiation codon can be identified. Translation from this codon results in a polypeptide sequence containing the MTO1 region.

[h] The genome sequence reported for exon 1 for this transcript contains a region for which the sequence has not been completely resolved (*i.e.* is reported as NN...NN). Based on the sequence available surrounding this area, this region encodes part of exon 1 of a CGS; however, until the sequence is resolved the presence of MTO1 sequence in this region will remain unclear.

Figure 13 – Phylogenetic tree for CGS and CGS-like sequences from annotated plant genomes. Annotations for each sequence are presented in Table 12 above. The yellow branch highlights those monocot sequences designated as "unknown", where the dark blue branch highlights the monocot sequences designated as CGS. All sequences above the dashed line do not contain the MTO1 region. The orange box surrounds those sequences in the *Brassicaceae* family which have the 10 exon, 9 intron structure (*i.e.* are missing the sequences equivalent to exon1), whereas the purple branch highlights the *Brasicaceae* family CGS sequences with the standard genomic structure. The green branch represents the algae sequences. Red stars indicate those sequences that appear truncated, yet have the sequences that would code for the MTO1 region upstream in the predicted mRNA sequence. The green star indicates the incomplete potato sequence in which some exon1 sequence has not been resolved. The blue branch highlights the *Fabaceae* family CGS sequences. Tree generated by the online tool Mobyle Archaeopteryx (Han and Zmasek, 2009).


Thalo2
Brapa3
Athal1
Zmays2
Sbico2
Rcomm
Smoel
Vcert
Crein
Osati4
Bdist2
Osati3
Osati2
Sital2
Sbico1
Zmays4
Zmays3
Osati1
Zmays1
Bdist1
Sital1
Vvini
Stube1
Slyco
Stube2
Mgutt2
Mgutt1
Acoer
Ppate2
Ppate1
Ptric
Mescu2
Mescu1
Csati
Ljapo
Mtrun
Gmax2
Gmax1
Ppers
Csine
Cclem
Egran
Cpapa
Thalo1
Brapa2
Brapa1
Alyra
Athal3

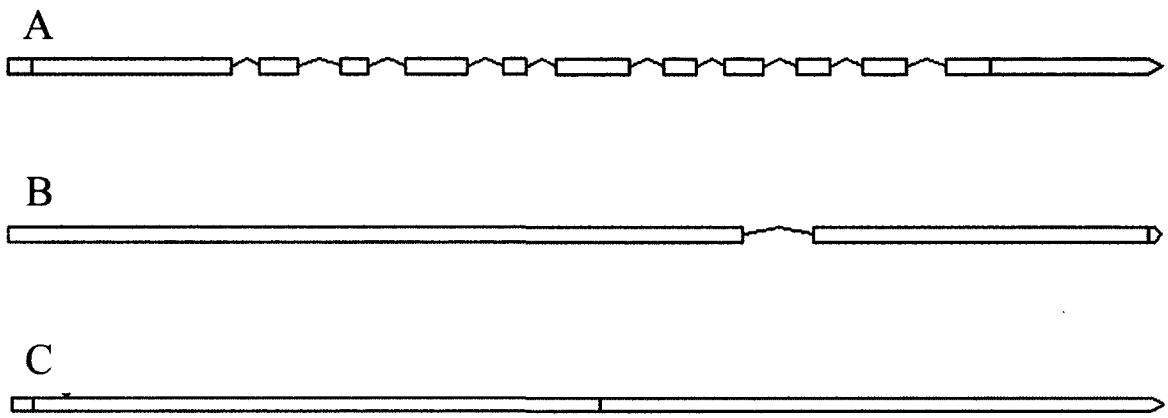
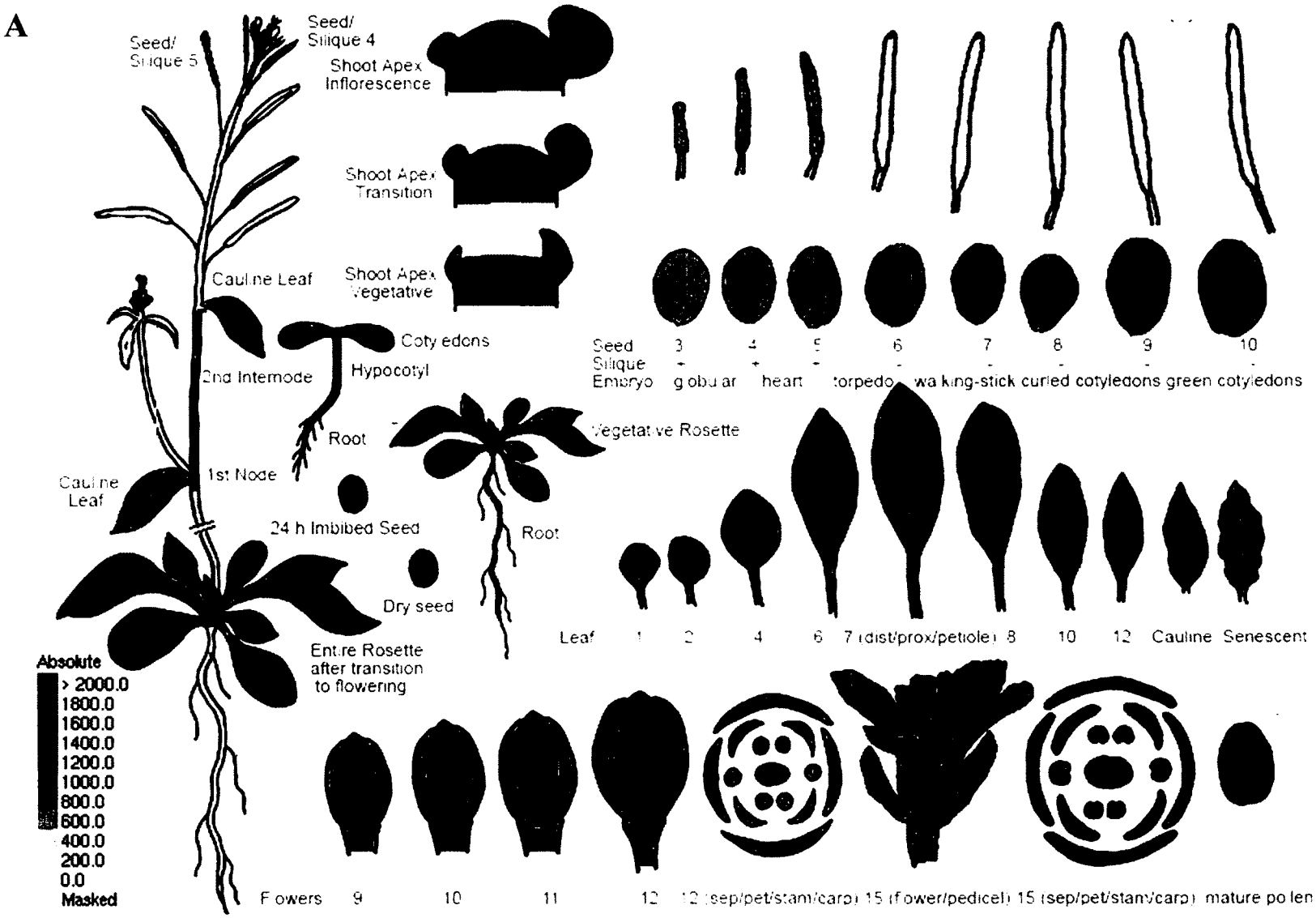


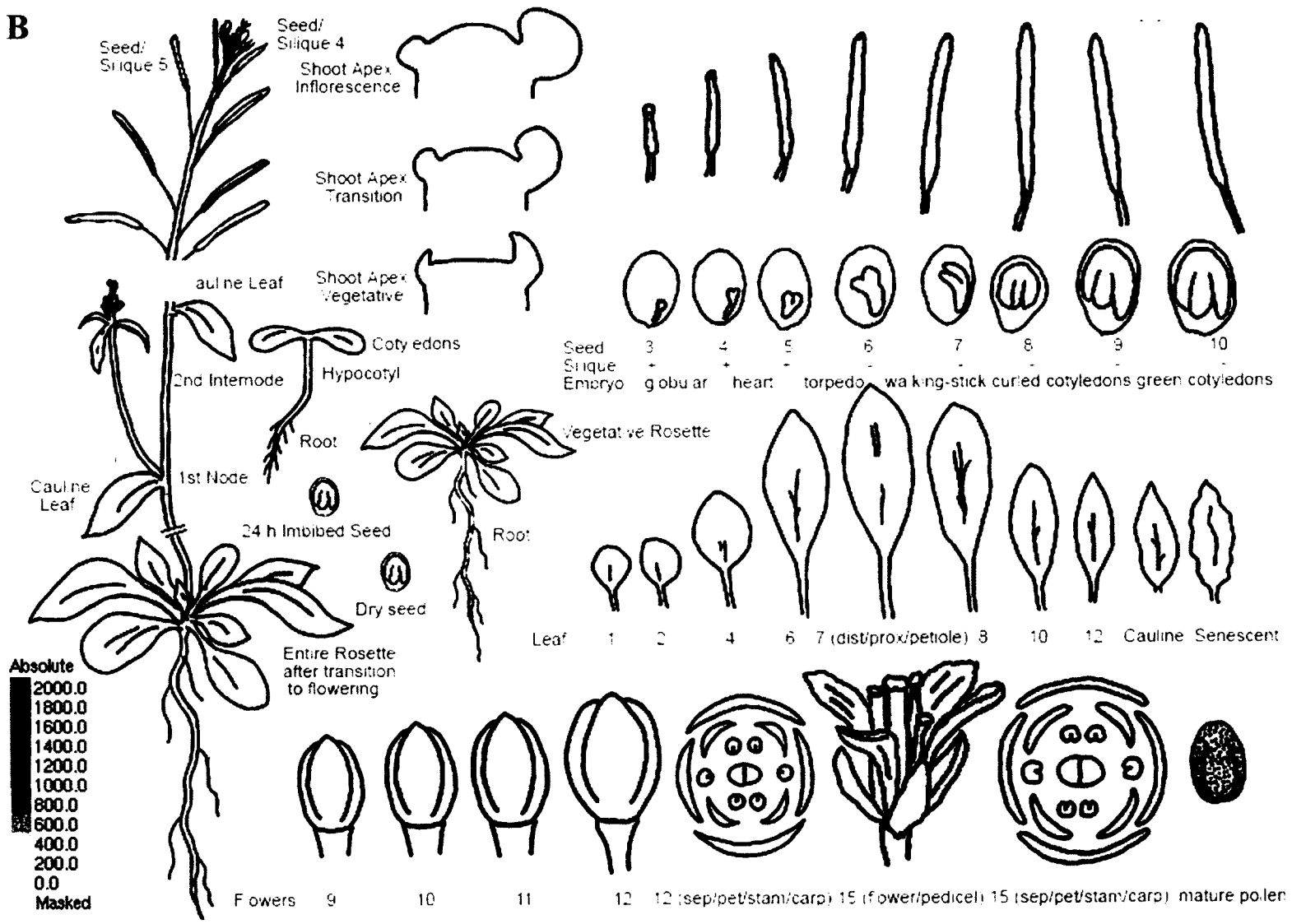
Figure 14 – Arrangement of introns and exons observed for plant CGS genes.
A – "standard" structure with 11 exons, 10 introns.
B – "monocot" structure with 2 exons, 1 intron.
C – "algae" structure with no introns.

3.4 T-DNA Insertion Lines for At1g33320 and At3g01120

A second putative CGS gene is present in *A. thaliana* at locus At1g33320, in addition to the characterized CGS locus (At3g01120). Although there are no expressed sequence tags for the At1g33320 locus, expression data using the ATH1 micro-array is available, and the expression patterns of At3g01120 and At1g33320 differ noticeably. The At3g01120 locus demonstrates increased expression in young and photosynthetic tissues, while the At1g33320 locus has approximately 100-fold lower expression than the A3g01120 locus (Figure 15). The disparity in the expression patterns of these two loci suggests that the At3g01120 locus may account for most, if not all, of the observable CGS activity in wild-type plants.

Figure 15 – The developmental map of expression of the *A. thaliana* loci At3g01120 (A), and At1g33320 (B). Figure created using the Arabidopsis electronic Fluorescent Pictograph (eFP) browser from the Bio-Array Resource at University of Toronto (http://bar.utoronto.ca; Winter *et al*, 2007). The signals represent the absolute expression of each loci as a gradient from red (high expression) to yellow (low expression), which compares each tissue's signal to the threshold signal of 2000.0. The highest value recorded for the At3g01120 locus is 2158.23, and for the At1g33320 locus is 25.01. Expression values with standard deviations greater than 50% are masked (*i.e.* appear white on the figure).


A
Seed/
Silique 5
Seed/
Silique 4
Shoot Apex
Inflorescence
Shoot Apex
Transition
Shoot Apex
Vegetative
Cauline Leaf
Cotyledons
Hypocotyl
2nd Internode
Root
1st Node
Cauline
Leaf
24 h Imbibed Seed
Dry seed
Entire Rosette
after transition
to flowering
Root
Vegetative Rosette
Seed 3 4 5 6 7 8 9 10
Silique
Embryo globular heart torpedo walking-stick curled cotyledons green cotyledons
Leaf 1 2 4 6 7 (dist/prox/petiole) 8 10 12 Cauline Senescent
Flowers 9 10 11 12 12 (sep/pet/stam/carp) 15 (flower/pedicel) 15 (sep/pet/stam/carp) mature pollen
Absolute
> 2000.0
1800.0
1600.0
1400.0
1200.0
1000.0
800.0
600.0
400.0
200.0
0.0
Masked


B
Seed/
Silique 5
Seed/
Silique 4
Shoot Apex
Inflorescence
Shoot Apex
Transition
Shoot Apex
Vegetative
aul ne Leaf
Coty edons
2nd Internode
Hypocotyl
Root
1st Node
Cauline
Leaf
24 h Imbibed Seed
Dry seed
Entire Rosette
after transition
to flowering
Vegetat ve Rosette
Root
Seed 3 4 5 6 7 8 9 10
Silique + + - - - - - -
Embryo g obu ar heart torpedo wa king-stick curled cotyledons green cotyledons
Leaf 1 2 4 6 7 (dist/prox/petiole) 8 10 12 Cauline Senescent
Absolute
2000.0
1800.0
1600.0
1400.0
1200.0
1000.0
800.0
600.0
400.0
200.0
0.0
Masked
F owers 9 10 11 12 12 (sep/pet/stam/carp) 15 (f ower/pedicel) 15 (sep/pet/stam/carp) mature pollen

Seed stocks for T-DNA insertion lines with reported insertions in either locus of interest were ordered (Table 5), screened for the zygosity of the insert using PCR based genotyping, and compared to wild type plants for observable phenotypes. Both lines obtained for the At1g33320 locus (SALK_058208 and SALK_128836C) were found to be homozygous for the insertion and displayed no discernable phenotype (Table 13, Figure 16). Of the five T-DNA insertion lines obtained for the At3g01120 locus, two were found to be homozygous for the insertion (SALK_001032 and SALK_064727) with no discernable phenotypes (Table 13, Figure 16). The lack of distinctive phenotype may be due to the locations of the insertions, which are situated approximately 1000 base pairs 5' of the start codon, and within the last exon, respectively (Table 5; Figure 8), which may make them incapable of sufficiently disrupting gene transcription and/or translation to produce a phenotype distinct from the wild-type. Genotyping of the At3g01120 T-DNA insertion lines SALK_124839 and SALK_049643 was inconclusive, with failure to detect any amplicons indicative of the presence of the expected T-DNA insertion (*i.e.* all plants genotyped demonstrated wild-type banding patterns; Figure 16), and did not display a distinct phenotype (Table 13). Three individual plants of the SALK_045913C T-DNA line were identified as heterozygous for the insertion (Figure 16). These three plants demonstrated delayed growth and were overall smaller than the wild-type, with fewer and smaller leaves at bolting, and less flowers and siliques (Table 13). The combination of the lack of phenotype for the At1g33320 homozygous T-DNA insertion lines and the observed phenotype for the At3g01120 heterozygous SALK_045913C line, suggests that the At3g01120 locus is the major source for CGS activity in *A. thaliana*, while the At1g33320 locus provides minor, if any, contribution to *in vivo* CGS activity.

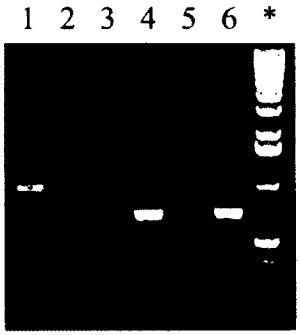
Table 13 – PCR-based genotyping and phenotypic observations for T-DNA insertion lines ordered from the Salk Institute Genomic Analysis Laboratory (Alonso *et al*, 2003).

Line	Insertion Genotype	Phenotype
At1g33320		
SALK_058208	Homozygous	Not different from wild type
SALK_128836C	Homozygous	Not different from wild type
At3g01120		
SALK_001032	Homozygous	Not different from wild type
SALK_064727	Homozygous	Not different from wild type
SALK_124839	No insertion	Not different from wild type
SALK_049643	No insertion	Not different from wild type
SALK_045913C	Heterozygous	Delayed growth, stunted size, fewer and smaller leaves, fewer flowers and siliques

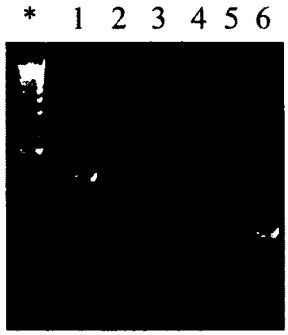
Figure 16 – Gels illustrating typical genotyping PCR results for T-DNA insertion lines. Lanes are labelled as: * - 1 kb ladder (Invitrogen); 1 – wild type gDNA template with locus specific left and right primers; 2 – wild type gDNA template with locus specific right primer and insert specific B primer; 3, 5, 7 – individual transgenic plant gDNA template with locus specific left and right primers; 4, 6, 8 – individual transgenic plant gDNA template with locus specific right primer and insert specific B primer.

At1g33320

SALK_058208

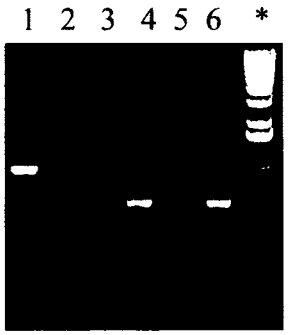


SALK_128836C

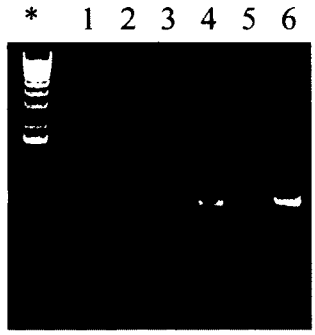


At3g01120

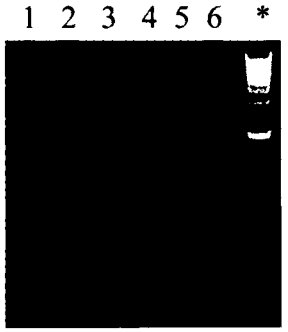
SALK_001032



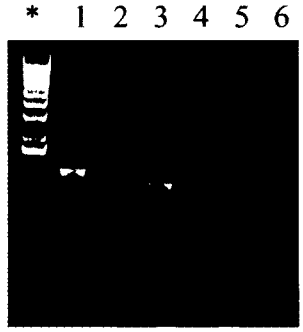
SALK_064727



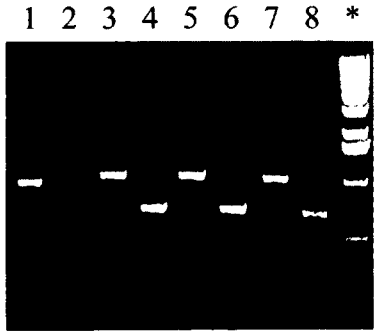
SALK_124839



SALK_049643



SALK_045913C



3.5 Gene Expression Analysis of CGS in Chickpea, Field Pea, and Lentil by Real Time PCR

The expression pattern of CGS over plant development in chickpea, field pea, and lentil was determined using real time PCR, and analyzed by the method described by Pfaffl (2001). The critical threshold levels of CGS from each tissue was normalized to EF1α, and set relative to the expression in the young leaf tissue using Equation [1] (Pfaffl, 2001). The normalized expression level for the young leaf tissue was arbitrarily designated a value of 1 (Figure 17). The highest levels of CGS expression were observed in young and mature leaf tissues in all three species, and the stem tissue of the field pea. All other tissues expressed significantly lower expression levels (approximately 10-fold) as analyzed by one-way ANOVA for each species. A similar expression pattern is observed for the characterized *A. thaliana* CGS locus At3g01120 by ATH1 micro-array (Figure 15; Winter *et al*, 2007).

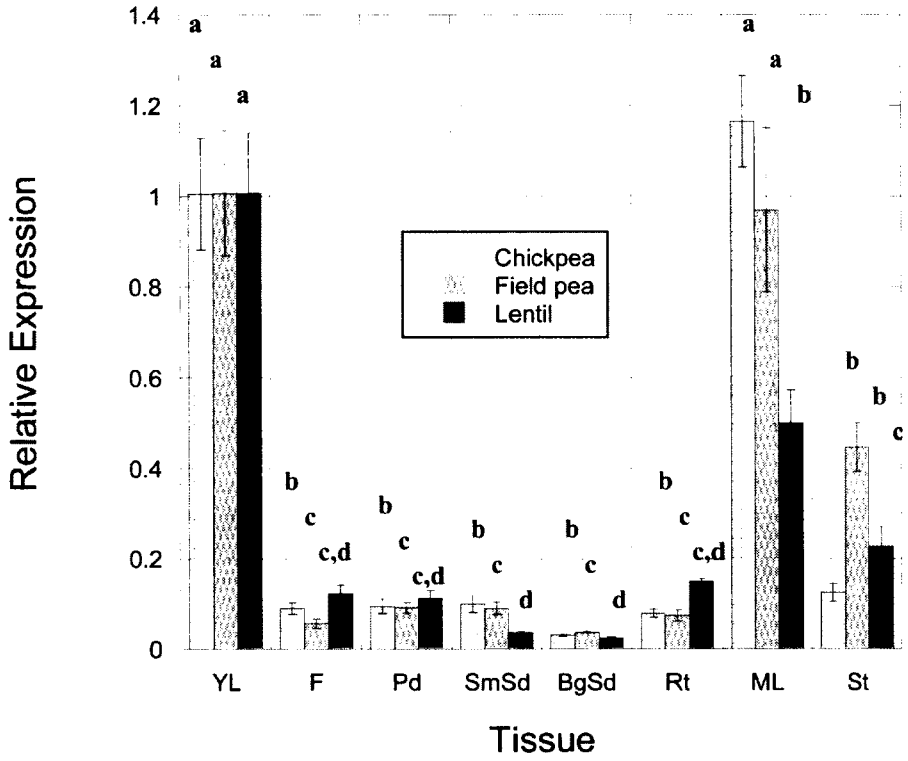


Figure 17 – Tissue distribution of CGS gene transcripts in chickpea (diagonal lines), field pea (hatched), and lentil (grey). Tissues surveyed: young leaf (YL), flower (F), forming seed pod (Pd), forming seed (SmSd), full-sized green seed (BgSd), root (Rt), mature leaf (ML), and stem (St) (Table 3). Real-time PCR analysis was performed using CGS and EF1α gene specific primers (Appendix 1), with CGS normalized against EF1α by the method described by Pfaffl (2001). Total RNA was isolated from at least 3 independent plants per tissue and then pooled prior to cDNA synthesis. Values are the means of technical replicates of at least triplicate runs (each with quadruplicate samples) +/- the standard deviation of the mean. Tissues that are statistically similar by 1-way ANOVA and Tukey's HSD test are indicated by letters.

3.6 Sequencing of the Exon 1/2 Boundary of Chickpea, Field Pea, and Lentil

The location of the exon 1/2 boundary in the coding sequence of CGS from chickpea, field pea, and lentil was obtained by amplification and sequencing a fragment of CGS from genomic DNA of each species. The fragments spanned exons 1-5, 1-3, and 1-2 for chickpea, field pea, and lentil respectively. In all three target species the exon 1/2 boundary occurs in an identical location, which aligns with the corresponding exon 1/2 boundaries of *A. thaliana*, tobacco, and soybean (Figure 11).

3.7 Constructs for Luciferase Assay for Post-Transcriptional SAM Regulation

In order to determine the effect of SAM mediated feed-back on CGS translation in chickpea, field pea, and lentil, constructs were created comprising the first exon of CGS from these plants and *A. thaliana* fused, in frame, with a luciferase reporter gene. As the MTO1 mutations in *A. thaliana* were reported to remove SAM-mediated post-transcriptional regulation, variants corresponding to the MTO1-1 mutation of the characterized *A. thaliana* CGS (G84→S, GGT →AGT; Ominato *et al*, 2002) were introduced into the chickpea, field pea, lentil, and *A. thaliana* exon 1 sequences, via overlap extension site-directed mutagenesis. These mutated exons were also fused in frame with a luciferase reporter. The control and wild-type plant constructs were assayed for successful reporter gene production (Table 14). Initial measurements demonstrate some variability both in replicates of a given construct and between constructs. This indicates that further optimization of the assay is required before beginning assays for regulation by SAM in this system.

Table 14 – Fluorescence of control and wild-type plant exon1::luciferase constructs following *in vitro* transcription/translation for 90 minutes.

Luciferase Construct	Fluorescence
Kit Control[a]	833,995
pGreen Control[b]	32,848
MAVS	8964 6882 5149
A. thaliana-Exon1	24,488 23,681 25,282
Chickpea-Exon1	127 131 155
Field pea-Exon1	548 1267 1448

4. DISCUSSION

Methionine is an essential amino acid for humans and non-ruminant animals, and as such, is required in the diet. Grain legumes, including pulses such as chickpea, lentil and field pea, are important food crops since they grow well in many areas of the world and, although deficient in methionine from a human nutritional perspective, are a rich source of protein (Gepts *et al*, 2005). The pivotal position of CGS as the committing step towards methionine biosynthesis in plants makes it an ideal candidate for studies with the goal of increasing the understanding of this pathway in plants. The objective of this thesis was to sequence and begin genetic characterization of cystathionine γ-synthase from the agronomically important pulse species chickpea, field pea, and lentil, and to provide some insight into the function of the second putative copy of CGS in *A. thaliana* at locus At1g33320.

4.1 Sequencing of CGS from Chickpea, Field Pea, and Lentil

The nucleotide sequences of the CGS transcripts from the three target pulse species are not present inGenBank. Therefore, a protocol for the amplification of unknown sequences was required for this study. While other techniques such as functional complementation of a bacterial knockout have been used to identify transcripts conferring CGS activity (Kim and Leustek, 1996), CGS nucleotide sequences from plant species demonstrate regions of high conservation which permit the design of degenerate oligonucleotide primers to amplify segments of the coding region. This simplifies the procedure to first strand cDNA synthesis, PCR, and sequencing, thereby obviating the need for time and resource intensive library screening. The sequences obtained with this method cover 87% of the coding region, when compared to a soybean CGS mRNA (GenBank accession no. AF141602; Figure 10), in two

overlapping fragments. The high sequence identity between the chickpea, field pea and lentil amplicons and those reported for *A. thaliana*, tobacco, and soybean (Table 7; GenBank accession nos. U83500, AF097180, and AF141602 respectively) support their preliminary assignment as encoding gene products with CGS activity.

Degenerate primer cloning was effective for amplification of fragments between well conserved regions; however the untranslated regions (UTRs) of mRNA are variable in sequence. Therefore, adapter-tagging of cDNA (Figure 6) and Genome Walking (Figure 7) were employed to amplify the 3' and 5' regions of the CGS transcripts, respectively. Both methods rely upon the addition of a designated sequence to the end of the fragment of interest, cDNA or digested gDNA, respectively, and the use of oligonucleotide primers specific to these regions in combination with transcript-specific primers. Amplification of the 3' ends of CGS transcripts through the use of adapter-tagged cDNA was successful, yielding amplicons with sufficient overlap to align with the previously obtained sequences, as well as the poly-adenylated tail sequence, demonstrating the acquisition of the complete 3' UTR. In the targeted pulse species, the 3' UTR region of the CGS transcripts, between the stop codon and the poly-adenylated tail, is 143 – 193 base pairs which agree well with the average plant 3' UTR length of 200 base pairs (Table 8). Transcript UTRs are frequently locations for regulation, and may warrant future investigation in the event that the pulse species do not follow the *Arabidopsis thaliana* model of CGS regulation (Mignone *et al*, 2002).

Genome Walking was used to obtain additional sequence in the region 5' of the segments acquired though the degenerate primer amplifications. Based on alignments with the available sequence genomic for the characterized *A. thaliana* CGS locus

(At3g01120; soybean genomic sequence not available at the time), degenerate primer cloning resulted in coverage into the first exon of the gene. Primers for Genome Walking were designed to be situated within the sequence of the first exon, corresponding to the *A. thaliana* At3g01120 locus, to avoid the possibility of amplification of introns or of a primer spanning an intron-exon boundary. The resulting amplicons ranged from 627 – 754 base pairs in length (Table 9). A putative translational start site was identified within each sequence and the downstream open reading frames contained MTO1 region sequences, indicating that this translational start site is probably correct. Genome Walking uses gDNA as a template, thus conclusions about exact length the 5' UTR cannot be made; however, the information obtained from this method was sufficient to provide sequence covering the proposed start codon of the gene, based on comparison to other plant CGS transcripts, thereby enabling the design of species-specific primers for coding region amplification.

The complete coding regions of CGS cDNA from chickpea, field pea, and lentil were amplified and inserted into the expression vector pTrc-99a2. The pulse CGS coding regions, including the start and stop codons, share 74.0 – 85.0 % and 70.0 – 85.0 % identity with the nucleotide and amino acid sequences, respectively, of CGS sequences from *A. thaliana*, tobacco, and soybean (Table 10; GenBank accession nos. U83500, AF097180 and AB035300, and AF141602 respectively). The slightly lower amino acid sequence identity can be explained through selective pressure to maintain the sequence as demonstrated by the Ka/Ks ratio of 0.2 (Siltberg and Liberles, 2002). Figure 11 details the conserved regions between these plant CGS sequences, with particular note of conservation between the target pulse species and model species in the MTO1 regulatory

region, at the exon 1/2 boundary, and the critical active-site lysine.

In plants, the CGS enzyme is localized to the chloroplast stroma (Wallsgrove *et al*, 1983), hence requiring a targeting sequence, which is cleaved to release the mature enzyme. The cleavage site for the targeting sequence in *A. thaliana* (At3g01120) was identified through Edman degradation, with the mature protein amino-terminus corresponding to valine-68 of the full-length CGS sequence (Ravanel *et al*, 1998). While a conserved glutamine residue is present at the corresponding position in CGS sequences of the *Fabaceae* species soybean, chickpea, field pea, and lentil (Figure 11) in all other plant species this position is a conserved as valine (Ominato *et al*, 2002). The conservation of residues surrounding this position suggests that the site of proteolytic cleavage to remove the plastid targeting peptide is conserved in the targeted pulse species chickpea, field pea, and lentil.

4.2 Complementation of Methionine Auxotrophy of the *E. coli* Strain ER1821*CGS::aadA*

The observed sequence similarity between the sequences obtained from the putative CGS coding regions and the characterized CGS sequences from *A. thaliana*, tobacco, and soybean is not sufficient evidence to assign CGS function to these sequences. Since bacteria produce methionine through a transsulphuration pathway which is similar to that of plants (Figure 1), a bacterial strain deficient in CGS activity may be used for functional complementation studies to confirm that the sequences acquired from the targeted pulse species demonstrate CGS activity (Kim and Leustek, 1996; Hacham *et al*, 2002).

The *metB* coding sequence of the strain ER1821 *CGS::aadA,* employed in this study, has been replaced with the sequence for *aadA*, conferring spectinomycin resistance, resulting in a methionine auxotroph (Lodha, 2011). Rescue of this phenotype can be achieved by complementation with the native *E. coli metB*, or a plasmid expressing a coding sequence with CGS activity (Figure 11; Lodha, 2011). As expected, a plasmid carrying sequence for the characterized *A. thaliana* CGS (locus At3g01120), with the plastid targeting sequence removed, was able to complement the methionine auxotrophy of the ER1821*CGS::aadA* strain. Likewise, the truncated sequences from chickpea, field pea, and lentil rescue the phenotype under the same conditions as the *A. thaliana* construct, demonstrating that these sequences confer also CGS activity (Figure 12, Table 11). The additional time required for complementation by the plant sequences, compared to the *E. coli metB* coding sequence, may be partially due to the codon bias of these sequences or to perturbations of the flux through the methionine/threonine pathways, as the plant CGS enzymes use OPHS as a substrate rather than OSHS, the preferred substrate of bacterial CGS, thus drawing the former away from threonine synthesis (Figure 1).

4.3 Bioinformatics for CGS Copy Number in Sequenced Plant Genomes

The rapid increase in the number of plant species with fully sequenced genomes has enabled investigation of the copy number of CGS in plants. At the time of writing 30 plant genome sequences had been annotated and made available for public use (Table 4; Table 12; Figure 13). Several trends, paralleling the phylogenetic relationships between these species, can be identified within this data. The CGS sequences from the diploid

algae *C reinhardtii* and *V certeri* are single copy, but are markedly different from those of land plants in that they lack both the MTO1 regulatory sequence and introns (green branch, Figure 13; Figure 14; Table 12). All other plant genomes analyzed contain at least one putative CGS coding region which contains the MTO1 regulatory region, and is organized into the "standard" CGS gene structure of 11 exons and 10 introns (Figure 14; Table 12). The non-flowering, non-vascular, tetraploid moss *P. patens* and non-flowering, vascular, diploid lycophte *S. moellendoffii* contain single copy of CGS, with a conserved MTO1 regulatory region and "standard" genome structure, per genome, suggesting that both the regulatory region and the gene structure are characteristics which evolved following the divergence of land plants, but are not limited to the flowering species. Due to the lack of available gymnosperm genomes no comment can be made about CGS in these species.

The tetraploid *Zea mays* (maize) and diploid monocot species *Sorghum bicolour* (sorghum), *Setaria italica* (foxtail millet), *Oryza sativa* (rice), and *Brachypodium distachyon* possess one copy per genome of a putative CGS sequence with an MTO1 regulatory region and "standard" genomic structure (dark blue branch, Figure 13; Table 12). Interestingly, these species also contain a sequence which is similar to CGS, at a single copy per genome, except in rice, where there are 3 copies. These sequences do not contain the MTO1 regulatory region, and have a distinct gene structure comprising two - four exons and one - three introns (yellow branch, Figure 13; Figure 14; Table 12). No investigation of these sequences has been performed, and while they might represent an additional CGS, it is also possible that they encode a protein with a distinct activity. Methionine γ-lyase (MGL) was identified in *A. thaliana* under similar circumstances,

where an unknown sequence shared significant similarity to CGS and CBL was found to be a missing link in methionine catabolism (Rebeille *et al*, 2006). While it is possible that these unidentified CGS-like sequences represent a monocot-specific MGL, other results from queries of these genomes returned sequences which were highly similar to the *A. thaliana* MGLs, suggesting that these, rather than the unknown sequence(s) encode this enzyme.

All of available genome sequences of eudicot species also contain a single CGS copy per genome (Table 12), with the notable exceptions of *A. thaliana, T. halophila,* and *B. rapa* of the *Brassicaceae* family. The diploid *A. thaliana* contains a characterized CGS at locus At3g01120 (Kim and Leustek, 1996; Ravanel *et al*, 1998; Hacham *et al*, 2002), and a second putative CGS at locus At1g33320. This second locus strongly resembles the At3g01120 locus, but lacks the equivalent of the first exon of the latter, which encodes the plastid targeting sequence and MTO1 region. Likewise, the *T. halophila* genome contains two putative CGS sequences, one resembling the At3g01120 and the other At1g33320 from *A. thaliana*. The *B. rapa* genome contains three putative CGS sequences – two that resemble the *A. thaliana* locus At3g01120 and one that resembles At1g33320, the characterized and putative CGS sequences, respectively. *B. rapa* is a paleo-alloploid, recently reverting to diploidy from an alloploid ancestor, which suggests that each of the "standard" CGS genes may have originated from the allopoloid ancestor. The presence of the At1g33320-like copy suggests that one of these ancestors may have been closely related to the *A. thaliana* linage, which possesses the second, putative CGS isoform. In contrast, only one putative copy of CGS was identified in the genome of *Arabidopsis lyrata* which is more closely related to *A. thaliana* than *B. rapa* or

T. halophila. However, this may be partially due to the recent release of this genome (October 2011; Hu *et al*, 2011), which may contain incomplete or incorrect gene notations at this time. Additional searches of this and other genomes may yet reveal a second putative CGS in these species or others closely related to *A. thaliana.*

The presence of a single copy per genome of CGS in the eurosid I clade, including the *Fabaceae* family species soybean, medic, and lotus, suggests that CGS is also present as a single copy in the diploid *Fabaceae* species chickpea, field pea, and lentil. Numerous other genomes are in the process of being sequenced, the most informative among which will be the *Fabaceae* family members *Phaseolus vulgaris* (common bean; Crosby *et al*, 2011), *Medicago sativa* (alfalfa; Alfalfa Genomics Network, 2011) and *Arachis hypogaea* (peanut; United States Department of Agriculture, 2011). Additional sequences in the *Brassicaceae* family, such as *Arabidopsis* and *Brassica* genus members, and increased annotation of the existing *A. lyrata* genome could shed light on the origins of the At1g33320-like CGS copies in these plants, and what, if any, function they possess.

4.4 T-DNA Insertion Lines for At1g33320 and At3g01120

The presence of a second, putative CGS sequence (At1g33320) within the *A. thaliana* genome raises questions about the methionine metabolic pathway and its regulation in this species. The absence of a putative plastid targeting sequence and MTO1 regulatory region in the sequence of the At1g33320 locus suggests that the protein would be localized to the cytosol of the cell, rather than the plastid, and that its expression would be insensitive to the SAM-mediated inhibition of translation observed

for the At3g01120 locus (Wallsgrove *et al*, 1983; Ominato *et al*, 2002). The cytosolic localization is of particular interest, as the substrates for CGS, OPHS and cysteine, are synthesized in the chloroplast, and processing of the cystathionine product by CBL would require its transport into the plastid (Amir, 2010). Also, the expression of this locus is constitutive and at a low level – approximately 100-fold lower than the characterized At3g01120, suggesting minimal contributions to methionine production (Figure 14).

Closer investigation of the amino acid sequences also reveals that where At3g01120 encodes a lysine at position 283, the At1g33320 and At1g33320-like loci encode glutamic acid (Figures 18 and 19). This residue is of particular importance for its location within the active site, where it has been proposed to be involved in substrate coordination. This position is generally conserved as a lysine in plant CGS sequences, whereas the corresponding residue in the *E. coli* CGS enzyme is a glycine (G103 based on the *E. coli* sequence; Steegborn *et al*, 1999). This is significant in the context of the activating group attached to the side-chain hydroxyl moiety of the homoserine substrate – succinate for bacteria and phosphate for plants. The glycine in the bacterial sequence is small and may accommodate the large succinyl group, while the phosphate group of the OPHS substrate of plant CGS is proposed to be coordinated by the lysine at this position (Steegborn *et al*, 1999). In contrast, in the amino acid sequence of the protein encoded by At1g33320 locus, this residue is a glutamic acid – whose negative charge may interfere with the coordination of the negatively charged phosphate group of the OPHS substrate. This suggests that if the At1g33320 locus is translated, the affinity of the resulting protein for the OPHS substrate may be reduced, resulting in minimal contribution to overall CGS activity in the cell.

Figure 18 – BoxShade sequence alignment of CGS amino acid sequences from *A. thaliana* (At3g01120 and At1g33320; At3g and At1g, GenBank accession no.s U83500 and NM_103059, respectively), tobacco (tob; GenBank accession nos. AF097180 and AB035300), and *E. coli* (ecoli; GenBank accession no. CP000948). Blue bar indicates MTO1 region, red arrow indicates the plant exon 1/2 boundary, purple arrow indicates the start of the available tobacco protein crystal structure. Stars indicate conserved key structural and catalytic residues: orange – between all CGS proteins, green – between all plant CGS proteins, purple – active site lysine bound to PLP cofactor.

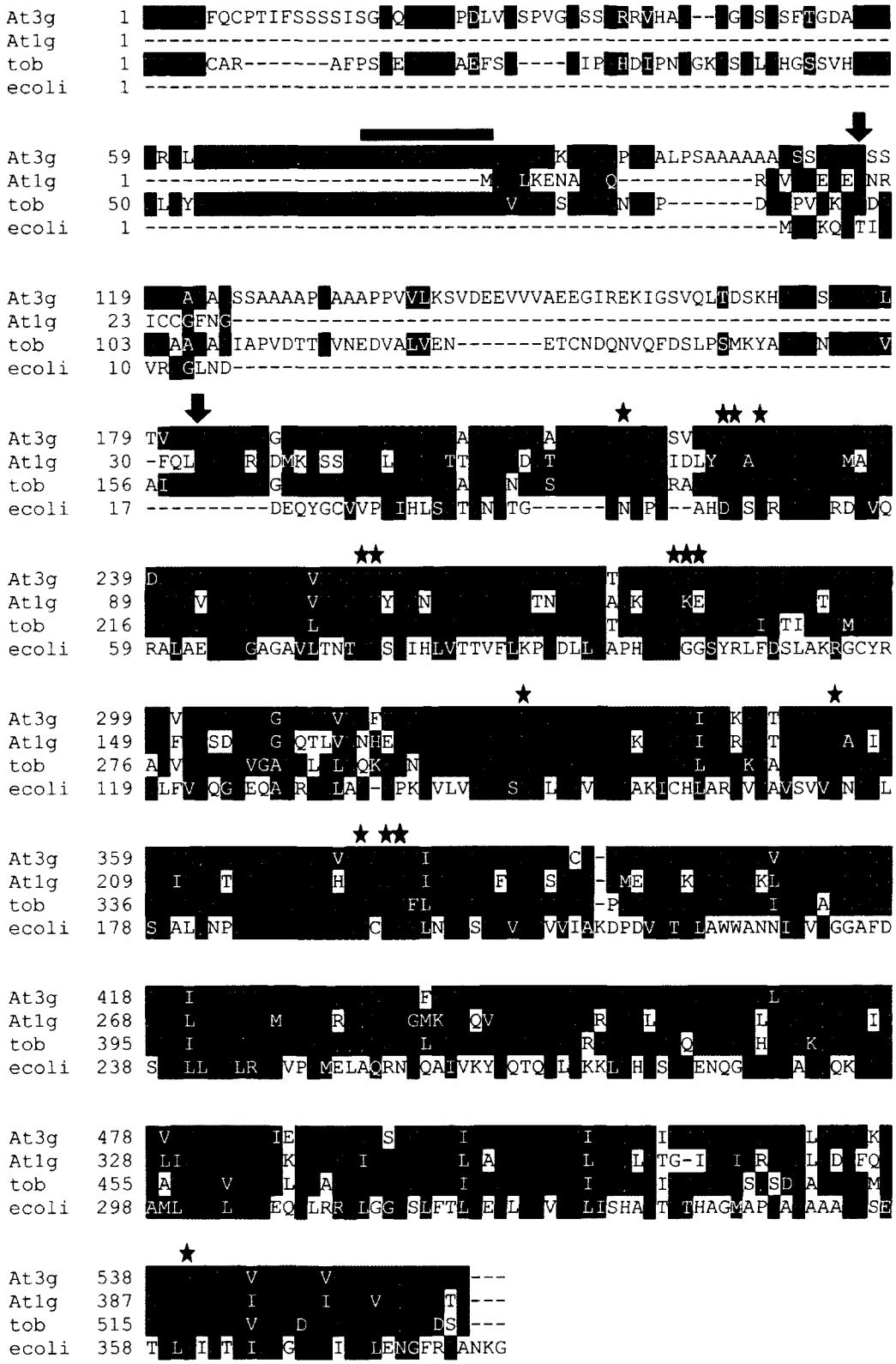
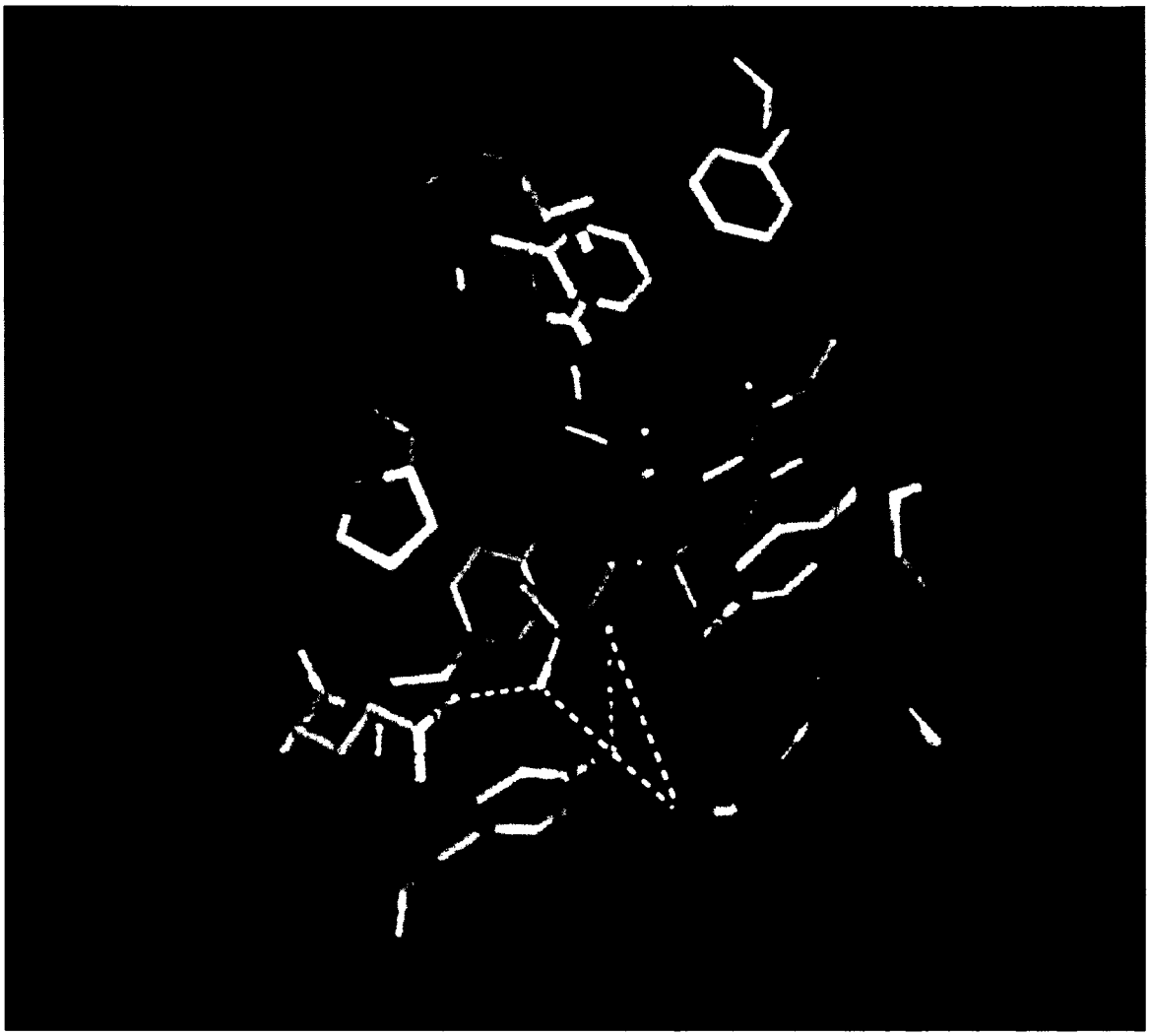

At3g
At1g
tob
ecoli

Figure 19 – Active site architecture of CGS from *N. tabacum* (Protein Data Bank ID 1QGN) binding the inhibitor APPA. The residues in cyan are involved in substrate coordination and binding (Steegborn *et al*, 1999; Steegborn *et al*, 2001). The magenta residue is the lysine at position 165 shown which is proposed to coordinate the phosphate group of the OPHS substrate. In contrast, the glycine at position 103 from *E. coli* CGS (Protein Data Bank ID 1CS1) is highlighted in red (Clausen *et al*, 1998). The PLP-inhibitor complex is shown in orange. The hydrogen bonding network coordinating the substrate is shown in dashed yellow lines. Figure created using Pymol.



To investigate whether the activity conferred to by the At1g33320 locus is required for viability or normal plant development of *A. thaliana*, T-DNA insertion lines for both of the loci of interest were obtained and phenotypically characterized (Figure 8; Table 5; Table 13). Both of the lines with insertions identified within At1g33320 were identified homozygous and displayed a phenotype undistinguishable from wild-type (Figure 16; Table 13). Based on the location of the insertions (approximately 300 base pairs upstream of the translational start site, and within the first exon, respectively) it is probable that either of these lines would reduce At1g33320 expression in these plants, however, both sequencing over the insertion site to confirm location, and mRNA and protein expression analysis to confirm elimination of expression would be required to confirm this.

The inability to isolate homozygous T-DNA insertion lines of At3g01120, combined with the altered phenotype of the three heterozygous plants (SALK_045913C, insertion within exon 3) are indicative of both the essential nature of the At3g01120 locus, which encodes the characterized CGS *A. thaliana* enzyme, for plant survival, and the lack of ability of At1g33320 to complement the loss of At3g01120 function in these lines. While genotyping revealed homozygous insertions for the lines SALK_001032 and SALK_064727, the location of these insertions (approximately 1000 base pairs upstream of the translational start site, and within the last exon, respectively; Figure 8; Table 5), combined with the lack of phenotype compared to wild-type plants, implies that the insertions in these lines do not sufficiently disrupt the transcription and translation of this locus to impair growth. Investigation of the expression levels of At3g01120 in these

lines through methods such as reverse-transcriptase or real time PCR and western blotting would demonstrate whether expression of this locus in these lines has been interrupted.

The lines SALK_124839, SALK_049643, and SALK_045913C are more likely to result in an interruption of At3g01120 expression, given their locations within the coding sequence (within exons 6, 8, and 3, respectively; Figure 8; Table 5). The lack of identification of plants which were homo- or hetero-zygous for the insertion for the lines SALK_124839 and SALK_049643 may indicate that: 1) the lines contain no insertions at the supposed insertion site, possibly due to miss-identification of the insertion; 2) the primers designed for genotyping failed, which is unlikely due to the ability to identify wild-type bands during genotyping; or, 3) the homo- or hetero-zygous phenotype lethal in the absence of some method to alleviate it, such as methionine supplementation. The last option seems the most probable, as two independent reports of anti-sense RNA lines to produce CGS knockdowns both required methionine supplementation to survive (Gakiere *et al*, 2000; Kim and Leustek, 2000).

This is further supported by the detrimental phenotype observed in the SALK_045913C plants heterozygous for the insertion, from which it can also be inferred that At1g33320 is incapable of rescuing even partial expression repression of At3g01120. Detailed investigation of a homozygous SALK_045913C line by phenotypic observation and expression analysis of both loci would enable the exploration of these hypotheses. The growth of the heterozygous and homozyous SALK_045913C plants under methionine supplementation may assist in alleviating any severely detrimental phenotypes by providing a metabolite downstream of the interruption in the pathway and supporting normal plant growth. These data suggest that At3g01120 may be essential for

normal plant growth and development, and that At1g33320 is incapable of rescuing a line in which the expression of At3g01120 has been reduced. This demonstrates that the expression of At1g33320 does not contribute CGS activity to methionine metabolism in a biologically relevant way, and that the characterized CGS (At3g01120) is the source of most if not all of the CGS activity in *A. thaliana*. Therefore, the previous characterization of the *A. thaliana* is likely an accurate model for CGS activity in plants, and the second locus may be dismissed as an evolutionary relic.

4.5 Gene Expression Analysis of CGS in Chickpea, Field Pea, and Lentil by Real Time PCR

Although the copy number of CGS in the targeted pulse species is unknown, analysis of the 30 available plant genomes suggests that it is single copy in these diploid legume species. Real time PCR was performed to determine the relative expression of CGS over plant development and analyzed using the method described by Pfaffl (2001; Equation [1]). This analysis method requires the selection of a housekeeping gene to which the target gene expression is normalized before relative comparison to a selected control. For this study, the housekeeping gene elongation factor 1α was selected, based on its previous use in expression analysis of target genes in different plant developmental stages (Nicot *et al*, 2005). Expression of CGS in the targeted pulse species was determined relative to young leaf, as this tissue was predicted to have relatively high expression of CGS, compared to other tissues, based on examination of the expression pattern of the *A. thaliana* locus At3g01120 (Figure 15; Winter *et al*, 2007).

The expression patterns for chickpea, field pea, and lentil CGS (Figure 17) were determined to generally similar to that observed for the A3g01120 locus, encoding the characterized CGS of *A. thaliana,* in ATH1 micro-array data (Winter *et al*, 2007). The statistically significant differences between the tissues should not be interpreted as "high" and "low" expression, but rather as tissues with "increased" or "decreased" expression relative to each other. The critical threshold values for CGS in all tissues examined were well above the values for negative controls indicating that the CGS transcript is present in all tissues (data not shown). The highest expression of CGS in each of the targeted pulse species was found to be in photosynthetic, rapidly dividing tissues (*i.e.* young leaves), followed by other photosynthetic tissues (*i.e.* mature leaves of all species, and the stem of the field pea), and CGS expression was lower in mature, non-photosynthetic tissues (*i.e.* full-size seeds, prior to desiccation, or roots). These patterns were expected, as transcript levels have been correlated to methionine and CGS protein levels during the developmental stages of *A. thaliana* (Inaba *et al*, 1994).

The expression patterns of CGS in *A. thaliana* and the targeted pulse species, coupled with the elucidation of a transport mechanism involving the cycling of methionine to SMM and back through the actions of MMT and HMT in *A. thaliana*, support a source/sink relationship for methionine in plants. In this model, methionine is synthesized in the photosynthetic tissues and transported for storage, incorporation into proteins, or further processing for the production of SAM and other downstream metabolites (reviewed in Jander and Joshi, 2010). This is contradicted by findings in potato methionine feeding experiments, in which it was found that methionine application did not affect CGS expression (*i.e.* did not decrease the observable levels of CGS mRNA)

or related metabolite levels (*e.g.* threonine, methionine, and SMM) (Kreft *et al*, 2003). This is unexpected, as both potato isoforms of CGS contain MTO1 regions, and would be predicted to be regulated by the SAM generated by the plant in methionine feeding experiments. Kreft *et al.* (2003) also observed that translation of *A. thaliana* CGS in an *in vitro* system was impaired in the presence of methionine, in direct contrast to the studies performed by Naito *et al*, in which it was found that SAM, and not methionine, is the effector in the posttranscriptional autoregulation of CGS (Kreft *et al*, 2003; Chiba *et al*, 2003). Taken together, this suggests an alternate, as yet unidentified, mechanism for the regulation of methionine biosynthesis in potato plants, unique from that observed in *A. thaliana*. *In vitro* studies of posttranscriptional CGS autoregulation in the targeted pulses would clarify which model (*e.g. A. thaliana* or potato) these plant species most closely resemble.

4.6 Sequencing of the Exon 1/2 Boundary of Chickpea, Field Pea, and Lentil and Constructs for Luciferase Assay for Post-Transcriptional SAM Regulation

Exon 1 of the *A. thaliana* CGS locus At3g01120 contains sequences which have been implicated in both localization of the protein to the chloroplast stroma (Ravanel *et al*, 1998; Wallsgrove *et al*, 1983) and SAM-mediated feedback inhibition of translation (Suzuki *et al*, 2001; Ominato *et al*, 2002). Complete removal of exon 1 from the *A. thaliana* sequence had no effect on the ability of the remainder of the sequence to complement a methionine auxotrophic bacterial strain (Hacham *et al*, 2002). While all plants contain at least one putative CGS with the MTO1 regulatory region intact, there have been conflicting reports as to whether or not this region confers the same regulation

in all plants (Kreft *et al*, 2003). This has led to intense study of the regulatory nature of this exon, particularly the MTO1 region, through its fusion to a luciferase reporter gene for the investigation of the mechanism of the SAM-mediated regulation conferred by this region (Suzuki *et al*, 2001; Ominato *et al*, 2002; Onoue *et al*, 2011).

The first step required for investigation of this region of the pulse sequence was to confirm the exon 1/2 boundary in CGS from these species. The exon 1/2 boundary was identified in the chickpea, field pea, and lentil sequences based on sequence alignments with *A. thaliana*, and later confirmed to be in the same location through sequencing of genomic DNA segments spanning this region. A series of plasmid constructs were subsequently generated to examine the post-transcriptional SAM-mediated feedback regulation of the targeted pulse species. Initial testing of these constructs demonstrated that luciferase activity from the control and wild-type plant exon1::luciferase fusions were measured following *in vitro* transcription and translation. Differences in the measured levels of luciferase activity between the various constructs may be attributed to the different sequences preceding the luciferase reporter gene – the longer plant exons take more time to transcribe and translate, resulting in lower reporter activity compared to the controls. Purity of the individual luciferase constructs may also impact the efficacy of the *in vitro* transcription/translation. Further optimization of this system is required prior to continuation of these experiments.

Additional *in vitro* transcription and translation experiments, in the presence and absence of SAM, will be required to demonstrate if SAM-mediated regulation occurs in these plant constructs. Constructs containing the MTO1-1 mutation were also prepared to confirm the role of the MTO1 region in any observed SAM-mediated regulation. It is

expected that the pulse species will demonstrate the SAM-mediated regulation, and that the MTO1-1 mutation will ease this regulation. In this case the *A. thaliana* regulatory model could be applied to the pulses, particularly in attempts to increase methionine content. However, should the pulse species resemble potato, *i.e.* not demonstrate the posttranscriptional control by SAM, then the use of *A. thaliana* as a model for the regulation of CGS in methionine biosynthesis would be inappropriate, and the development of a novel regulatory model for the pulse species would be required.

5. CONCLUSIONS AND FUTURE DIRECTIONS

This study represents the first genetic characterization of CGS from pulse species, in which methionine is a limiting nutrient from a human and non-ruminant animal nutritional perspective. The sequence of CGS from the pulse species *Cicer arietinum* (chickpea), *Pisum sativum* (field pea), and *Lens culinaris* (lentil) were obtained by PCR based methods, and found to be similar to previously characterized CGS transcripts from plants (*e.g. Arabidopsis thaliana, Nicotiana tabacum,* and *Glycine max*). Conserved features include: genomic structure of the first 3-5 introns/exons, the presence of a putative plastid targeting sequence and MTO1 regulatory region in the first exon, key active site and structural residues, and expression pattern over plant development. The pulse CGS coding sequences were also able to functionally complement an *E. coli* CGS replacement strain, thereby demonstrating CGS activity.

The significance of the multiple copies of CGS in *A. thaliana* with respect to other plant species was clarified through a bioinformatics study for the copy number of CGS in sequenced and annotated plant genomes. Generally CGS is present in plants at single copy per genome, and that a putative second copy of CGS arose during the divergence of the *Brassicaceae* family. However the role of the second putative CGS remains in question, as its activity has not been demonstrated and an active site residue of the *A. thaliana* At1g33320 locus which is proposed to interfere with substrate coordination is not conserved. Additionally, investigation of T-DNA insertion lines for both the characterized *A. thaliana* locus At3g01120, and the putative CGS at locus At1g33320 call into question the *in vivo* contribution of CGS activity by At1g33320 as

demonstrated by the detrimental phenotype of a heterozygous line with an insertion in At3g01120, and lack of phenotype in lines homozygous for insertions in At1g33320.

The regulation of CGS in *A. thaliana* has been well characterized, and appears to involve a combination of SAM feedback inhibition of CGS translation coupled and allosteric activation of TS, which competes with CGS for the branch-point metabolite OPHS. However, work in the non-model species potato has suggested that the SAM–mediated post-transcriptional regulation of CGS may be unique to *A. thaliana*. The generation of pulse species exon1::luciferase fusion constructs was performed as a first step towards determining if chickpea, field pea, and lentil share the SAM-mediated feedback, or if other regulatory mechanisms may be at work.

Future directions for this project include extensions of the genetic characterization of CGS in pulse crops by Southern blotting to determine gene copy number. It is predicted that CGS will be present at single copy per genome in these diploid legume species, based on the analysis of available sequenced genomes of plant species. Southern blotting has been used in a number of other species for the determination of copy number, and has generally indicated that CGS is present in a single copy per genome. Examples include potato (Riedel *et al*, 1999), soybean (Hughes *et al*, 1999), and strawberry (Marty *et al*, 2000). However, while CGS was also found to be single copy gene in *A. thaliana* by Southern blot analysis (Kim and Leustek, 1996) a second putative isoform at locus At1g33320 was later identified when the *A. thaliana* genome was sequenced (Arabidopsis Genome Initiative, 2000). Southern blots of the pulses would act merely as support for assignment of CGS copy per genome until complete genome sequences for these species become available, permitting a definitive assignment.

While CGS has been confirmed to be localised to the chloroplast in peas (Wallsgrove *et al*, 1983), and the cleavage site of a targeting sequence has been identified in *A. thaliana* and is conserved in the pulse species, the plastid localization of CGS in chickpea and lentil, and the verification of the targeting sequence cleavage site in the targeted pulses has not been performed. Confirmation of targeting to the chloroplast may be performed using targeting sequence::reporter gene fusions (Xiao *et al*, 2008). Identification of the plastid targeting sequence in these species will be important for future studies, should it be desirable to target modified CGS sequences or other methionine biosynthetic pathway alterations to the chloroplast in these plants.

Elucidation of the ability of SAM to posttranscriptionally regulate CGS through feedback inhibition of translation in the pulse species will be critical to application of the *A. thaliana* model to these species. The conflicting reports of SAM-mediated feedback in potato in comparison to *A. thaliana* suggest that additional regulatory mechanisms may be at work. Understanding all of the regulatory factors involved is essential for future studies with the goal of increasing the methionine content of plants, as it is likely that any increase in methionine will lead to a corresponding increase in cellular SAM levels. Should the wild-type exon1::luciferase constructs demonstrate SAM-mediated feedback inhibition, the mechanism of this action will be probed through investigation of mutant constructs. The MTO1-1 mutation in *A. thaliana* has been demonstrated to alleviate the SAM-mediated feedback leading to increases in soluble methionine in plants (Inaba *et al*, 1994), and if the pulse species are regulated in a similar manner introduction of an equivalent mutation in the pulse exon 1 sequence is predicted to reduce the ability of SAM to inhibit translation. However, should the pulse sequences be found to be SAM-

insensitive, it will be necessary to explore alternative regulatory mechanisms, such as the increased importance of threonine activation by SAM in potato.

The CGS enzyme from a variety of plant species has been characterized, including enzymes purified from native tissue (*Lemna* (Thompson *et al*, 1981), wheat (Kreft *et al*, 1995), and spinach (Ravanel *et al*, 1995), and recombinantly expressed (*A. thaliana* (Ravanel *et al* 1998) and tobacco (Clausen *et al*, 1999). The biochemical characterization of the recombinant pulse CGS enzymes would provide enzyme kinetic data which will be critical for potential future applications if the metabolic flux towards methionine will be increased.

The applicability of *A. thaliana* as a model for the mechanism and regulation of methionine metabolism in plants has been challenged with the discovery of a second putative CGS at locus At1g33320. Analysis of T-DNA insertion lines for both CGS loci in *A. thaliana* has suggested that the putative CGS (At1g33320) may not provide much if any CGS activity to the cell, since homozygous insertion lines in this locus do not demonstrate a phenotype discernable from wild-type. Also, heterozygous insertion lines for the At3g01120 demonstrate a disturbed phenotype, suggesting that the At1g33320 loci cannot rescue. More in depth characterization of these lines through mRNA and protein expression studies will confirm if the insertions in these lines successfully interrupt CGS expression. The inability to identify homozygous insertion lines for the At3g01120 locus may be due to the severe phenotypes expected with the interruption of CGS expression. Growing future generations with methionine feeding should alleviate severe phenotypes, and would be recommended for future attempts at isolating homozygous insertions at the At3g01120 locus.

The studies presented here, combined with the future directions for this research, will represent the characterization of CGS from pulse species. This data, combined with information from the model species *A. thaliana* and non-model species potato provide a look at the complexities of methionine biosynthesis in plants. Any attempts to increase methionine in pulses through increasing the flux of metabolites towards methionine will require a detailed understanding of regulation of this pathway, especially through CGS which catalyzes the committing step towards methionine. The success of strategies to increase nutrient levels in crop plants using metabolic engineering has been demonstrated through the introduction of a high lysine mays line in which the lysine metabolic pathway was modified to include a feed-back insensitive DHDPS, leading to an increase in lysine levels (Huang *et al*, 2005). Likewise, an increase in methionine content in pulse crop seeds will result in a nutritionally enhanced, value-added crop, with both agronomical and economical benefits.

6. REFERENCS

Agriculture and Agri-Food Canada. 2010. Overview of the Canadian pulse industry, 2009. Accessed online: http://www.ats-sea.agr.gc.ca/can/4753-eng.pdf

Alfalfa Genomics Network. 2011. Overview. Accessed online: http://www.alfalfa-genome.org

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., Ecker, J.R. 2003. Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*. *Science* **301**:653-657

Alvarez, C., Calo, L., Romero, L.C., Garcia, I., Gotor, C. 2010. An *O*-acetylserine (thiol) lyase homolog with L-cysteine desulfhydrase activity regulates cysteine homeostasis in Arabidopsis. *Plant Physiology* **152**:656-669

Amir, R. 2010. Current understanding of the factors regulating methionine content in vegetative tissues of higher plants. *Amino Acids* **39**:917-931

Applied Biosystems. 2005. Sequence Scanner Software v1.0. Build ID: 20050914-1.0fc11

Arabidopsis Genome Initiative. 2000. Analysis of the genome of the flowering plant *Arabidopsis thaliana*. *Nature* **408**(6814):796-815

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., de Pamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., Ashton, N.W., Axtell, M.J., Barker, E., Barker, M.S., Bennetzen, J.L., Bonawitz, N.D., Chapple, C., Cheng, C., Correa, L.G.G., Dacre, M., De Barry, J., Dreyer, I., Elias, M., Engstrom, E.M., Estelle, M., Feng, L., Finet, C., Floyd, S.K., Frommer, W.B., Fujita, T., Gramzow, L., Gutensohn, M., Harholt, J., Hattori, M., Hey, A., Hirai, T., Hiwatashi, Y., Ishikawa, M., Iwata, M., Karol, K.G., Koehler, B., Kolukisaoglu, U., Kubo, M., Kurata, T., Lalonde, S., Li, K., Li, Y., Litt, A., Lyons, E., Manning, G., Maruyama, T., Michael, T.P., Mikami, K., Miyazaki, S., Morinaga, S., Murata, T., Mueller-Roeber, B., Nelson, D.R., Obara, M., Oguri, Y., Olmstead, R.G., Onodera, N., Petersen, B.L., Pils, B., Prigge, M., Rensing, S.A., Riaño-Pachón, D.M., Roberts, A.W., Sato, Y., Scheller, H.V., Schulz, B., Schulz, C., Shakirov, E.V., Shibagaki, N., Shinohara, N., Shippen, D.E., Sørensen, I., Sotooka, R., Sugimoto,

N., Sugita, M., Sumikawa, N., Tanurdzic, M., Theißen, G., Ulvskov, P., Wakazuki, S., Weng, J.-K., Willats, W.W.G.T., Wipf, D., Wolf, P.G., Yang, L., Zimmer, A.D., Zhu, Q., Mitros, T., Hellsten, U., Loqué, D., Otillar, R. Salamov, A., Schmutz, J., Shapiro, H., Lindquist, E., Lucas, S., Rokhsar, D., Grigoriev, I.V. 2011. The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. *Science* **332**:960-963

Bartlem, D., Lambien, I., Okamoto, T., Itaya, A., Uda, Y., Kijima, F., Nambara, E., Naito, S. 2000. Mutation in threonine synthase gene results in an over-accumulation of soluble methionine in Arabidopsis. *Plant Physiology* **123**:101-110

Baum, H.J., Madison, J.T., Thompson, J.F. 1983. Feedback inhibition of homoserine kinase from radish leaves. *Phytochemistry* **22**:2409-2412

Bergen Center for Computational Science. 2011. Ka/Ks Calculation Tool. Accessed online: http://services.cbu.uib.no/tools/kaks

Bouvier, F., Linka, N., Isner, J.-C., Mutterer, J., Weber, A.P.M, Camara, B. 2006. *Arabidopsis* SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. *The Plant Cell* **18**(11):3088-3105

The Brassica rapa Genome Sequencing Project Consortium, Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.-H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Wang, J., Wang, X., Freeling, M., Pires, J.C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.-S., Weisshaar, B., Liu, B., Li, B., Liu, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Wang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H., Jin, H., Parkin, I.A.P., Batley, J., Kim, J.-S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Li, J., Yu, J., Meng, J., Wang, J., Min, J., Poulain, J., Wang, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S., Sato, S., Sun, S., Kwon, S.-J., Choi, S.-R., Lee, T.-H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y., Wang, Z., Li, Z., Wang, Z., Xiong Z., Zhang Z. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43**:1035-1039

Cannon, S.B., Sterck, L., Rombauts, S., Sato, S., Cheung, F., Gouzy, J., Wang, S., Mudge, J., Vasdewani, J., Schiex, T., Spannagl, M., Monaghan, E., Nicholson, C., Humphray, S.J., Schoof, H., Mayer, K.F.X., Rogers, J., Quétier, F., Oldroyd, G.E., Debellé, F., Cook, D.R., Retzel, E.F., Roe, B.A., Town, C.D., Tabata, S., Van de Peer, Y., Young, N.D. 2007. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of the National Academy of Science* **103**(40):14959-14964

Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J., Chen, G., Cahoon, E.B., Gedil, M., Stanke, M., Haas, B.J., Wortman, J.R., Fraser-Liggett, C.M., Ravel, J., Rabinowicz, P.D. 2010. Draft genome sequence of the oilseed species Ricinus communis. *Nature Biotechnology* **28**:951-956

Chiba, Y., Ishikawa, M., Kijima, F., Tyson, R.H., Kim, J., Yamamoto, A., Nambara, E., Leustek, T., Wallsgrove, R.M., Naito, S. 1999. Evidence for autoregulation of cystathionine γ-synthase in *Arabidopsis*. *Science* **286**:1371-1374

Chiba, Y., Sakurai, R., Yoshino, M., Ominato, K., Ishikawa, M., Onouchi, H., Naito, S. 2003. *S*-adenosyl-$_L$-methionine is an effector in the posttranscriptional autoregulation of the cystathionine γ-synthase gene in Arabidopsis. *Proceedings of the National Academy of Science* **100**(18):10225-10230

Clausen, T., Huber, R., Prade, L., Wahl, M.C., Messerschmidt, A. 1998. Crystal structure of *Escherichia coli* cystathionine gamma-synthase at 1.5 A resolution. *EMBO* **17**:6827-6838

Clausen, T., Wahl, M.C., Fuhrmann, J.C., Laber, B., Streber, W., Steegborn. C. 1999. Cloning, purification, and characterisation or cystathionine gamma-synthase from *Nicotiana tabacum*. *Biological Chemistry* **380**(10):1237-1242

Crosby, W., Cao, R., Power, K.,Yu, K, Paule, K.P., Bozzo, G., Maraolals, F., Navabi, A., Weersink, A. 2011. Draft genome sequence for common bean *P. vulgaris*. Accessed online: http://www.beangenomics.ca/research/projects/view/draft-genome-sequence-for-common-bean-i-p-vulgaris-i

Curien, G., Dumas, R., Ravanel, S., Douce, R. 1996. Characterization of an *Arabidopsis thaliana* cDNA encoding an *S*-adenosylmethionine-sensitive threonine synthase. *FEBS Letters* **390**:85-90

Curien, G., Laurencin, M., Robert-Genton, M., Dumas, R. 2007. Allosteric mono-functional aspartate kinases from *Arabidopsis*. *FEBS Journal* **274**(1):164-176

Datko, A.H., Giovanelli, J., Mudd, S.H. 1974. Homocysteine biosynthesis in green plants. *Journal of Biological Chemistry* **249**(4):1139-1155

Eichel, J., Gonzolez, J.C., Hotez, M, Matthews, R.C., Schroder, J. 1995. Vitamin-B12-independent methionine synthase from a higher plant (*Cartharanthus roseus*). *European Journal of Biochemistry* **230**:1053-1058

Farsi A., Lodha P.H., Skanes J.E., Los H., Kalidindi N., Aitken S.M. 2009. Interconversion of a pair of active-site residues in *Escherichia coli* cystathionine γ-synthase, *E. coli* cystathionine β-lyase, and *Sacchromyces cerevisiae* cystathionine γ-lyase and development of tools for the investigation of their mechanisms and reaction specificity. *Biochemistry and Cell Biology* **87**:445-457

Food and Agriculture Organization. 1994. Definition and classification of commodities. 4. *Pulses and derived products*. Accessed online: http://www.fao.org/es/faodef/fdef04e.htm

Galili, G., and Hoefgen, R. 2002. Metabolic engineering of amino acids and seed storage proteins in plants. *Metabolic Engineering* **4**:3-11

Gakiere, B., Ravanel, S., Droux, M., Douce, R., Job. D. 2000. Mechanisms to Account for Maintenance of the Soluble Methionine Pool in Transgenic Arabidopsis Plants Expressing Antisense Cystathionine γ-Synthase cDNA. *C. R. Academy of Science III* **323**:841-851

Gakiere, B., Denis, L., Droux, M., Job. D. 2002. Over-exression of cystathionine γ-synthase in *Arabidopsis thaliana* leads to increased levels of methionine and *S*-methyl-methionine. *Plant Physiology and Biochemistry* **40**:119-126

Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F., Young, N.D. 2005. Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. *Plant Physiology* **137**:1228-1235

Giovanelli, J. and Mudd, S.H. 1971. Transsulfuration in higher plants: Partial purification and properties of β-cystathionase of spinach. *Biochimica Biophysica Acta* **227**:654-670

Giovaneli, J., Mudd, S., Datko, A.H. 1985. *In vivo* regulation of *de novo* methionine biosynthesis in a higher plant (*Lemna*). *Plant Physiology* **77**(2):450-455

Goto, D.B., Ogi, M., Kijima, F., Kumagai, T., van Werven, F., Onouchi, H. Naito, S. 2002. A single-nucleotide mutation in a gene encoding *S*-adenosylmethionine synthase is associated with methionine over-accumulating phenotype in *Arabidopsis thaliana*. *Genes and Genetic Systems* **77**:89-95

Goyer, A., Collakova, E., Shachar-Hill, Y., Hanson, A.D. 2007. Functional characterization of a methionine γ-lyase in Arabidopsis and its implication in an alternative to the reverse trans-sulfuration pathway. *Plant and Cell Physiology* **48**(2):232-242

Hacham, Y., Avraham, T., Amir. R. 2002. The N-Terminal Region of Arabidopsis Cytathionine γ-Synthase Plays an Important Regulatory Role in Methionine Metabolism. *American Society of Plant Biologists* **128**:454-462

Hacham, Y., Gophna, U., Amir, R. 2003. *In vivo* analysis of various substrates utilized by cystathionine γ-synthase and *O*-acetylhomoserine sulphydylase in methionine biosynthesis. *Molecular Biology and Evolution* **20**(9):1513-1520

Hacham, Y., Schuster, G., Amir, R. 2006. An *in vivo* internal delection in the N-terminus region of Arabidopsis cystathionine γ-synthase results in CGS expression that is insensitive to methionine. *The Plant Journal* **45**:955-967

Hacham, Y., Song, L., Schuster, G., Amir, R. 2007. Lysine enhances methionine content by modulating the expression of *S*-adenosylmethionine synthase. *The Plant Journal* **51**:850-861

Han, M.V., and Zmasek, C.M. 2009. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Biochemistry* **10**:356

Haughn, G.W. and Somerville, C. 1981. Sulfonylurea-resistant mutants of *Arabidopsis thaliana*. *Molecular and General Genetics* **204**(3):430-434

Hesse, H., Kreft, O., Maimann, S., Zeh, M., Hoefgen, R. 2004. Current understanding of the regulation of methionine biosynthesis in plants. *Journal of Experimental Botany* **55**(404):1799-1808

Ho, C.-L., Noji, M., Saito, M., Saito, K. 1999. Regulation of serine biosynthesis in Arabidopsis: Crucial role of plastidic 3-phosphoglycerate dehydrogenase in non-photosynthetic tissues. *Journal of Biological Chemistry* **274**(1):397-402

Hoagland D.R., and Arnon D.I. 1950. The water-culture method for growing plants without soil. *In* California Agricultural Experiment Station Circular 347. The College of Agriculture, University of California, Berkeley, CA, pp 1–39

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottilar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F., Van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., Guo, Y.L. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics* **43**(5):476-481

Huang, S., Kruger, D.E., Frizzi, A., D'Ordine, D.L., Florida, C.A., Adams, W.R., Brown, W.E., Luethy, M.H. 2005. High-lysine corn produced by the combination of enhanced lysine biosynthesis and reduced zein production. *Plant Biotechnology* **3**(6):555-569

Huang X., and Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Research* **9**:868-877

Hughes, C., Gebhardt, J., Reuss, A., Matthews. B. 1999. Identification of a cDNA encoding cystathionine γ-synthase in soybean. *Plant Science* **146**(2):69-79

Inaba, K., Fujiwara, T., Hatashi, H., Chino, M., Komeda, Y., Naito, S. 1994. Isolation of an *Arabidopsis thaliana* mutant, *mto1*, that overaccumulates soluble methionine. *Plant Physiology* **104**:881-887

Inoue H., Nojima H., Okayama H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**:23-28

Integrated DNA Technologies. 2011. OligoAnalyzer 3.1. Accessed online: http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/

The International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**:763-768

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., Billault, A., Segurens, C., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pè, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.-F., Weissenbach, J., Quétier, F., Wincker, P. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**:463-467

Jander, G., and Joshi, V. 2010. Recent Progress in Deciphering the Biosynthesis of Aspartate-Derived Amino Acids in Plants. *Molecular Plant* **3**(1):54-65

Jarvis, P. 2008. Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytologist* **179**(2):257-285

Joshi, V. and Jander, G. 2009. Arabidopsis methionine γ-lyase is regulated according to isoleucine biosynthesis needs but plays a subordinate role to threonine deaminase. *Plant Physiology* **151**(1):367-378

Katz, Y.S, G. Galili, R. Amir. 2006. Regulatory role of cystathionine gamma-synthase and *de novo* synthesis of methionine in ethylene production during tomato fruit ripening. *Plant Molecular Biology*. **61**:255-268

Kim, J, and Leustek. T. 1996. Cloning and Analysis of the Gene for Cystathionine Gamma-Synthase from *Arabidopsis thaliana*. *Plant Molecular Biology* **32**(6):1117-1124

Kim, J. and Leustek. T. 2000. Repression of Cystathionine γ-synthase in *Arabidopsis thaliana* produces partial methionine auxotrophy and developmental abnormalities. *Plant Science* **151**:9-18

Kreft, O., Hoefgen, R., and Hesse, H. 2003. Functional analysis of cystathionine γ-synthase in genetically engineered potato plants. *Plant Physiology* **131**:1843-1854

Kusano, M., Fukushima, A., Redstig H., Kobayashi, M., Otsuki, H., Onouchi, H., Naito, S., Hirai, M.Y., Saito, K. 2010. Comparative metabolomics charts the impact of genotype-dependant methionine accumulation in *Arabidopsis thaliana*. *Amino Acids* **39**:1013-1021

Laber, B., Maurer, W., Henke, C., Grafe, S., Ehlert, S., Messerschmidt, A., Clausen, T. 1999. Characterization of recombinant *Arabidopsis thaliana* threonine synthase. *European Journal of Biochemistry* **263**:212-221

Larkin, M.A., Blacksheilds, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. Thompson, J.D., Gibson T.J. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21):2947-2948

Lee, M., and Leustek, T. 1999. Identification of the gene encoding homoserine kinase from *Arabidopsis thaliana* and characterization of the recombinant enzyme derived from the gene. *Archives of Biochemistry and Biophysics* **372**(1):135-142

Lee, M., Martin, M.N., Hudson, A.O., Lee, J., Muhitch, M.J., Leustek, T. 2005. Methionine and threonine synthesis are limited by homoserine availability and not the activity of homoserine kinase in *Arabidopsis thaliana. The Plant Journal* **41**(5):685-696

Lodha, P. 2011. Exploration of the determinants of substrate and reaction specificity in the pyridoxal 5'-phosphate-dependent enzymes of the transsulfuration pathways. Ph.D. Thesis – Carleton University, Ottawa, Canada

Macnicol, P.K., Datko, A.H., Giovanelli, J., Mudd, S.H. 1981. Homocysteine biosynthesis in green plants: Physiological importance of the transsulphuration pathway in *Lemna paucicostata. Plant Physiology* **68**(3):619-615

Mandel M. and Higa A. 1970. Calcium-dependent bacteriophage DNA infection. *Journal of Molecular Biology* **53**:159-162

Marty, I., Douat, C., Tichit, L., Jungsup, K., Leustek, T., Abegnac, G. 2000. The cystathionine-γ-synthase gene involved in methionine biosynthesis is highlt expressed and auxin repressed during wild strawberry (*Fragaria vesca* L.) fruit ripening. *Theoretical and Applied Genetics* **100**:1129-1136

Mazelis, M., Whatley, F., Whatley, J. 1977. The enzymology of lysine biosynthesis in higher plants; the occurrence, characterization and some regulatory properties of dihydrodipicolinate synthase. *FEBS Letters* **84**: 236–240

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L., Marshall, W.F., Qu, L.H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S.,

Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A.V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riano-Pachon, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Martinez, D., Ngau, W.C., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I.V., Rokhsar, D.S., Grossman, A.R. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science **318**(5848):245-50.

Mignone, F., Gissi, C., Liuni, S., Pesole, G. 2002. Untranslated regions of mRNAs. *Genome Biology* **3**(3):reviews0004.1-0004.10

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V., Lewis, K.L.T., Salzberg, S.L., Feng, L., Jones, M.R., Skelton, R.L., Murray, J.E., Chen, C., Qian, W., Shen, J., Du, P., Eustice, M., Tong, E., Tang, H., Lyons, E., Paull, R.E., Michael, T.P., Wall, K., Rice, D.W., Albert, H., Wang, M.-L., Zhu, Y.J., Schatz, M., Nagarajan, N., Acob, R.A., Guan, P., Blas, A., Wai1, C.M., Ackerman, C.M., Ren, Y., Liu, C., Wang, J., Wang, J., Na, J.-K., Shakirov, E.V., Haas, B., Thimmapuram, J., Nelson, D., Wang, X., Bowers, J.E., Gschwend, A.R., Delcher, A.L., Singh, R., Suzuki, J.Y., Tripathi, S., Neupane, K., Wei, H., Irikura, B., Paidi, M., Jiang, N., Zhang, W., Presting, G., Windsor, A., Navajas-Pérez, R., Torres, M.J., Feltus, A., Porter, B., Li, Y., Burroughs, A.M., Luo, M.-C., Liu, L., Christopher, D.A., Mount, S.M., Moore, P.H., Sugimura, T., Jiang, J., Schuler, M.A., Friedman, V., Mitchell-Olds, T., Shippen, D.E., dePamphilis, C.W., Palmer, J.D., Freeling, M., Paterson, A.H., Gonsalves, D., Wang, L., Alam, M. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452**:991-996

Mudd, S.H., and Datko, A.H. 1990. The *S*-methylmethionine cycle in *Lemna paucicostata*. *Plant Physiology* **93**(2):623-630

Muntz, K., Christov, V., Saalbach, G., Saalbach, I., Waddell, D., Pickardt, T., Schieder, O., Wustenhagen, T. 1998. Genetic engineering for high methionine grain legumes. *Modification by Genetic Engineering* **42**:125-127

National Center for Biotechnology Information. 2010. VecScreen. Accessed online: http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html

National Center for Biotechnology Information. 2011. BLASTP Suite. Accessed online: http://blast.ncbi.nlm.nih.gov/Blast.cgi

Negrutiu, I., Brouwer, D., Dirks, R., Jacobs, M. 1985. Amino acid auxotrophs for protoplast cultures of *Nicotiana plumbaginifolia*, Viviani. *Molecular and General Genetics* **199**(2):330-337

Nicot, N., Hausman, J.-F., Hofmann, L., Evers., D. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany* **56**(421): 2907–2914

Nikiforova, V., Kempa, S., Zeh, M., Maimann, S., Kreft, O., Casazza, A.P., Riedel, K., Tauberger, E., Hoefgen, R., Hesse, H. 2002. Engineering of cysteine and methionine biosynthesis in potato. *Amino Acids* **22**:259-278

Noji, M., Inoue, K., Kimura, N., Gouda, A., Saito, K. 1998. Isoform-dependant differences in feedback regulation and subcellular localization of serine acetyltransferase involved in cysteine biosynthesis from *Arabidopsis thaliana*. *Journal of Biological Chemistry* **273**(49):32739-32745

Ominato, K., Akita, H., Suzuki, A., Kijima, F., Yoshino, T., Yoshino, M., Chiba, Y., Onouchi, H., Naito, S. 2002. Identification of a short highly conserved amino acid sequence as the functional region required for posttranscriptional autoregulation of the cystathionine γ-synthase gene in *Arabidopsis*. *Journal of Biological Chemistry* **277**(39):36380-36386

Onouchi, H., Nagami, Y., Haraguchi, Y., Nakamoto, M., Nishimura, Y., Sakurai, R., Nagao, N., Kawasaki, D., Kadokura, Y., Naito., S. 2005. Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the *CGS1* gene of Arabidopsis. *Genes and Development* **19**:1799-1810

Onoue, N., Yamashita, Y., Nagao, N., Goto, D.B., Onouchi, H., Naito, S. 2011. *S*-adenosyl-$_L$-methionine induces compaction of nascent peptide chain inside the ribosomal exit tunnel upon translation arrest in the *Arabidopsis CGS1* gene. *Journal of Biological Chemistry* **286**(17):14903-14912

Paris, S., Wessel, P.M., Dumas, R. 2002a. Overproduction, purification, and characterization of recombinant bi-functional threonine-sensitive aspartate kinase-homoserine dehydrogenase from *Arabidopsis thaliana*. *Protein Expression and Purification* **24**(1):105-110

Paris, S., Wessel, P.M., Dumas, R. 2002b. Overproduction, purification, and characterization of recombinant aspartate semialdehyde dehydrogenase from *Arabidopsis thaliana*. *Protein Expression and Purification* **24**(1):99-104

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A., K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J., Rokhsar, D.S. 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**:551-556

Pfaffl, M.W. 2001. A New Mathematical Model for Relative Quantification In Real-Time RT-PCR. *Nucleic Acids Research* **29**(9):2002-2007.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S.A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A., Shapiro, H., Schmutz, J., Grimwood, J., Lindquist, E., Lucas, S., Grigoriev, I.V., Schmitt, R., Kirk, D., Rokhsar, D.S. 2010. Genomic Analysis of Organismal Complexity in the Multicellular Green Alga *Volvox carteri*. *Science* **329**(5988):223-226

Ravanel, S., Droux, M., Douce. R. 1995. Methionine biosynthesis in higher plants. I. Purification and characterization of cystathionine gamma synthase from spinach chloroplast. *Archives of Biochemistry and Biophysics* **316**(1):572-584

Ravanel, S., Gakiere, B., Job, D., Douce, R. 1998. Cystathionine-γ Synthase from *Arabidopsis thaliana*: Purification and Biochemical Characterization of the Recombinant Enzyme Overexpressed in *Escherichia coli*. *Biochemistry Journal* **331**:639-648

Ravanel, S., Block, M.A., Rippert, P., Jabrin, S., Curien, G., Rebeille, F., Douce, R. 2004. Methinoine metabolism in plants. Chloroplasts are autonomous for *de novo* methionine synthesis and can import S-adenosylmethinoine from the cytosol. *Journal of Biological Chemistry* **279**:22548-22557

Rebeille, F., Jabrin, S., Bligny, R., Loizeau, K., Gambonnet, B., Van Wilder, V., Douce, R., Ravanel, W. 2006. Methionine catabolism in *Arabidopsis* cells is initiated by a γ-cleavage process and leads to *S*-methylcysteine and isoleucine synthesis. *Proceedings of the National Academy of Science* **103**(42):15687-15692

Rensing, S., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.-H., Stueber, K., Theodoulou, F.-L., Tu, H., Van de Peer, Y., Verrier, P.-J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S., Boore, J.L. 2007. The *Physcomitrella* Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science* **319**(5859):64-69

Reidel, K., Mangelsdorf, C., Streber, W., Willmitzer, L., Hofgen, R., Hesse. H. 1999. Isolation and characterization of a cDNA encoding cystathionine gamma-synthase from potato. *Plant Biology* **1**:638-644

Rognes, S.E., Lea, P.J., Miflin, B.J. 1980. *S*-adenosylmethionine anovel regulator of aspartate kinase. *Nature* **287**:357-359

Rozen, S., and Skaletsky H.J. 1996, 1997, 1998. Primer3. Code available at: http://www-genome.wi.mit.edu/genome_software/other/primer3.html

Saito, K. 2004. Sulfur assimilatory mechanism. The long and smelling road. *Plant Physiology* **136**:2443-2450

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J.,

Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., Jackson, S.A. 2008. Genome sequence of the paleopolyploid soybean. *Nature* **463**(7278):178-183

Shen, B., Li, C., Tarczynski, M.C., 2002. High free-methionine and decreased lignin content result from a mutation in the Arabidopsis *S*-adenosylmethionine synthase 3 gene. *The Plant Journal* **29**(3):371-380

Steegborn, C., Messerschmidt, A., Laber, B., Streber, W., huber, B., Clausen, T. 1999. The crystal structure of cystathionine γ-synthase from *Nicotiana tabacum* reveals its substrate and reaction specificity. *Journal of Molecular Biology* **290**:983-996

Steegborn, C., Messerschmidt, A., Huber, R., Clausen, T. 2001. Crystal structures of cystathionine gamma-synthase inhibitor complexes rationalize the increased affinity of a novel inhibitor. *Journal of Molecular Biology* **311**:789-801

Stothard, P. 2000. The sequence manipulation suite: Reverse complement. Accessed online: http://www.bioinformatics.org/sms/rev_comp.html

Suzuki, A., Shirata, Y., Ishida, H., Chiba, H., Onouchi H., Naito, S. 2001. The first exon coding region of cystathionine γ-synthase is necessary and sufficient for down regulation of its own mRNA accumulation in transgenic *Arabidopsis thaliana. Plant Cell and Physiology* **42**(10):1174-1180

Tabe, L. and Higgins, T. 1998. Engineering Plant Protein Composition for Improved Nutrition. *Trends in Plant Science* **3**(7):282-286

Takahashi, H., Kopriva, S., Giordano, M., Saito, K., Hell, R. 2011. Sulfur assimilation in photosynthetic organisms: Molecular functions and regulations of transporters and assimilatory enzymes. *Annual Reviews in Plant Biology* **62**:157-184

Tang, G., Zhu-Shimoni, J.X., Amir, R., Zchori, I.B.-T., Galili, G. 1997. Cloning and expression of an *Arabidopsis thaliana* cDNA encoding a monofunctional aspartate kinase homologous to the lysine-sensitive enzyme of *Escherichia coli. Plant Molecular Biology* **34**:287-294

Tewari-Singh, N., Sen, J., Kiesecker, H., Reddy, V.S., Jacobsen, H.J., Guha-Mukherjee, S. 2004. Use of a herbicide or lysine plus threonine for non-antibiotic selection of transgenic chickpea. *Plant Cell Reports* **22**(8):576-583

Thoen, A., Rognes, S.E., Aarnes, H. 1978. Biosynthesis of threonine from homoserine in pea seedlingsL homoserine kinase. *Plant Science Letters* **13**:103-112

Thompson, G.A., Datko, A.H., Mudd, H., Giovanelli. J. 1981. Methionine biosynthesis in *Lemna*: Studies on the regulation of cystathionine gamma-syntahse, *O*-phosphohomoserine sulfhydrolase, and *O*-acetylserine sulfhydrylase. *Plant Physiology* **69**:1077-1083

Tuskan, G.A, DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, A., dePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, S., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, H., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer10, Y., Rokhsar, D. 2006. The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**(5793):1596-1604

Ufaz, S. and Galili, G. 2008. Improving the content of essential amino acids in crop plants: Goals and opportunities. *Plant Physiology* **147**:954-961

United States Department of Agriculture. 2011. International Peanut Genomic Research Initiative. Strategic plant for 2012 – 2016 – Characterization of the peanut genome. Accessed online: http://www.peanutbioscience.com/images/IPGRI_StratPlan_v4_3_Sep11b.pdf

Wallsgrove, R.M., Lea, P.J., Milfin, B.J. 1983. Intracellular localization of aspartate kinase and enzymes of threonine and methionine biosynthesis in green leaves. *Plant Physiology* **71**(4):780-784

Wanatabe, M., Hubberten, H.-M., Saito, K., Hoefgen, R. 2010. General regulatory patterns of plant mineral nutrient depletion as revealed by *serat* quadruple mutants disturbed in cysteine synthesis. *Molecular Plant* **3**(2):438-466

Wang, T.L., Domoney, C., Hadley, C.L., Casey, R., Grusak, M.A. 2003. Can we improve nutritional quality of legume seeds? *Plant Physiology* **131**:886-891

Wilkie, S.E., Roper, J.M., Smith, A.G., Warren, M.J. 1995. Isolation, characterization and expression of a cDNA clone encoding plastid aspartate aminotransferase from *Arabidopsis thaliana. Plant Molecular Biology* **27**(6):1227-1233

Winter, D., Vinegar, B., Nahal, H., Ammer, R., Wilson, G.V., Provart, N.J., 2007. An "Electronic Fluorescent Pictogram" browser for exploring and analyzing large-scale biological data sets. *PLOS ONE* **2**(8):e718

Wirtz, M., Droux, M., Hell, R. 2004. *O*-acetylserine (thiol) lyase: and enigmatic enzyme of plant cysteine biosynthesis revisited in *Arabidopsis thaliana. Journal of Experimental Botany* **55**(404):1781-1798

Xaio, H., Tattersall, E.A., Siddiqua, M.K., Cramer, C.R., Nassuth, A. 2008. CBF4 is a unique member of the CBF transcription factor family of *Vitis vinifera* and *Vitis riparia. Plant, Cell and Environment* **31**(1):1-10

Xu, X., Pan, S.K., Cheng, S.F., Zhang, B., Bachem, C.W.B., de Boer, J.M., Borm, T.J.A., Kloosterman, B.A., van Eck, H.J., Datema, E., Goverse, A., van Ham, R.C.H.J., Visser, R.G.F. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* **475**:189-195

Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., Yang, H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **5**:79-92

Zbio.net. 2006. Six-frame translation. Accessed online: http://molbiol.ru/eng/scripts/01_13.html

Zeh, M., Casazza, A.P., Kreft, O., Roessner, U., Biebrich, K., Willmitzer, L., Hoefgen, R., Hesse, H. 2001. Antisense inhibition of threonine synthase lease to high methionine content in transgenic potato plants. *Plant Physiology* **127**(3):792-802

Zhou, M.-Y., Gomez-Sanchez, C.E. 2000. Universal TA cloning. *Current Issues in Molecular Biology* **2**(1):1-7

APPENDIX 1 – LISTS OF PRIMERS

Table 1 – Primers used for PCR amplification and sequencing

Name	Methods Section	Sequence	T_M (°C) [a]
First Strand cDNA Synthesis			
$dT_{(18)}$	2.3, 2.7, 2.8	TTTTTTTTTTTTTTTTTT	34.3
polyT-adpt	2.7	GTCGGATCCACCTGGAGATCG $T_{(18)}$	61.5
Colony PCR and Sequencing			
pGEM-SeqF	2.4, 2.5, 2.7, 2.13	CGACGGCCAGTGAATTGTAATACG	58.0
pGEM-SeqR	2.4, 2.5, 2.7, 2.13	CTCAAGCTATGCATCCAACGCGT	59.9
pSEQ-seq0	2.4, 2.8, 2.9	GGCGTCAGGCAGCCATCGGAAGCTG	67.8
pSEQ-seq7r	2.4, 2.8, 2.9	GCCCGCCACCCTCCGGGCCGTTGCTTCGC	76.2
AtCGS-intF	2.9	CCCTACGACTGTGGTACTTGAAG	56.8
AtCGS-intR	2.9	TCAGCAGGATCAATCACAGTGACA	58.2
cCGS-intF	2.8	GCGAGCAAAACGATTCATGCCGCTG	63.2
cCGS-intR	2.8	TCATGATGTCCACCCATGTACTTTG	57.0
pCGS-intF	2.8	TAGTTATGAATATGGGCGCTATGGG	56.4
pCGS-intR	2.8	AAATTTGGCCATCCCCATTCCAGT	59.5
tCGS-intF	2.8	GGCTAGTTATGAATATGGGCGCT	57.0
tCGS-intR	2.8	GCAGGGTTTTCAAGCCTCTGATGA	59.5
pSP64-SEQ-F	2.14	ACA ATT TCA CAC AGG AAA CAG CTA TGA	57.0
M13-	2.14	CAG GAA ACA GCT ATG AC	47.0
LUC-SEQ-R	2.14	AGGAACCAGGGCGTATCTCTTCATAG	59.5

Name	Methods Section	Sequence	T_M (°C) [a]
Degenerate Primer Cloning			
pCGS-F2[b]	2.7	AAGCTTAAYTGYAGCAAYATYGGYGT	53.5 – 63.2
pCGS-42[b]	2.7	AAGCTTCTYAAYCCDAAYGCTGCRTAYCT	56.4 – 66.4
CGS-R5[b]	2.7	AAGCTTTCTCCATCAAYCTCRAAACTGAC	57.5 – 60.4
CGS-R7[b]	2.7	AAGCTTCTAKATDKYYTCSAGRGCYTG	53.2 – 65.4
Amplification of the 3' End of CGS from Pulse Species			
Adapter	2.7	GTCGGATCCACCTGGAGATCG	59.2
cCGS-cF4	2.7	GTCAGTTTTGAGATAGATGGAGA	51.8
pCGS-fF4	2.7	GTTTTGGAATTGAAGATTTTGAGGA	52.4
tCGS-tF4	2.7	CTTAACCCGAATGCTGCATACCT	57.3
Genome Walker for 5' End Amplification of CGS from Pulse Species			
GW-A1	2.7	GTAATACGACTCACTATAGGGC	51.9
GW-A2	2.7	ACTATAGGGCACGCGTGGT	58.8
Ca-GW1	2.7	CAACGGCGACAACATTTTCGTCGGCGAC	65.4
Ca-GW2	2.7	GGAGGCGCTGTAGTCGCATCAACAGTTG	64.5
Ps-GW1	2.7	GTACAGCCCCATTTACCCCAGCATCGAC	64.2
Ps-GW2	2.7	GTCATCGGCGTCGAGGTCTAAGGGGA	65.0
Lc-GW1	2.7	CTACCGAGTCTCTCAGCGGCATGAATCG	63.5
Lc-GW2	2.7	CAAGGTCTACGGGGAGCGGTGCAGTC	66.3
Cloning of Full Length Plant CGS and Truncations for Complementation			
cCGS-start	2.8	GACATATGGCCGTTTCAAGTTTCCACCG	61.6
cCGS-stop	2.8	GGGAGCTCCTATATGGCTTCGAGAGCTTGCAG	65.4
cCGS-Q60	2.8	GAA CATATG CAGCGTCAAC TCAGCACCAA AGCTC	64.8

Name	Methods Section	Sequence	T_M (°C) [a]
pCGS-start	2.8	GACATATGGCCGTTTCAAGCCTCCAC	61.7
pCGS-stop	2.8	GGGAGCTCCTAGATGGTCTCGAGGGCTTGCA	68.1
pCGS-Q58	2.8	GAA CATATG CAGC GTCAGCTCAG TACTAAAGCT C	62.6
tCGS-start	2.8	GACATATGGCCGTTTCAAGCCTCCAC	61.7
tCGS-stop	2.8	GGGAGCTCCTATATGACTTCCAGTGCCTGCACG	66.6
tCGS-Q58	2.8	GAACATATGCAGCGTCAGCTAAGCACTAAAGCTC	63.1
aCGS-Fc-NdeI	2.9	GTACCATATGGCCGTCTCATCATTCC	59.0
aICGS-NdeI-start	2.9	GTACCATATGGTCGCTCAGCTGAGCA	61.2
aCGS-BanHI-stop	2.9	GTACGGATCCTCAGATGGCTTCGAGA	61.3
Amplification Over the Exon 1/2 Boundary in Pulse Crop CGS			
cCGS-cR5	2.13	AGG AGT GGT AAT TCC ATC AGT CAC G	58.3
pCGS-fR1	2.13	CCCATAGCGCCCATATTCATAACTA	56.4
tCGS-tR5	2.13	GCAGGGTTTTCAAGCCTCTGATGA	59.5
Pulse Crop Exon1::Luciferase Fusion Constructs			
Luc-Start-XmaI	2.14	AGCCCGGGATGGAAGACGCCAAAAACATAAAGA	65.6
Luc-KZ-XbaI	2.14	CATCTAGAGCCACCATGGAAGACGCCAAAAACA	64.4
Luc-MAVS-XbaI	2.14	CATCTAGAGCCACCATGGCCGTCTCAATGGAAGACGCCAAAAACA	69.3
Luc-Stop-SacI	2.14	GAGAGCTCCAATTTGGACTTTCCGCCCTTCT	64.6
AtEx-kz-XbaI	2.14	CATCTAGAGCCACCATGGCCGTCTCATCATTCCAGT	66.8
AtEx-HAG-XmaI	2.14	CACCCGGGACCGGCATGAACAGTGAGGC	69.6
CpEx-HAA-XmaI	2.14	CATCTAGAGCCACCATGGCCGTTTCAAGTTTCCACC	66.6
Ca-Exon1-R	2.14	CACCCGGGAGCGGCATGAATCGTTTTGCTTG	67.6
PtEx-kz-XbaI	2.14	CATCTAGAGCCACCATGGCCGTTTCAAGCCTCCACC	68.9

Name	Methods Section	Sequence	T_M (°C) [a]
LtEx-HAA-XmaI	2.14	CACCCGGGAGCGGCATGAATCGTTTTACTTG	65.6
At-MTO1-F	2.14	CAACATCAGTGTTGCACAGATC	54.4
At-MTO1-R	2.14	GATCTGTGCAACACTGATGTTG	54.4
Cpt-MTO1-F	2.14	CAACATCAGTGTCGCGCAAATC	57.3
Cpt-MTO1-R	2.14	GATTTGCGCGACACTGATGTTG	57.3

[a] As determined by the online tool "OligoAnalyzer 3.1" (Integrated DNA Technologies, 2011)

[b] Degenerate base notation: R = A, G; Y = C, T; M = A, C; K = G, T; S = C, G; W = A, T; H = A, C, T; B = C, G, T; V = A, C, G; D = A, G, T (Integrated DNA Technologies, Coralville, USA)

Table 2 - Primers used for investigation of pulse CGS expression levels over plant development by real time PCR

Target	Species	Name	Sequence	T_M (°C) [a]	Product Length (bp)	Product T_M (°C)
CGS	*E. coli*	eCGS-RT-F	AAATACCTGCAAACCCAGCCGTTG	60.4	315	88.7
		eCGS-RT-R	TTCGCCATCTTCAATACCGGTGGA	60.4		
	Chickpea	cCGS-RT-F	TGATCCTGCTGATGTTGGAGCCTT	60.3	350	84.0
		cCGS-RT-R	ATGCAGCATTCGGGTTAAGAGCAC	59.7		
	Field pea	pCGS-RT-F	CACATCCCAAGGTGAAGCGTGTAT	59.0	294	83.4
		pCGS-RT-R	CTCCGAAGCTGAAGCGAACCAAAT	59.5		
	Lentil	tCGS-RT-F	ATCCCAAGGTGAAGCGTGTCTACT	59.9	170	84.0
		tCGS-RT-R	AGGTTGGTCCACAATGCTCTCACA	60.6		
EF1α	Chickpea	cEF1a-RT-F	TTGACCAGATCAACGAGCCCAAGA	60.0	286	86.0
		cEF1a-RT-R	GATGCAACGAAACCACGCTTGAGA	59.8		
	Field pea	pEF1a-RT-F	TTGAAACTGGTGTTGTGAAGCCCG	59.9	264	85.0
		pEF1a-RT-R	TTCCAATCTGTCCAGGGTGGTTCA	60.0		
	Lentil	tEF1a-RT-F	TTGAAACTGGTGTTGTGAAGCCCG	59.9	263	85.2
		tEF1a-RT-R	ACCAATCTGTCCTGGGTGGTTCAT	60.1		

[a] As determined by the online tool "OligoAnalyzer 3.1" (Integrated DNA Technologies, 2011)

Table 3 - Primers for genotyping of *Arabidopsis thaliana* T-DNA insertion lines

Line	Name	Sequence	T_M (°C) [a]	L/R Product Length (bp)	B3/R Product Length (bp)
Insert specific	B	ATTTTGCCGATTTCGGAAC	51.5		
At1g33320					
SALK_058208.23.80	Left	AAATGCATCATTTTGTATTGTGG	50.6	1053	564
	Right	TTGTTGGGTTTCCATAACGAG	53.1		
SALK_128836.55.50C	Left	ATGGTGAACAAGATCAGCACC	55.4	1246	633
	Right	CCTCGCTCATCACTTTCAGTC	55.5		
At3g01120					
SALK_001032.53.70	Left	GTTGCTACAGTTTCTACGGGC	56.0	1011	557
	Right	TTGGACCCAGACCATTATCAC	54.4		
SALK_014068.56.00C	Left	AAAGCATTGTTGACCAACCTG	54.1	1148	595
	Right	TGCACAACGTTTTCTGATCTG	53.7		
SALK_045913.36.85C	Left	GACAAGATCAGCACCAAGAGC	56.0	1119	545
	Right	TACAGCTGACGGATTCCAAAC	55.0		
SALK_049643.17.95	Left	TCGCGGAACAATAGTTCAAAG	53.3	1157	665
	Right	GCTCTTGGTGCTGATCTTGTC	56.0		
SALK_064727.38.85	Left	CATTGGGATTCAAAAACAACG	50.8	1147	635
	Right	TTTGGAGGTGTGGTCAGTTTC	55.2		
SALK_124839.35.05	Left	TAATTGCACTTGCCTTTCCTG	53.7	1061	536
	Right	TTGCATACCTTAGGATGTGCC	55.0		

[a] As determined by the online tool "OligoAnalyzer 3.1" (Integrated DNA Technologies, 2011)

APPENDIX 2 - BIOINFORMATICS TOOLS AND SETTINGS AND PRIMER DESIGN PARAMETERS

Table 1 – Bioinformatics tools and settings

Tool	Use	Section	Settings		Reference
VecScreen	Removal of vector sequence from sequencing results	2.6, 2.7, 2.8, 2.9, 2.13, 2.14	Gap open Gap extend Nucleotide mismatch Nucleotide match Expect value Word size Dropoff extension	3 3 -5 1 700 11 $1.75e^{12}$	National Center for Biotechnology Information, 2010
ClustalW	Multiple sequence alignment – nucleotide	2.6, 2.7, 2.8, 2.9, 2.13, 2.14	DNA weight matrix Gap open Gap extend Gap distance No end gaps Iteration Numiter Clustering	IUB 10 0.2 5 NO NONE 1 NJ	Larkin *et al*, 2007
BLASTP Genome Databases	Querying genome databases	2.10	Expect value Comparison matrix Word length Allow gaps	-1 BLOSUM62 3 YES	Table 4

Tool	Use	Section	Settings	Reference
BLASTP National Center for Biotechnology Information	Protein sequence identification	2.10	Search set – non-redundant protein sequences Organism taxid 3702 Expect value 10 Word size 3 Matrix BLOSUM 62 Gap existence 11	National Center for Biotechnology Information, 2011
Mobyle Archaeopteryx	Phylogenetic tree generation	2.10	Sequence type Protein Gap Open 10 Gap Extend 5	Han and Zmasek, 2009
Ka/Ks Calculation Tool	Ka/Ks ration calculations	3.1	Weighting none Tree method parsimony Submatrix discrete_Grantham LI rate moderate	Bergen Center for Computational Science, 2011

Table 2 – Parameters for Genome Walking primer design[a]

Parameter	Acceptable	Ideal Primer
Length (base pairs)	26-30	27
T_M (°C)	> 60.0	67.0
G/C Content (%)	40-60	50
Hairpin (°C)	< 40	< 10
Self Dimer (-ΔG, kcal/mole)	> -7.0	> -3.0

[a] Integrated DNA Technologies, 2011

Table 3 - Parameters for real time PCR primer design in Primer Quest[a] and ideal primers.

Primer	Parameter	Primer Quest Design	Ideal Primer
Individual	Length (base pairs)	22 – 26	24
	T_M (°C)	58.0 – 62.0	60.0
	G/C Content (%)	45 – 55	50
	Hairpin (°C)	N/A	< 40.0
	Self Dimer (-ΔG, kcal/mole)	N/A	> -7.0
Pair	Product Length (base pairs)	200 – 350	250
	Heterodimer (-ΔG, kcal/mole)	N/A	> -7.0

[a] Rozen and Skaletsky, 1996-1998

APPENDIX 3 – CGS SEQUENCES FROM TARGET PULSE SPECIES

3.1 CHICKPEA

All sequences compiled using CAP3 (Huang and Madan, 1999). The 5' genomic sequence and 3' untranslated regions, from the mRNA, are underlined. The translation initiation and stop codons are bolded. Intron sequences and 5' genomic sequence from genomic DNA are highlighted in grey. Sequence with coverage from both genomic DNA and cDNA amplification is in red text. Sequence with coverage from only cDNA amplification is in *italics*.

TACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTAAAATAAAATAATATAAATAAGA
GGTAAAATTTAATAGAAAAACTATTACACCCTACTATTCTACCTCCACTACAATTGGAAA
TTGTTTAATTTTTTTTTTTATGACAACTACAATTCAACTTGTATTAATAAGGTTCAAAGT
AAATAGTGACTAAATTTGAAAATTAATAAACAAAATGAGTGAGTAAAGTAACAAACCGAC
AAAGAAGGGGACGCGTATGCCAGGTGTAAGTTACTATTAATATTAAATCATCACTAGTTA
ATATAAGGTTTGAATTCAACAAGGTTATTAAAATAAAACCAAAAGTTCCCCGAGAAGATA
TATTCATTCATTTGTGTTAGTTTGGTACAGAAGGGATAGTAGGTGTAGGAAGGAACAGAA
ACACAAAAACTCACCAAACTATCAAAACGACGACGAAAATG

ATATTAAATTTCATTTTCAATGATTAGTTCCTTTTATTAGGTT
TTCGTATTAATTGTTTCAGTTACATATTTTATCCTCTTTTTTCAAAATTAGGTTTTTTGG
GTTTATATTATTTAAATTTTATATTATGTGACTAGTTTACAAATATATTGTTTTTTTTATT
GAAGTATGTTTTTGTTTACTAAGTT

GTATACGATTCACTGATTTGGATCTTTAGTATGGTGCTTGAAAAACTG
TTATTTTTAAATGTTGTCAGTGGTTCTTTAGGTAACTCAAAGAGTAGTGCAAATGTTTTA
ATTGAAAGCGTGGTTAAAGTTTGTTTTGCATTTTCATTTTCATATTCTGCTCCAACTTGT
ATTCCCTGGGATTTTTTATTAGAGTAATAAAAACTCTGCTTTGAAACTTGAAAGTGAATG
GTTGATAGAGGCACTTAGTGTACATATCTTGCATAAAAATTTGTTGGGGTACTTAATGCT
TAGATGTCTTTTATGCTTCCTTTCTGAACTGGGCACTTACAGCTATAAGTATAATTTGTA
ACAATTTTTATTTTTCTGAGCAG

GTAAGTGCTGGGGAGAGTAATTGGTTTCCTTTTTGT
GCATTTCTCTTATTTATTCAAATTAATAGATAGTGTTGAATTCAGATGGTATTATTCATA
CGTGCAG

TGTCCTTGGTGGATGCATAAGTGGCTCTGA
TAAGTTGATTTCACAAATTCGAATTTTGCATCATATTTTGGGCGGTGCTCTTAACCCGAA
TGCTGCATACCTATTCATCAGAGGCATGAAAACATTGCATCTTCGTGTACAGCAGCAGAA
TTCAACCGGAATGAGGATGGCCAAACTTTTAGAGGCACATCCCAAGGTGAAGCGTGTATA
CTATCCAGGCTTGCCTAGTCATCCTGAACACGAGCTTGCCATGAGGCAAATGACTGGTTT
TGGGGGCGTTGTCAGTTTTGAGATAGATGGAGATCTAACAACCACAATAAAATTTGTGGA

TTCACTTAAAATCCCATATATTGCAGCCTCTTTTGGTGGCTGTGAGAGCATTGTGGATCA
ACCGGCCATTTTGTCTTACTGGGATCTTCCTGCATCAGAAAGGGCCAAATATAAGATTTA
TGACAATTTGGTTCGCTTCAGCTTCGGAGTTGAAGATTTTGAGGATTTGAAGGCTGATAT
*CCTGCAAGCTCTCGAAGCCATA**TAG**GCAGTATTCATTGTTCTCCCCAAGCTTTTTTCTTT*
TTCAGTCTTCAATCGTGTTTTACGTTCTTTTGTTTGAACATAAATCGGGTACTCGTTGCA
TGTAATTGCAGTTATTATTTTTGCAAATTGAGTAGAGTGTTTTTTGTTAAAAAAAAAAAA
AAAAAAAAAAA

Nucleotide coding sequence, untranslated regions and exons removed. The translation initiation and stop codons are underlined and bold.

ATGGCCGTTTCAAGTTTCCACCGCGTTTTAACTTTTGAGTGTCGTTCAGATCCCGATTTC
GCATCTCTTCCTTCCTTCGACAACCGTCGTACTCGTCATTTCCCAACCGCTGCCGTTTCC
GTCGCGTTTTACGCCGCTTCATCTCCAATCCTTCGTTTTCCTCCCAACTTTCAGCGTCAA
CTCAGCACCAAAGCTCGTCGTAACTGTAGCAACATCGGTGTCGCGCAAATCGTAGCTGCT
TCGTGGTCTAATGAAGGTTCCGGTAACCCCGCCGCTGGGGCTCCTCCGGTGCCTGCCGTT
GCATCAACTGTTGATGCGACTACAGCGCCTCCCCCGTTAGACCTCGACGCCGTCGCCGAC
GAAAATGTTGTCGCCGTTGATGAAAATGGGGCTGTACAGACTAACCGTAGTTCTTATTCT
AAGTTTTTGAAATCCGATGCAAGCAAAACGATTCATGCCGCTGAAAGACTAGGTAGGGGT
ATCGTGACTGATGGAATTACCACTCCTGTGGTGAACACTTCTGCCTACTTTTTTAAGAAA
ACCGCTGATCTTATTGATTTTAAGGAGAAACGTCAGACTAGTTATGAATATGGGCGCTAT
GGCAATCCAACCAGTACTGTTTTAGAGGAGAAGATAAGTGAACTCGAAGGAGCTGAATCA
ACTGTCTTAATGGCATCTGGAATGTGTGCTAGCATAGTCTTGTTGATGGCGCTGGTTCCA
GCTGGTGGACATCTTGTGACTACTACAGATTGTTACCGGAAGACTAGGATTTTTATAGAG
ACTGTGCTTCCAAAGATGGGGATCACGACCACCGTGGTTGATCCTGCTGATGTTGGAGCC
TTGGAATCTGCATTGGAGCAGAACAATGTGTCTCTTTTCTTCACTGAGTCTCCTACCAAT
CCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTTTGCCATAAGCATGGGGCT
TTGCTCTGTATTGATGGTACATTTGCAACTCCTTTGAACCAGAAGGCCCTTGCCCTTGGT
GCTGATTTGATTATGCACTCTTGTACAAAGTACATGGGTGGACATCATGATGTCCTTGGT
GGATGCATAAGTGGCTCTGATAAGTTGATTTCACAAATTCGAATTTTGCATCATATTTTG
GGCGGTGCTCTTAACCCGAATGCTGCATACCTATTCATCAGAGGCATGAAAACATTGCAT
CTTCGTGTACAGCAGCAGAATTCAACCGGAATGAGGATGGCCAAACTTTTAGAGGCACAT
CCCAAGGTGAAGCGTGTATACTATCCAGGCTTGCCTAGTCATCCTGAACACGAGCTTGCC
ATGAGGCAAATGACTGGTTTTGGGGGCGTTGTCAGTTTTGAGATAGATGGAGATCTAACA
ACCACAATAAAATTTGTGGATTCACTTAAAATCCCATATATTGCAGCCTCTTTTGGTGGC
TGTGAGAGCATTGTGGATCAACCGGCCATTTTGTCTTACTGGGATCTTCCTGCATCAGAA
AGGGCCAAATATAAGATTTATGACAATTTGGTTCGCTTCAGCTTCGGAGTTGAAGATTTT
GAGGATTTGAAGGCTGATATCCTGCAAGCTCTCGAAGCCATA**TAG**

Translated amino acid sequence. The truncation site for the removal of the putative plastid targeting sequence is underlined and bold. The putative MTO1 regulatory region is highlighted in grey. The exon 1/2 boundary is indicated by ^.

MAVSSFHRVLTFECRSDPDFASLPSFDNRRTRHFPTAAVSVAFYAASSPILRFPPN**FQ**RQ
LSTK[redacted]VAASWSNEGSGNPAAGAPPVPAVASTVDATTAPPPLDLDAVAD
ENVVAVDENGAVQTNRSSYSKFLKSDASKTIHA^AERLGRGIVTDGITTPVVNTSAYFFK
KTADLIDFKEKRQTSYEYGRYGNPTSTVLEEKISELEGAESTVLMASGMCASIVLLMALV
PAGGHLVTTTDCYRKTRIFIETVLPKMGITTTVVDPADVGALESALEQNNVSLFFTESPT
NPFLRCVDIKLVSELCHKHGALLCIDGTFATPLNQKALALGADLIMHSCTKYMGGHHDVL
GGCISGSDKLISQIRILHHILGGALNPNAAYLFIRGMKTLHLRVQQQNSTGMRMAKLLEA
HPKVKRVYYPGLPSHPEHELAMRQMTGFGGVVSFEIDGDLTTTIKFVDSLKIPYIAASFG
GCESIVDQPAILSYWDLPASERAKYKIYDNLVRFSFGVEDFEDLKADILQALEAI

3.2 FIELD PEA

All sequences compiled using CAP3 (Huang and Madan, 1999). The 5' genomic sequence and 3' untranslated regions, from the mRNA, are underlined. The translation initiation and stop codons are bolded. Intron sequences and 5' genomic sequence from genomic DNA are highlighted in grey. Sequence with coverage from both genomic DNA and cDNA amplification is in red text. Sequence with coverage from only cDNA amplification is in *italics*.

AAATCGAAAATCTATAACGATTGGTACAACCAAATACATATGTCCTATATCCTAATTAAA
ATTAATAAGTTGAATATTTTTAACATTATGAGTAGAGTTTGATTATTATGAGTAGTATAA
AATAATTTTATAATATTATTTTATAAAAAAATATTATAGTTTGATGTCATCGGATATACG
ATGGTAATTATTTTTTATTTTTTATTATTAGTGTATATTCTATTTTCACAACCGACAAAG
AATCCACGCGTATATTTAACTATTAATAAGGGTTGAATTGAATGAACAAAAGTTAGAATT
CAGAGGGAGTAGGTACAGAAAAAAAAAAGAAAAAAACCTAGCCTCACCAATCCCAAAACG
ACGTCAAAA**ATG**[illegible]**GGTATGCGA**
ATTTTTTTCTTTGTTTTTTGTAATAAATTAATACTGATTCAAATTTGAGTTTCCATGATT
AGGGGTTTGATAAGATTTAATCCTGTAGAATTTTCATTTTTTTTTGAGTTTAGCAACTAT
ATTCAATTTCAAATCCTATTATTGATTATGAATGGGTTTAATTGGATGTGACTTGCTTTG
CTTATAATTCTTTTTGGATTTTATGTGTTCTCGTTCAGATTTTGATTTTGTCTCAAGATT
TTTGTTTTGTATGAAGTTACTTTCATTTGTTTGTAAATTATTGTTAGTGGTTTCGTTTTT
GCTAATTTGATTGATTAGTTTTTGTATTTGTGTTACCATGCA[illegible]
[illegible]
AATACTGCC[illegible]**GTATTGTTTGTTGGGTTGTTTCGATTTTACCTG**
TGCAGTGAATATGTGAGGTGGAAGAATGAGGAATAGAATGTGAATGACTAGTGTGTTATG
TAGCATGTGCCGGAGTAGGCTCGACTGTGATTCCGTTCATTTGTTCATTCCTTCATTCAC
TTCCGTTTTCTCAAACAAATATTTGTGATGTGTTAGTGTTGTGTCTGTTTATATAGTGTG
GATAAAGTTTTCATCATTTTGCAG[illegible]
[illegible]*AATCCAACCAGTACTGTTTTAGAGGAGAAATTAAGTGAACTGGAAGGGGCTGAGTCA*
ACTGTCTTATTGGCATCTGGAATGTGCGCTAGCATAGTTATGCTCATGGCGCTGGTTCCA
GCCGGTGGACATCTTGTGACTACAACTGATTGTTACAGGAAGACTAGGATATTTATAGAG
ACTGTGCTTCCGAAGATGGGGATCACGGCCTCTGTGATTGATCCTGCTGATGTTGGAGCC
TTGGAATCTGCATTGGAGCAGAACAAGGTGTCTCTATTCTTCACCGAGTCTCCTACCAAT
CCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTTTGCCACAAGCATGGGGCT
TTGCTCTGCATTGATGGTACATTTGCTACGCCTTTGAACCAGAAGGCCCTTGCCCTTGGT
GCTGATTTGGTTATGCACTCTTGTACGAAGTACATTGGTGGACATCATGATGTAATTGGT
GGTTGCATAAGTGGCTCTGCTAAGCTGATTTCAGAAATTCGTATTTTGCATCATATTTTG
GGCGGAGTTCTTAACCCGAATGCTGCATACCTATTCATCAGAGGCATGAAAACACTGCAT
CTTCGTGTACAGCAGCAGAATTCAACTGGAATGGGGATGGCCAAATTTTTAGAGGCACAT
CCCAAGGTGAAGCGTGTCTACTATCCAGGCTTGCCAAGTCATCCTGAACATGAGCTTGCC
ATGAGGCAGATGACTGGTTTTGGAGGTGTCGTCAGTTTTGAGATTGATGGAGATCTAGAA
ACCACAATAAAATTTATTGATTCACTGAAAATCCCATATATTGCTCCCTCTTTTGGTGGC
TGTGAGAGCATTGTGGACCAACCAGCCATTTTGTCTTACTGGGATCTTCCTGCATCTGAA
AGGGCCAAGTGGAAAATTTATGACAATTTGGTTCGCTTCAGTTTTGGAATTGAAGATTTT
*GAGGATTTGAAGGCTGATATCGTGTGCAAGCCCTCGAGACCATC***TAA***GCAGCAGAGTTCA*
TTATTCACCGCAAGCTATTTTCCTTTTCTTGTTTATTTAATCGTGTTTGGTTCCTGTATG

TGCGGTGTGCAACTGCAGGTATTAATCAGTTTCTAGTGAGTTGAGGGTGTTTCTTGTTTC
TTGTTTGTTTGGTGAACCTAGTGAATTATGATTGCCTAGTGAAATCTGTTGGGTCTGTTT
AAAAAAAAAAAAAAAAAAA

Nucleotide coding sequence, untranslated regions and exons removed. The translation initiation and stop codons are underlined and bold.

ATGGCCGTTTCAAGCCTCCACCGCGTCTTCACCTTCGAGTGCCGCTCCGATCCCGATTTC
GCCTCCCTCCCATCCACCGAAAACCTCCGTAGCAGCCGCCGTCATTTCCCCGCCGCCGGT
CTCTCCGCCGCGTTCTACGGCATCTCTTCTCCAATTCTTCGTTTCCCCCCAAACTTTCAG
CGTCAGCTCAGTACTAAAGCTCGTCGTAACTGCAGCAACATCGGTGTCGCGCAAATCGTC
GCTGCTTCGTGGTCTAATGAAGGGATCGGCCACCCCGCCGCTGGGGTCCCACCGGTGCCG
GCCGTTGCATCAGCTGTTGATGCTGCGACGGCGCCGCTCCCCTTAGACCTCGACGCCGAT
GACGAGAACGTTGTCGATGCTGGGGTAAATGGGACTGTACAGACTAACAGTAGGTCTTAT
TCTTCGTTTTTGAAAACCGATGCAAGCAAAACGATTCATGCCGCTGAAAGACTTGGTAGG
GCTATTGTGACTGATGCAATTACCACTCCTGTGGTGAACACTTCTGCCTACTTTTTCAAG
AATACTGCCGATCTTATTGATTATAAGGAGAAACGTCTGGCTAGTTATGAATATGGGCGC
TATGGGAATCCAACCAGTACTGTTTTAGAGGAGAAATTAAGTGAACTGGAAGGGGCTGAG
TCAACTGTCTTATTGGCATCTGGAATGTGCGCTAGCATAGTTATGCTCATGGCGCTGGTT
CCAGCCGGTGGACATCTTGTGACTACAACTGATTGTTACAGGAAGACTAGGATATTTATA
GAGACTGTGCTTCCGAAGATGGGGATCACGGCCTCTGTGATTGATCCTGCTGATGTTGGA
GCCTTGGAATCTGCATTGGAGCAGAACAAGGTGTCTCTATTCTTCACCGAGTCTCCTACC
AATCCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTTTGCCACAAGCATGGG
GCTTTGCTCTGCATTGATGGTACATTTGCTACGCCTTTGAACCAGAAGGCCCTTGCCCTT
GGTGCTGATTTGGTTATGCACTCTTGTACGAAGTACATTGGTGGACATCATGATGTAATT
GGTGGTTGCATAAGTGGCTCTGCTAAGCTGATTTCAGAAATTCGTATTTTGCATCATATT
TTGGGCGGAGTTCTTAACCCGAATGCTGCATACCTATTCATCAGAGGCATGAAAACACTG
CATCTTCGTGTACAGCAGCAGAATTCAACTGGAATGGGGATGGCCAAATTTTTAGAGGCA
CATCCCAAGGTGAAGCGTGTCTACTATCCAGGCTTGCCAAGTCATCCTGAACATGAGCTT
GCCATGAGGCAGATGACTGGTTTTGGAGGTGTCGTCAGTTTTGAGATTGATGGAGATCTA
GAAACCACAATAAAATTTATTGATTCACTGAAAATCCCATATATTGCTCCCTCTTTTGGT
GGCTGTGAGAGCATTGTGGACCAACCAGCCATTTTGTCTTACTGGGATCTTCCTGCATCT
GAAAGGGCCAAGTGGAAAATTTATGACAATTTGGTTCGCTTCAGTTTTGGAATTGAAGAT
TTTGAGGATTTGAAGGCTGATATCGTGCAAGCCCTCGAGACCATC**TAA**

Translated amino acid sequence. The truncation site for the removal of the putative plastid targeting sequence is underlined and bold. The putative MTO1 regulatory region is highlighted in grey. The exon 1/2 boundary is indicated by ^.

MAVSSLHRVFTFECRSDPDFASLPSTENLRSSRRHFPAAGLSAAFYGISSPILRFPPN**FQ**
RQLSTK[redacted]VAASWSNEGIGHPAAGVPPVPAVASAVDAATAPLPLDLDAD
DENVVDAGVNGTVQTNSRSYSSFLKTDASKTIHA^AERLGRAIVTDAITTPVVNTSAYFF
KNTADLIDYKEKRLASYEYGRYGNPTSTVLEEKLSELEGAESTVLLASGMCASIVMLMAL
VPAGGHLVTTTDCYRKTRIFIETVLPKMGITASVIDPADVGALESALEQNKVSLFFTESP
TNPFLRCVDIKLVSELCHKHGALLCIDGTFATPLNQKALALGADLVMHSCTKYIGGHHDV
IGGCISGSAKLISEIRILHHILGGVLNPNAAYLFIRGMKTLHLRVQQQNSTGMGMAKFLE
AHPKVKRVYYPGLPSHPEHELAMRQMTGFGGVVSFEIDGDLETTIKFIDSLKIPYIAPSF
GGCESIVDQPAILSYWDLPASERAKWKIYDNLVRFSFGIEDFEDLKADIVQALETI

3.3 LENTIL

All sequences compiled using CAP3 (Huang and Madan, 1999). The 5' genomic sequence and 3' untranslated regions, from the mRNA, are underlined. The translation initiation and stop codons are bolded. Intron sequences and 5' genomic sequence from genomic DNA are highlighted in grey. Sequence with coverage from both genomic DNA and cDNA amplification is in red text. Sequence with coverage from only cDNA amplification is in *italics*.

TCACAACCGACAAAGAATCCACGCCTATTCTCAAGTATATTTAACTATTAATAATAAGGC
TTGAATTGAATGCAGAGATTCATCATTCATTCATTTCCTTGTTACGGAGGAAGTAGGTAG
AGAAAAAAAAAGAAAAAACCAAACCTCACAAAAACGACGACAAAA**ATG**

GGTATGCGATTTTTTCCTTAGATTTTGTTAATTT
ATTATCAGTTATTCAAATTTGAGTTTCCATGATTAGGGTTTGATAAGATTTAATCATGTA
AAATTTTCATTTTTTTTTTGGGTTTGATAAGATTCAATACTGTAAAGTTTTCATTTTTTTG
TGAGTTTAGCAACTGTACCCAATTTTAAATCCTATTGATTATGAAACTATATGGTTTTAA
TCCTCATTAGAAACTCAAACATGTTCGATCTTATATTTTAATTAGATGTGACTTGGTTAT
AATTTTCTTTTTGGATTTAATATTTTTTTTTTCTGATTTAGATTTTGTCTCTTATGATTT
TTGTTTTAATGAAGTTACTTTCATTTGTTTGTAAATTATTGTTGGTGGTTTTGTTTTTGC
TAATTTGATCGATAAGTTTTTGTGTTTTGTGTTACTTTGCA

ATGGAAATCCAACCAGTACTGTTTTAGAGGAGAAATTAAGTGAACTGGAAGGGGCTGAGT
CAACTGTCTTATTGGCATCTGGAATGTGTGCTAGCATAGTTTTGCTCATGGCGCTGATTC
CAGCCGGTGGACATCTTGTGACTACAACTGATTGTTACAGGAAGACTAGGGTATTTATAG
AGACTGTGCTGCCAAAGATGGGGATCACGGCCTCTGTGATTGATCCTGCTGATGTTGGAG
CCTTGGAATCTGCTTTGGAGCAGAACAAGGTGTCTCTATTCTTCACCGAGTCTCCTACCA
ATCCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTTTGCCACAAGAATGGGG
CTTTGCTCTGCATTGATGGTACATTTGCGACGCCTTTGAACCAGAAGGCCCTTGCCCTTG
GTGCTGATTTGGTTATGCACTCTTGTACGAAGTACATTGGTGGACATCATGATGTCATTG
GTGGTTGCATAAGTGGCTCTGATAAGCTGATTTCTGAAATTCGCATTTTGCATCATATTT
TGGGCGGTGTTCTTAACCCGAATGCTGCATACCTATTCATCAGAGGCTTGAAAACCCTGC
ATCTGCGTGTACAGCAGCAGAATTCAACTGGAATGGGGATGGCCAAATTTTTAGAGGCAC
ATCCCAAGGTGAAGCGTGTCTACTATCCCGGCTTGCCAAGTCATCCTGAACATGAGCTTG
CCATGAGGCAGATGACTGGTTTTGGGGGTGTCGTCAGTTTTGAGATTGATGGAGATCTAG
AAACCACAATAAAATTTGTCGATTCACTGAAAATCCCATATATTGCTCCCTCTTTTGGTG
GCTGTGAGAGCATTGTGGACCAACCTGCCATTTTGTCTTACTGGGATCTTCCTGCATCTG
AAAGGGCTAAGTGGAAAATTTATGACAATTTGGTTCGCTTCAGTTTTGGAATTGAAGATT
*TTGAGGATTTGAAGGCTGATATCGTGCAGGCACTGGAAGTCATA**TAG**GCAGCAGAGTTCA*
CTATTCACCGCAAGCTTTTTTCCTTCTCTCGTTTATTTAATCGTGTTTGGTTCTGGTATA
TGCGGTGTGAAACTGCGGATATTTAATTTCTACTGAGATGTGGGTGTTGTTTGTTCTTGT
TTGTTTGGTGAACCTAGTGAATTATGACTGCCTAATGAAATCTGTTGGGTTTGTTTAATA
AAAAAAAAAAAAAAAAAAA

Nucleotide coding sequence, untranslated regions and exons removed. The translation initiation and stop codons are underlined and bold.

ATGGCCGTTTCAAGCCTCCACCGTGTCTTCACCTTCGAGTGCCGCTCCGATCCCGATTTC
GCCTCCCTCCCATCCACCGAAAACCTCCGCAGCACCCCTCGCCATTTCCCCGCCGCCGGT
CTCTCCGCCGCGTTCTACGGCGTCTCATCTCCAATCCTTCGTTTCCCCCCAAACTTTCAG
CGTCAGCTAAGCACTAAAGCTCGCCGTAACTGCAGCAACATCGGTGTCGCGCAAATCGTC
GCTGCTTCGTGGTCTAATGAAGGTGCCGGTAGCCCCGCCGCTGGGGCGAATCCGGTGCCG
GCCGTTGCATCAGCTGTTGATGCCGCGACTGCACCGCTCCCCGTAGACCTTGGCGCCGAT
GACGAGAGTGTTGTCGATGCTGGGAAAAATGGGGCTGTACAGACTAACCGTAGTTCTTAT
TCTTCGTTTTTGAAAACCGATGCAAGTAAAACGATTCATGCCGCTGAGAGACTTGGTAGG
GCTATAGTGACTGATGCAATTACCACTCCTGTGGTGAACACTTCTGCCTACTTTTTTAAG
AAGACTGCCGATCTTATTGATTATAAGGAGAAACGTCTGGCTAGTTATGAATATGGGCGC
TATGGAAATCCAACCAGTACTGTTTTAGAGGAGAAATTAAGTGAACTGGAAGGGGCTGAG
TCAACTGTCTTATTGGCATCTGGAATGTGTGCTAGCATAGTTTTGCTCATGGCGCTGATT
CCAGCCGGTGGACATCTTGTGACTACAACTGATTGTTACAGGAAGACTAGGGTATTTATA
GAGACTGTGCTGCCAAAGATGGGGATCACGGCCTCTGTGATTGATCCTGCTGATGTTGGA
GCCTTGGAATCTGCTTTGGAGCAGAACAAGGTGTCTCTATTCTTCACCGAGTCTCCTACC
AATCCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTTTGCCACAAGAATGGG
GCTTTGCTCTGCATTGATGGTACATTTGCGACGCCTTTGAACCAGAAGGCCCTTGCCCTT
GGTGCTGATTTGGTTATGCACTCTTGTACGAAGTACATTGGTGGACATCATGATGTCATT
GGTGGTTGCATAAGTGGCTCTGATAAGCTGATTTCTGAAATTCGCATTTTGCATCATATT
TTGGGCGGTGTTCTTAACCCGAATGCTGCATACCTATTCATCAGAGGCTTGAAAACCCTG
CATCTTCGTGTACAGCAGCAGAATTCAACTGGAATGGGGATGGCCAAATTTTTAGAGGCA
CATCCCAAGGTGAAGCGTGTCTACTATCCCGGCTTGCCAAGTCATCCTGAACATGAGCTT
GCCATGAGGCAGATGACTGGTTTTGGGGGTGTCGTCAGTTTTGAGATTGATGGAGATCTA
GAAACCACAATAAAATTTGTCGATTCACTGAAAATCCCATATATTGCTCCCTCTTTTGGT
GGCTGTGAGAGCATTGTGGACCAACCTGCCATTTTGTCTTACTGGGATCTTCCTGCATCT
GAAAGGGCTAAGTGGAAAATTTATGACAATTTGGTTCGCTTCAGTTTTGGAATTGAAGAT
TTTGAGGATTTGAAGGCTGATATCGTGCAGGCACTGGAAGTCATA**TAG**

Translated amino acid sequence. The truncation site for the removal of the putative plastid targeting sequence is underlined and bold. The putative MTO1 regulatory region is highlighted in grey. The exon 1/2 boundary is indicated by ^.

MAVSSLHRVFTFECRSDPDFASLPSTENLRSTPRHFPAAGLSAAFYGVSSPILRFPPN**FQ**
RQLSTK[redacted]VAASWSNEGAGSPAAGANPVPAVASAVDAATAPLPVDLGAD
DESVVDAGKNGAVQTNRSSYSSFLKTDASKTIHA^AERLGRAIVTDAITTPVVNTSAYFF
KKTADLIDYKEKRLASYEYGRYGNPTSTVLEEKLSELEGAESTVLLASGMCASIVLLMAL
IPAGGHLVTTTDCYRKTRVFIETVLPKMGITASVIDPADVGALESALEQNKVSLFFTESP
TNPFLRCVDIKLVSELCHKNGALLCIDGTFATPLNQKALALGADLVMHSCTKYIGGHHDV
IGGCISGSDKLISEIRILHHILGGVLNPNAAYLFIRGLKTLHLRVQQQNSTGMGMAKFLE
AHPKVKRVYYPGLPSHPEHELAMRQMTGFGGVVSFEIDGDLETTIKFVDSLKIPYIAPSF
GGCESIVDQPAILSYWDLPASERAKWKIYDNLVRFSFGIEDFEDLKADIVQALEVI

3.4 MULTIPLE SEQUENCE ALIGNMENT

Nucleotide multiple sequence alignment of CGS sequences from *A. thaliana* locus At3g01120 (At; GenBank accession no. U83500), tobacco (Nt; GenBank accession nos. AF097180 and AB035300), soybean (Gm; GenBank accession no. AF141602), chickpea (Ca), field pea (Ps), and lentil (Lc) generated by ClustalW (Larkin *et al*, 2007).

```
Ps  ATGGCCGTTTCA-------AG----------------CCTCCAC-CGCG--TCTTCA-CC 33
Lc  ATGGCCGTTTCA-------AG----------------CCTCCAC-CGTG--TCTTCA-CC 33
Ca  ATGGCCGTTTCA-------AG----------------TTTCCAC-CGCG--TTTTAA-CT 33
Gm  ATGGCCGTTTCG-------AG----------------CTCGCACATGCG--T-TTCA-CC 33
At  ATGGCCGTCTCATCATTCCAGTGCCCTACCATCTTCTCCTCCTCCTCAA--TCTCCG-GC 57
Nt  ATGGCGGTCTCA-------AG----------------CTGTGCTAGGGCTTTTCCATCC 36
    ***** ** **         **                      *       * *

Ps  TTCGAGTGCCGCTCCGATCCCGATTTC-------GCCTCCCTC-CCATC-CACCGAAAAC 84
Lc  TTCGAGTGCCGCTCCGATCCCGATTTC-------GCCTCCCTC-CCATC-CACCGAAAAC 84
Ca  TTTGAGTGTCGTTCAGATCCCGATTTC-------GCATCTCTT-CCTTC-CTTCGACAAC 84
Gm  TTTGAGTGCCGCTCCGATCCCGATTTCTC-----GCCCCCCCCGCCGTC-CTTCGACAAC 87
At  TTTCAATGCCGTTCTGATCCAGATCTC-------GTCGGTTCTCCCGTC-GGTGGATCAT 109
Nt  TTTGAGTGCCGTTCGGATGCCGAATTCTCTGGCGGTATTCCTCGCCACGACATTCCTAAC 96
    **  * ** ** ** *** * **  **       *         **            *

Ps  CTCCGTAGCAGCCGCCGTCATTTCC---CCGCCGCCGGTCTCTCCGCCGCGTTCTACGG- 140
Lc  CTCCGCAGCACCCCTCGCCATTTCC---CCGCCGCCGGTCTCTCCGCCGCGTTCTACGG- 140
Ca  CGTCGTA------CTCGTCATTTCC---CAACCGCTGCCGTTTCCGTCGCGTTTTACGC- 134
Gm  CTCCGC------CGCCGAAACTTCCGCTCCTCCGCAGATTCCGGCGCGGCGTTTCACGG- 140
At  CTC----------GCCGTCGTGTCCATGCCTCCGCCGGGATTTCTTCCTCATTCACCGGG 159
Nt  TCCGGCAAG----GCCAGCATTTTGA----GTCACGGAT-----CGTC-CGTTC-ACGG- 140
                  *      *         * * *           * **   **

Ps  --CATC------TCTTCTCCAATTCTTCGTTTCCCCCCAAACTTTCAGCGTCAGCTCAGT 192
Lc  --CGTC------TCATCTCCAATCCTTCGTTTCCCCCCAAACTTTCAGCGTCAGCTAAGC 192
Ca  --CGCT------TCATCTCCAATCCTTCGTTTTCCTCCCAACTTTCAGCGTCAACTCAGC 186
Gm  --CATC------TCCTCCCTCATCCTCCGCTTCCCTCCCAACTTCCAGCGCCAGCTAAGC 192
At  GACGCTGGATTATCCTCCAGGATCTTAAGATTTCCTCCTAATTTCGTCCGTCAGCTGAGC 219
Nt  --CCTG------TCTTCTCTTATTTATAGGTTTCCTCCGAACTTTGTTAGGCAGTTGAGC 192
      *          ** **    **     * ** ** ** ** **     * **  * **

Ps  ACTAAAGCTCGTCGTAACTGCAGCAACATCGGTGTCGCGCAAATCGTCGCTGCTTCGTGG 252
Lc  ACTAAAGCTCGCCGTAACTGCAGCAACATCGGTGTCGCGCAAATCGTCGCTGCTTCGTGG 252
Ca  ACCAAAGCTCGTCGTAACTGTAGCAACATCGGTGTCGCGCAAATCGTAGCTGCTTCGTGG 246
Gm  ACCAAGGCGCGCCGCAACTGCAGCAACATCGGCGTCGCGCAAATCGTCGCCGCTTCGTGG 252
At  ATTAAAGCCCGTAGAAACTGTAGCAACATCGGTGTTGCACAGATCGTGGCGGCTAAGTGG 279
Nt  ATTAAGGCTCGGAGGAACTGCAGCAACATCGGTGTGGCTCAAGTTGTGGCGGCTTCGTGG 252
    *  ** ** **  * ***** ********** ** ** **  * ** ** ***  ****

Ps  TCTAATGAA----------------------GGGATCGGCCACC----------CCGCC 279
Lc  TCTAATGAA----------------------GGTGCCGGTAGCC----------CCGCC 279
Ca  TCTAATGAA----------------------GGTTCCGGTAACC----------CCGCC 273
Gm  TCGAACAAC----------------------AGCGACAACTCTC----------CGGCC 279
At  TCCAACAACCCATCCTCCGCGTTACCTTCGGCGGCGGCGGCTGCTGCTACCTCGTCTGCA 339
Nt  TCTAACAAT--------------------------------AACT-----------CATC 269
    ** **  *
```

```
Ps  GCTGGGGTCCC-ACCGGTG-C-----CGGCCGTTGCATCAGCTGTTGATGCTGCGACGGC 332
Lc  GCTGGGGCGAA-TCCGGTG-C-----CGGCCGTTGCATCAGCTGTTGATGCCGCGACTGC 332
Ca  GCTGGGGCTCC-TCCGGTG-C-----CTGCCGTTGCATCAACTGTTGATGCGACTACAGC 326
Gm  GCCGGGGC----TCCGGCG-C-----CGCCCGCGGCCACCGCCACGGACGCCGCTACGGT 329
At  TCTGCGGTTTCTTCCGCCG-CATCTGCAGCCGCAGCCTCGTCCGCCGCCGCCGCCCCTGT 398
Nt  TCCTGATTTCACTCCGGTGGCTAAGGCCGTCGATGCCGCCGCCGCCGCCGCCGCCGCTAT 329
     *           ***  * *     *   **  **  *  *    *  **  *  *

Ps  GCCG------CTCCCCTTAGACCTCGA---CGCCGATG---ACGAGAACGTTGTCGATGC 380
Lc  ACCG------CTCCCCGTAGACCTTGG---CGCCGATG---ACGAGAGTGTTGTCGATGC 380
Ca  GCCT------CCCCCGTTAGACCTCGA---CGCCGTCGCCGACGAAAATGTTGTCGCCGT 377
Gm  GCCT------CTCCCCGTCGTCGTCGC---CGCCAACG---AGGACGTCGTTGTCTCCGC 377
At  GGCTGCCGCGCCTCCCGTCGTGCTGAAAAGCGTCGATG---AGGAGGTTGTGGTGGCCGA 455
Nt  TGCT---------CCTGTTGATACCACG---GTAGTTA---ACGAGGACGTGG-CGTTGG 373
      *          **  * *           *          * **    ** *     *

Ps  TG-----------GGGTAAATGGG---------------ACTGTACAGACTAACAGT--- 411
Lc  TG-----------GGAAAAATGGG---------------GCTGTACAGACTAACCGT--- 411
Ca  TG-----------ATGAAAATGGG---------------GCTGTACAGACTAACCGT--- 408
Gm  CGCGGCA-----GACGAGAACGGG---------------GCTGTACAGTTAAACAGT--- 414
At  GGAGGGGATCAGGGAGAAGATAGGT--------------AGTGTACAG-CTGACGG---- 496
Nt  TG-------------GAAAATGAGACTTGTAATGATCAAAATGTACAGTTTGACAGTTTG 420
     *                 *   *                  *******    ** *

Ps  ---AGGTCTTATT---CTTCGTTTTTGAAAACCGATGCAAGCAAAACGATTCATGCCGCT 465
Lc  ---AGTTCTTATT---CTTCGTTTTTGAAAACCGATGCAAGTAAAACGATTCATGCCGCT 465
Ca  ---AGTTCTTATT---CTAAGTTTTTGAAATCCGATGCAAGCAAAACGATTCATGCCGCT 462
Gm  ---AGTTCTTATT---CTTCATTTTTGAAATCCGATGCAAGCAAAACGATTCATGCCGCT 468
At  ----ATTCCAAAC---ATTCTTTCTTGAGCTCCGATGGGAGCCTCACTGTTCATGCCGGT 549
Nt  CCAAGTATGAAATACGCTTCGTTTTTGAATTCTGATGGGAGTGTCGCTATTCATGCCGGT 480
              *      *   ** ****   * ****  **     *  ********* *

Ps  GAAAGACTTGGTAGGGCTATTGTGACTGATGCAATTACCACTCCTGTGGTGAACACTTCT 525
Lc  GAGAGACTTGGTAGGGCTATAGTGACTGATGCAATTACCACTCCTGTGGTGAACACTTCT 525
Ca  GAAAGACTAGGTAGGGGTATCGTGACTGATGGAATTACCACTCCTGTGGTGAACACTTCT 522
Gm  GAAAGACTGGGTAGGGGTATTGAGACTGATGGAATTACCACCCCTGTGGTTAACACTTCT 528
At  GAAAGATTAGGCCGTGGTATAGTGACGGATGCTATCACTACTCCTGTAGTCAACACATCT 609
Nt  GAAAGATTGGGACGTGGCATCGTTACGGATGCAATTACTACCCCAGTAGTTAACACATCT 540
    ** *** * **  * *  ** *  ** ****  ** ** ** ** ** ** ***** ***

Ps  GCCTACTTTTTCAAGAATACTGCCGATCTTATTGATTATAAGGAGAAACGTCTGGCTAGT 585
Lc  GCCTACTTTTTTAAGAAGACTGCCGATCTTATTGATTATAAGGAGAAACGTCTGGCTAGT 585
Ca  GCCTACTTTTTTAAGAAAACCGCTGATCTTATTGATTTTAAGGAGAAACGTCAGACTAGT 582
Gm  GCCTACTTTTTTAAGAAAACCGCTGATCTCATTGATTTCAAGGAGAATCGTCAAGTGAGT 588
At  GCCTACTTCTTCAAGAAAACTGCTGAGCTTATTGACTTCAAGGAGAAAAGGAGTGTGAGT 669
Nt  GCTTATTTCTTCAACAAGACTTCTGAGCTCATTGATTTTAAGGAGAAAAGACGCGCAAGT 600
    ** ** ** ** ** ** **  * ** ** ***** *  ********  *       ***

Ps  TATGAATATGGGCGCTATGGGAATCCAACCAGTACTGTTTTAGAGGAGAAATTAAGTGAA 645
Lc  TATGAATATGGGCGCTATGGAAATCCAACCAGTACTGTTTTAGAGGAGAAATTAAGTGAA 645
Ca  TATGAATATGGGCGCTATGGCAATCCAACCAGTACTGTTTTAGAGGAGAAGATAAGTGAA 642
Gm  TATGAATACGGGCGCTATGGAAACCCAACGACGGTGGTTCTGGAGGAGAAGATAAGTGCA 648
At  TTCGAGTATGGTCGTTATGGTAACCCTACGACTGTGGTACTTGAAGATAAGATAAGTGCA 729
Nt  TTTGAATATGGGCGCTATGGGAACCCAACTACTGTTGTTTTAGAGGAGAAGATAAGCGCA 660
    *  ** ** ** ** ***** ** ** ** *     **  * ** ** **  **** * *
```

```
Ps  CTGGAAGGGGCTGAGTCAACTGTCTTATTGGCATCTGGAATGTGCGCTAGCATAGTTATG 705
Lc  CTGGAAGGGGCTGAGTCAACTGTCTTATTGGCATCTGGAATGTGTGCTAGCATAGTTTTG 705
Ca  CTCGAAGGAGCTGAATCAACTGTCTTAATGGCATCTGGAATGTGTGCTAGCATAGTCTTG 702
Gm  TTGGAGGGGGCCGAATCAACTGTGATAATGGCGTCTGGGATGTGTGCTAGCGTAGTCCTG 708
At  CTTGAAGGGGCTGAATCAACTTTGGTTATGGCATCTGGGATGTGTGCAAGCACTGTTATG 789
Nt  CTGGAGGGTGCTGAATCAACCTTGTTAATGGCATCTGGAATGTGCGCCAGTACTGTAATG 720
     * ** ** ** ** *****  *  *  **** ***** ***** ** **    **  **

Ps  CTCATGGCGCTGGTTCCAGCCGGTGGACATCTTGTGACTACAACTGATTGTTACAGGAAG 765
Lc  CTCATGGCGCTGATTCCAGCCGGTGGACATCTTGTGACTACAACTGATTGTTACAGGAAG 765
Ca  TTGATGGCGCTGGTTCCAGCTGGTGGACATCTTGTGACTACTACAGATTGTTACCGGAAG 762
Gm  TTTATGGCACTGGTTCCAGCTGGTGGACATCTTGTGACCACTACGGATTGTTATAGGAAG 768
At  CTTTTGGCATTGGTTCCTGCTGGTGGACACATTGTCACTACTACTGATTGCTACAGGAAG 849
Nt  CTGTTGGCTTTGGTTCCTGCTGGAGGCCATATTGTTACAACCACAGACTGCTATAGGAAG 780
     *  ****  ** **** ** ** ** **  **** ** ** ** ** ** **  *****

Ps  ACTAGGATATTTATAGAGACTGTGCTTCCGAAGATGGGGATCACGGCCTCTGTGATTGAT 825
Lc  ACTAGGGTATTTATAGAGACTGTGCTGCCAAAGATGGGGATCACGGCCTCTGTGATTGAT 825
Ca  ACTAGGATTTTTATAGAGACTGTGCTTCCAAAGATGGGGATCACGACCACCGTGGTTGAT 822
Gm  ACTAGAATATTCATTGAGACTTTTCTTCCTAAGATGGGGATCACGACCACTGTAATTGAT 828
At  ACTAGGATCTTCATGGAGAATTTTCTTCCCAAGTTGGGGATCACTGTCACTGTGATTGAT 909
Nt  ACCCGGATCTTTATTGAAACTATACTTCCTAAAATGGGGATTACGGCCACGGTCATTGAC 840
     **  *  * ** ** ** * * * ** ** **  ******* **   * * **  ****

Ps  CCTGCTGATGTTGGAGCCTTGGAATCTGCATTGGAGCAGAACAAGGTGTCTCTATTCTTC 885
Lc  CCTGCTGATGTTGGAGCCTTGGAATCTGCTTTGGAGCAGAACAAGGTGTCTCTATTCTTC 885
Ca  CCTGCTGATGTTGGAGCCTTGGAATCTGCATTGGAGCAGAACAATGTGTCTCTTTTCTTC 882
Gm  CCAGCAGATGTTGGAGCCTTGGAATCTGCATTGGAGCAGCACAATGTGTCTCTATTCTTC 888
At  CCTGCTGATATCGCAGGGCTTGAAGCTGCAGTGAATGAGTTCAAGGTATCTCTGTTCTTC 969
Nt  CCAGCCGATGTTGGAGCTCTGGAGTTGGCTCTAAATCAGAAGAAAGTGAATCTTTTCTTT 900
     ** ** *** * * **   * **    **  *  *  **   ** **   *** *****

Ps  ACCGAGTCTCCTACCAATCCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTT 945
Lc  ACCGAGTCTCCTACCAATCCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTT 945
Ca  ACTGAGTCTCCTACCAATCCATTCCTGAGATGTGTTGACATTAAGCTGGTTTCAGAGCTT 942
Gm  ACTGAGTCTCCTACCAATCCATTCCTGAGATGTGTTGATATTAAGCTGGTTTCAGAGCTT 948
At  ACTGAGTCCCCGACAAACCCATTCCTTAGATGTGTCGACATTGAGCTAGTTTCAAAAATA 1029
Nt  ACCGAGTCTCCAACAAATCCATTCCTCAGATGTGTAGACATTGAGCTGGTTTCAAAGCTT 960
     ** ***** ** ** ** ******** ******** ** *** **** ****** *  *

Ps  TGCCACAAGCATGGGGCTTTGCTCTGCATTGATGGTACATTTGCTACGCCTTTGAACCAG 1005
Lc  TGCCACAAGAATGGGGCTTTGCTCTGCATTGATGGTACATTTGCGACGCCTTTGAACCAG 1005
Ca  TGCCATAAGCATGGGGCTTTGCTCTGTATTGATGGTACATTTGCAACTCCTTTGAACCAG 1002
Gm  TGCCACAAGAAGGGGACTTTGCTCTGTATTGATGGTACATTTGCAACTCCATTGAACCAG 1008
At  TGCCACAAGAGGGGAACTCTGGTTTGCATTGATGGCACCTTTGCAACACCTCTGAATCAG 1089
Nt  TGTCATGAAAAAGGAGCCTTGGTTTGCATAGATGGGACTTTTGCAACTCCTCTTAATCAA 1020
     ** **  *    **  *  ** * ** ** ***** ** ***** ** **  * ** **

Ps  AAGGCCCTTGCCCTTGGTGCTGATTTGGTTATGCACTCTTGTACGAAGTACATTGGTGGA 1065
Lc  AAGGCCCTTGCCCTTGGTGCTGATTTGGTTATGCACTCTTGTACGAAGTACATTGGTGGA 1065
Ca  AAGGCCCTTGCCCTTGGTGCTGATTTGATTATGCACTCTTGTACAAAGTACATGGGTGGA 1062
Gm  AAGGCCCTTGCCCTTGGCGCTGATCTGATTCTGCACTCCTTAACAAAATACATGGGTGGA 1068
At  AAAGCCCTTGCTCTTGGTGCTGATCTTGTCGTGCACTCTGCTACAAAGTACATTGGAGGA 1149
Nt  AAGGCCCTTGCCCTTGGGGCTGACCTCGTTCTGCACTCTGCAACAAAATTCCTTGGTGGA 1080
     ** ******** ***** *****  *  *  *******    ** ** * * * ** ***
```

```
Ps  CATCATGATGTAATTGGTGGTTGCATAAGTGGCTCTGCTAAGCTGATTTCAGAAATTCGT 1125
Lc  CATCATGATGTCATTGGTGGTTGCATAAGTGGCTCTGATAAGCTGATTTCTGAAATTCGC 1125
Ca  CATCATGATGTCCTTGGTGGATGCATAAGTGGCTCTGATAAGTTGATTTCACAAATTCGA 1122
Gm  CATCATGATGTCCTTGGTGGTTGCATAAGTGGTTCAATTAAGGTGGTTTCGCAAATTCGG 1128
At  CACAATGATGTTCTTGCTGGATGCATCTGTGGTTCACTGAAGTTGGTTTCAGAAATTCGC 1209
Nt  CATAATGATGTTCTCGCAGGTTGCATTAGTGGTCCTTTAAAGTTAGTTTCAGAAATACGT 1140
    **  *******  * *  ** *****  ****  *    *** *  ****  **** **

Ps  ATTTTGCATCATATTTTGGGCGGAGTTCTTAACCCGAATGCTGCATACCTATTCATCAGA 1185
Lc  ATTTTGCATCATATTTTGGGCGGTGTTCTTAACCCGAATGCTGCATACCTATTCATCAGA 1185
Ca  ATTTTGCATCATATTTTGGGCGGTGCTCTTAACCCGAATGCTGCATACCTATTCATCAGA 1182
Gm  ACTTTGCACCATGTTTTGGGTGGTACACTTAACCCGAATGCTGCATACCTATTCATCAGA 1188
At  AATCTGCATCATGTGTTGGGAGGAACACTTAACCCAAACGCTGCGTACCTAATCATCCGA 1269
Nt  AACCTGCATCACATCCTGGGTGGTGCTCTCAATCCGAATGCCGCATATCTGATTATCCGA 1200
    *   **** **  *  **** **    ** ** ** ** ** ** ** **  * *** **

Ps  GGCATGAAAACACTGCATCTTCGTGTACAGCAGCAGAATTCAACTGGAATGGGGATGGCC 1245
Lc  GGCTTGAAAACCCTGCATCTTCGTGTACAGCAGCAGAATTCAACTGGAATGGGGATGGCC 1245
Ca  GGCATGAAAACATTGCATCTTCGTGTACAGCAGCAGAATTCAACCGGAATGAGGATGGCC 1242
Gm  GGCATGAAAACGCTGCATCTCCGTGTACAGCAGCAGAATTCAACAGGAATGAGGATGGCC 1248
At  GGCATGAAGACATTGCATCTTCGTGTACAGCAACAGAATTCGACCGCTTTTAGAATGGCC 1329
Nt  GGCATGAAGACGTTGCATCTTCGTGTACAGCAACAAAACTCAACTGCATTGAGGATGGCC 1260
    *** **** **  ******* ********** ** ** ** ** *   *  * ******

Ps  AAATTTTTAGAGGCACATCCCAAGGTGAAGCGTGTCTACTATCCAGGCTTGCCAAGTCAT 1305
Lc  AAATTTTTAGAGGCACATCCCAAGGTGAAGCGTGTCTACTATCCCGGCTTGCCAAGTCAT 1305
Ca  AAACTTTTAGAGGCACATCCCAAGGTGAAGCGTGTATACTATCCAGGCTTGCCTAGTCAT 1302
Gm  AAACTTTTAGAGGCACATCCCAAGGTGAAGCGGGTCTACTATCCAGGCTTGCCGAGTCAC 1308
At  GAAATTTTAGAGGCACATCCTAAGGTGAGTCATGTGTACTATCCAGGCCTTCCAAGTCAT 1389
Nt  GAGATTTTAGAGGCTCATCCCAAGGTGAGACATGTCTATTATCCAGGCCTGCAAAGTCAT 1320
     *  ********** ***** *******  *  ** ** ***** *** * *  *****

Ps  CCTGAACATGAGCTTGCCATGAGGCAGATGACTGGTTTTGGAGGTGTCGTCAGTTTTGAG 1365
Lc  CCTGAACATGAGCTTGCCATGAGGCAGATGACTGGTTTTGGGGGTGTCGTCAGTTTTGAG 1365
Ca  CCTGAACACGAGCTTGCCATGAGGCAAATGACTGGTTTTGGGGGCGTTGTCAGTTTTGAG 1362
Gm  CCTGAACATGAGCTTGCCAAGAGGCAGATGACTGGTTTCGGTGGTGTTGTCAGTTTTGAG 1368
At  CCCGAACATGAACTCGCCAAGCGACAAATGACTGGTTTTGGAGGTGTGGTCAGTTTCGAG 1449
Nt  CCAGAACATCACATTGCAAAGAAACAAATGACAGGTTTTGGTGGTGCGGTCAGTTTTGAG 1380
    ** *****  *  * ** * *   ** ***** ***** ** ** *  ******** ***

Ps  ATTGATGGAGATCTAGAAACCACAATAAAATTTATTGATTCACTGAAAATCCCATATATT 1425
Lc  ATTGATGGAGATCTAGAAACCACAATAAAATTTGTCGATTCACTGAAAATCCCATATATT 1425
Ca  ATAGATGGAGATCTAACAACCACAATAAAATTTGTGGATTCACTTAAAATCCCATATATT 1422
Gm  ATTGATGGAGATCTACATACCACAATAAAATTTATTGATTCATTGAAAATCCCATATATT 1428
At  ATTGATGGAGACATTGAAACGACAATCAAGTTTGTGGATTCTCTAAAGATTCCTTACATT 1509
Nt  GTTGACGGAGACCTGCTTACAACCGCAAAATTTGTGGATGCTCTGAAAATTCCGTATATT 1440
     * ** *****  *    ** **    ** *** * *** *  * ** ** ** ** ***

Ps  GCTCCCTCTTTTGGTGGCTGTGAGAGCATTGTGGACCAACCAGCCATTTTGTCTTACTGG 1485
Lc  GCTCCCTCTTTTGGTGGCTGTGAGAGCATTGTGGACCAACCTGCCATTTTGTCTTACTGG 1485
Ca  GCAGCCTCTTTTGGTGGCTGTGAGAGCATTGTGGATCAACCGGCCATTTTGTCTTACTGG 1482
Gm  GCGGCCTCGTTTGGTGGCTGTGAGAGCATTGTGGATCAACCTGCTATTTTGTCTTACTGG 1488
At  GCACCATCCTTCGGTGGCTGCGAAAGCATTGTTGACCAACCTGCTATCATGTCCTACTGG 1569
Nt  GCTCCATCATTTGGAGGTTGTGAGAGCATTGTGGACCAACCAGCAATAATGTCTTACTGG 1500
    **  * ** ** ** ** ** ** ******** ** ***** ** **  **** ******
```

```
Ps  GATCTTCCTGCATCTGAAAGGGCCAAGTGGAAAATTTATGACAATTTGGTTCGCTTCAGT 1545
Lc  GATCTTCCTGCATCTGAAAGGGCTAAGTGGAAAATTTATGACAATTTGGTTCGCTTCAGT 1545
Ca  GATCTTCCTGCATCAGAAAGGGCCAAATATAAGATTTATGACAATTTGGTTCGCTTCAGC 1542
Gm  GATCTTCCTCAGTCAGAAAGGGCCAAGTACAAGATTTATGACAACCTGGTTCGCTTCAGC 1548
At  GATCTGCCGCAAGAGGAGAGACTAAAGTATGGAATCAAAGATAACTTGGTTCGTTTCAGC 1629
Nt  GATCTTAGCCAGTCAGATAGAGCCAAGTATGGCATCATGGACAATTTGGTTAGGTTCAGC 1560
    *****          ** **    ** *     **    ** **  ***** * *****

Ps  TTTGGAATTGAAGATTTTGAGGATTTGAAGGCTGATATCGTGCAAGCCCTCGAGACCATC 1605
Lc  TTTGGAATTGAAGATTTTGAGGATTTGAAGGCTGATATCGTGCAGGCACTGGAAGTCATA 1605
Ca  TTCGGAGTTGAAGATTTTGAGGATTTGAAGGCTGATATCCTGCAAGCTCTCGAAGCCATA 1602
Gm  TTTGGAGTTGAAGATTTTGAGGATTTGAAGGCTGATGTCCTGCAAGCTCTGGAAGCTATA 1608
At  TTTGGAGTTGAAGACTTTGAAGATGTCAAAGCTGACATTCTTCAAGCTCTCGAAGCCATC 1689
Nt  TTTGGTGTGGAAGATTTTGATGACTTGAAAGCTGATATTCTTCAGGCTTTGGACTCAATA 1620
    ** **  * ***** ***** **  * ** *****  *  * ** **  * **    **

Ps  TAA 1608
Lc  TAG 1608
Ca  TAG 1605
Gm  TAG 1611
At  TGA 1692
Nt  TAG 1623
    *
```