

**Expression of the rice cystatin,
Oryzacystatin-I (OC-I) influences the plant
growth and development**

Sarah Owdah Alomrani

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
Centre for Plant Sciences

September 2020

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2020, The University of Leeds, Sarah Alomrani

Acknowledgment

Today, I have completed the writing of my thesis during the lockdown period which has been enforced due to the COVID-19 pandemic. Before I thank those who have support me through my journey, I would like to mention that, when I got divorced, around five years ago, I felt broken and I will always remember that my father told me: “You can start on your next chapter as an independent Saudi women”. I thought that the only thing that people will know about me was that I had been divorced, but now that is not the case. I have worked hard to obtain a scholarship to study in UK in my favourite subject, plant sciences, and the most important thing about me now is that I’m a strong women who no longer hopes for someone who would bring flowers for her, but I am able to plant my own flowers. My journey in the UK has been a very challenging experience, for which I am thankful. I am especially grateful to the Ministry of Education in Saudi Arabia for funding my PhD studentship.

It would have been impossible to complete the PhD process without the considerable help and support of a number of people. Firstly, I would like to thank my supervisor, Professor Christine Foyer, whose guidance and support have been invaluable to improving my skills and helping me to become a real scientist. Secondly, I would also like to thank my co-supervisor, Dr. Yoselin Benitez-Alfonso for all of her advice and help. Moreover, I am grateful to her for accepting me as a member of her team. Thirdly, I would also like to thank the members of the Foyer and Baker labs for enabling me to work in such a delightful and diverse setting.

It is my belief that true happiness is the result of having a loving and supportive network of family and friends, and I have been blessed with both. I am truly grateful to my wonderful parents who have been a source of constant support, guidance and inspiration. Without them, I would ever have been able to reach this stage.

In conclusion, I would like to add that I am immensely proud of this achievement. Despite the considerable difficulties I have faced, I fervently believe that I have made the right decision to study abroad, and I am eternally grateful to have had the chance to go through this difficult, yet rewarding, experience.

Abstract

Plants contain large numbers of proteases that fulfil a wide range of functions. Cysteine proteases and their endogenous inhibitors, phytocystatins, are involved in the control of protein turnover, but their precise functions remain poorly characterized. To study cysteine protease/phytocystatin functions in detail, the properties of *Arabidopsis*, wheat and soybean seeds expressing *Oryzacystatin I* (OC-I) were investigated. All of the transgenic seeds contained significantly more protein than the wild type, but germination was similar in all lines. The protein profiles of the seeds were broadly similar to the wild type, but the wheat and soy flour made from the transgenic seeds showed some differences in protein composition compared to that of the wild type. Moreover, a proteomic analysis of the transgenic wheat seeds revealed some differences in the accumulation of specific proteins, particularly storage proteins. The growth and development of three independent transgenic *Arabidopsis* lines that express the cystatin *Oryzacystatin I* (OC-I) in the cytosol (CYS lines) and three independent lines that express OC-I in the chloroplasts (PC lines) were characterised. The CYS and PC lines had a smaller rosette diameter with fewer leaves, and they accumulated less biomass than the wild type during vegetative growth. However, the transgenic lines had significantly more biomass than the wild type at the later (reproductive) stages of development. The CYS lines had less leaf chlorophyll and carotenoid pigments than the wild type, particularly at the later stages of development. In contrast, the PC rosettes accumulated more leaf pigments than the wild type at the later stages of leaf development. The CYS and PC rosettes flowered significantly later than the wild type. The abundance of leaf transcripts and photosynthetic proteins was changed in the CYS and PC lines. In addition, chloroplast-to-nucleus signalling, analysed by the changes in the levels of specific transcripts that encode photosynthetic proteins in the presence of chloroplast inhibitors, was extensively modified in the CYS and PC lines compared to the wild type. Moreover, photosynthetic carbon assimilation was less inhibited after exposure to high light stress in the CYS and PC lines than in the wild-type. Taken together, these data demonstrate that OC-I and its target cysteine proteases play important roles in the regulation of photosynthesis, as well as vegetative and reproductive development.

Table of contents

Acknowledgment	iii
Abstract	iv
Table of contents	v
List of Abbreviations	ix
List of Tables	x
List of Figures	xi
Chapter 1 . Introduction	1
1.1 Photosynthesis	1
1.2 The impact of stress conditions in plant growth and development	3
1.3 Chloroplast protein degradation	4
1.3.1 Rubisco degradation	5
1.4 Proteases and their Classification	8
1.4.1 Intracellular localization of proteases	10
1.4.2 Cysteine proteases (CPs)	11
1.4.3 The roles of CPs in plant growth and development.....	12
1.5 Plant protease inhibitors	15
1.5.1 Phytocystatins	16
1.6 Hypothesis and objectives.....	23
Chapter 2 . Materials and Methods	25
2.1 <i>Oryzacystatin-I</i> (OC-I) Constructs	25
2.1.1 Overexpression of OC-I in <i>Arabidopsis</i> lines.....	25
2.1.2 Overexpression of OC-I in wheat lines	28
2.1.3 Overexpression of OC-I in soybean lines.....	30
2.2 Seed sterilization	30
2.3 Growth conditions.....	30
2.3.1 The growth of transgenic <i>Arabidopsis</i> lines on agar plates	30
2.3.2 The growth of transgenic <i>Arabidopsis</i> lines on soil	31

2.3.3	The growth of transgenic wheat lines on soil	31
2.4	Phenotypic analysis.....	31
2.4.1	Germination efficiency and seedling survival	31
2.4.2	Root architecture	32
2.4.3	Shoot growth analysis	32
2.5	Seed yield.....	33
2.6	Selection on kanamycin plates	38
2.7	Leaf pigments	33
2.8	Total protein content extracted from Arabidopsis leaves.....	33
2.9	Total protein content extracted from seeds	34
2.10	Production of soy flour and isolation of soy flour protein isolates (SPI).....	35
2.11	SDS gel electrophoresis and protein staining.....	35
2.12	Western Blotting.....	36
2.13	High light treatments	38
2.14	The use of norflurazon and lincomycin to inhibit chloroplast processes	39
2.15	Photosynthetic carbon assimilation.....	39
2.16	Total protease activity and cysteine protease activity.....	40
2.17	Nucleic acid extraction	42
2.17.1	DNA extraction from transgenic Arabidopsis lines expressing OC-I.....	42
2.17.2	DNA extraction from wheat leaves and seeds.....	42
2.17.3	RNA extraction.....	44
2.17.4	Quantitative real-time reverse transcription polymerase chain reaction	46
2.17.5	Semi-quantitative reverse transcription-polymerase chain reaction	47
2.18	Determination of copy number in transformed wheat plants.....	47
2.19	PCR primers design.....	48
2.20	Label-free quantitative proteomics analysis.....	50
2.21	Bioinformatics analysis and construction of phylogenetic tree.....	51
2.22	Statistical analysis.....	51

Chapter 3 : The characterisation of transgenic <i>Arabidopsis</i> plants expressing the rice cystatin, <i>oryzacystatin-I</i> (OC-I), in the cytosol.....	52
3.1 Introduction.....	52
3.2 Results.....	54
3.2.1 Confirmation of OC-I expression in <i>Arabidopsis</i> plants.....	54
3.2.2 Seed germination.....	56
3.2.3 Root architecture of transgenic line seedlings.....	57
3.2.4 Rosette morphology.....	59
3.2.5 The role of OC-I in tolerance to high light stress.....	63
3.3 Discussion.....	70
Chapter 4 .The characterisation of transgenic <i>Arabidopsis</i> plants expressing the rice cystatin, <i>oryzacystatin-I</i> (OC-I) in the chloroplasts.....	76
4.1 Introduction.....	76
4.2 Results.....	79
4.2.1 Confirmation of OC-I expression in <i>Arabidopsis</i> plants.....	79
4.2.2 Germination efficiency and plant development after germination.....	81
4.2.3 Root architecture of transgenic line seedlings.....	82
4.2.4 Rosette morphology.....	84
4.2.5 A role for OC-I in tolerance to high light stress.....	88
4.3 Discussion.....	95
Chapter 5 . Selection of T4 generation transformed wheat plants over-expressing OC-I..	102
5.1 Introduction.....	102
5.2 Results.....	105
5.2.1 The overall procedure for generation selection.....	105
5.2.2 Selection of kanamycin-resistant T2 seeds from T1 seeds.....	108
5.2.3 Selection of T3 seeds and confirmation of the presence of the OC-I transgene in T2 plants.....	111
5.2.4 Confirmation of the presence of OC-I gene in T3 plants and selection of T4 generation seeds.....	122
5.3 Discussion.....	126

Chapter 6 : Properties of <i>Arabidopsis</i> , soybean and wheat seeds over-expressing <i>Oryzacystatin I</i> (OC-I)	128
6.1 Introduction	128
6.1.1 Dicot and monocot seed storage proteins	129
6.1.2 The degradation of storage proteins	130
6.2 Results.....	132
6.2.1 Transgenic <i>Arabidopsis</i> seeds overexpression OC-I	132
6.2.2 The effects of <i>Oryzacystatin I</i> (OC-I) expression in soybean seeds.....	135
6.2.3 Wheat lines overexpressing OC-I.....	139
6.2.4 Proteomic analysis of wheat seeds	143
6.3 Discussion	149
Chapter 7 . Genome-wide analysis of cysteine proteases in <i>Arabidopsis</i> and wheat.....	154
7.1 Introduction	154
7.2 Results.....	156
7.2.1 Phylogenetic analysis of cysteine protease in <i>Arabidopsis</i>	156
7.2.2 Phylogenetic analysis of well-characterized cysteine proteases	158
7.2.3 Phylogeny of selected representative cysteine proteases from different subfamilies of C1A encoded by <i>Arabidopsis</i>	160
7.2.4 Homology between protein sequences in <i>Arabidopsis</i> and the barley PAP14 cysteine protease	162
7.2.5 Phylogenetic analysis of cysteine protease in wheat	166
7.3 Discussion	168
Chapter 8 : General Discussion	171
8.1 Plant growth and development	172
8.2 Seed quality and properties.....	177
8.3 Identification of papain-like cysteine proteases (PLCPs)	179
8.4 Conclusion and perspective	181
References	182
Appendix	201

List of Abbreviations

°C	Degrees Celsius
µg	microgram
µL	microliter
BLAST	Basic local alignment search tool
bp	Base pairs
CaMV	Cauliflower mosaic virus
cm	Centimetre
CP	Cysteine protease
CPI	Cysteine protease inhibitor
C _T	Thermocycle number
CYS	<i>Arabidopsis thaliana</i> line expressing the OC-I protein in the cytosol
DNA	Deoxyribonucleic acid
E-64	Trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane
h	hours
HL	high light
kDA	Kilodalton
LHC	Light harvesting complex
LL	Low light
mM	milimolar
MS	Murashige and Skoog
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
OCE	OC-I expressing
OC-I	Oryzacystatin I
PAGE	PAGE Polyacrylamide gel electrophoresis
PC	<i>Arabidopsis thaliana</i> lines expressing the OC-I protein either in the chloroplast
PCD	programmed cell death
PCR	polymerase chain reaction
qRTPCR	quantitative real-time PCR
RNA	Ribonucleic acid
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SDS	SDS Sodium dodecyl sulfate
TAE	Tris-acetate EDTA
TBT	TBT Tris-buffered tween
v/v	Volume per volume
WT	Wild type

List of Tables

Table 1-1: Plant Protease Classification.	9
Table 2-1: Transgenic wheat lines (T1) were provided by the National Institute of Agricultural Botany (NIBA).....	29
Table 2-2: List of forward and reverse primers used to amplify OC-I plasmid in <i>Arabidopsis</i> and wheat.	49
Table 2-3: Lists of forward and reverse <i>Arabidopsis</i> primers used in qPCR and Semi-quantitative (RT-PCR).....	49
Table 2-4: Lists of forward and reverse wheat primers used in qPCR and copy number.....	49
Table 5-1: T-DNA copy number of transformed T1 plants of transgenic wheat expressing OC-I and WT..	110
Table 5-2: Overview of the segregation analysis of T2 seeds on kanamycin media and Chi-square test for the segregation of 3 Resistant: 1 Sensitive.....	112
Table 5-3: Overview of segregation analysis of T3 seeds on kanamycin media.	122
Table 6-1: The number of proteins identified in wheat lines... Error! Bookmark not defined.	
Table 6-2: The differentially expressed proteins in wheat lines relative to the WT..... Error! Bookmark not defined.	
Table 6-3: A comparison of differential abundance of storage proteins that can cause wheat allergy and coeliac disease between the WT and OCI-expressing lines. Error! Bookmark not defined.	
Table 6-4: A comparison of differential abundance of storage proteins that are important for wheat quality between the WT and WOC lines..... Error! Bookmark not defined.	
Table 7-1: Selected representative cysteine proteases from different subfamilies of C1A involved in different stages of growth in <i>Arabidopsis</i>	161
Table 7-2: Degree of similarity between selected representative cysteine proteases from different subfamilies of C1A in <i>Arabidopsis</i> and HvPAP14.....	163
Appendix I. T-DNA copy number of transformed T2 plants of transgenic wheat expressing OC-I and WT.....	201
Appendix II. T-DNA copy number of transformed T3 plants of transgenic wheat expressing OC-I and WT.....	202
Appendix III. Differential abundance of storage proteins expressed in the WT and OCI-expressing line (WOC).....	203
Appendix IV. Examples of the original gels used for Western-blot analysis of the Rubisco, D1 and phosphorylated D1 proteins in the leaves of the CYS and PC lines compared to WT <i>Arabidopsis</i> plants grown under moderate light (LL) and high light (HL) conditions. This	

analysis included a gel stained with Commassie Brilliant Blue and gel of the loading control to ensure appropriate quantitation204

List of Figures

Figure 1.1: The diagram illustrates the chloroplast electron transport chain pathway and the interaction between photosystem I and photosystem II which occurs in the thylakoid membranes.	2
Figure 1.2: A scheme showing the degradation pathways of chloroplast proteins during senescence.....	7
Figure 1.3: A scheme showing the localization of different protease types (cysteine, serine, metallo and aspartic proteases) in different subcellular compartments.	10
Figure 1.4: A scheme showing the mechanism of cysteine protease activity (adapted from van der Hoorn, 2008).	11
Figure 1.5: Scheme showing the mechanism of phytoecystatin action.....	18
Figure 1.6: Scheme showing the 3D structure of OC-I.	20
Figure 1.7: Amino acid sequence of <i>oryzacystatin-I</i> (OC-I) showing the location of the secondary structures (a-helix and b-sheets).....	20
Figure 1.8: OC-I expression of tobacco delays senescence and increases biomass production and seed yields.	22
Figure 2.1: Expression cassettes in destination vectors used to target OC-I to (A) the cytosol (pLBRCys-1) and to (B) chloroplast (pLBRPRKCys-1).....	27
Figure 2.2: Transient expression of yellow fluorescent protein in chloroplasts, demonstrating the ability of the PRK signal peptide to target transgenic proteins to plastids.	27
Figure 2.3: Representative image of transgenic <i>Arabidopsis</i> plants selected using BASTA leaf painting.	27
Figure 2.4: The destination vector, pSc4ActR1R2, used with the pENTR1A-OC-I to create the final binary expression vector pRMH052	29
Figure 2.5: A flowchart of the steps in the protein extraction method and western blot procedure used in this study.....	37
Figure 2.6: The FITC curve and equation used to calculate protease activity.	41
Figure 3.1: Agarose gel electrophoresis of DNA extracted from the leaves of wild type (WT) and 3 independent transgenic <i>A. thaliana</i> cell lines (CYS1, CYS3, and CYS4).....	54
Figure 3.2: Expression analysis of OC-I in four-week-old <i>Arabidopsis</i> plants (CYS lines and WT plants) compared with SAND as a housekeeping gene.....	55
Figure 3.3: Comparison of seed germination and seedling survival rate in transgenic lines overexpressing OC-I (CYS1, CYS3, and CYS4) and wild type (WT)	56
Figure 3.4: A comparison of root architecture in wild-type (WT) <i>Arabidopsis thaliana</i> and transgenic lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4) in 10-day-old seedlings:.....	57
Figure 3.5: A comparison of the representative phenotype of plants of the transgenic lines overexpressing OC-I in the cytosol (CYS1, CYS3 and CYS4) to that of WT.....	59
Figure 3.6: A comparison of shoot phenotype in the wild-type (WT) <i>Arabidopsis thaliana</i> and lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4) compared to the WT plants...61	61
Figure 3.7: A comparison of the amount of (A) chlorophyll, and (B) carotenoid pigments and (C) protein in the leaves of the wild-type (WT) <i>Arabidopsis thaliana</i> and lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4).	62

Figure 3.8: The effects of the expression of OC-I on photosynthetic CO ₂ assimilation rates in the leaves of CYS lines and WT <i>Arabidopsis</i> plants grown under low light (LL) and high light (HL) conditions.....	63
Figure 3.9: (A) Total protease activities and (B) cysteine protease activity in the leaves of CYS lines and WT <i>Arabidopsis</i> plants grown under low light (LL) and high light (HL) conditions...64	64
Figure 3.10: Western-blot analysis of Rubisco, D1 and D1 phosphorylated PSII proteins extracted from 5-week-old plants of the CYS lines and the WT grown under LL and then transferred to HL for 8 hours.....	65
Figure 3.11: The effect of OC-I expression on the transcripts level of nuclear-encoded chloroplast and chloroplast-encoded photosynthesis genes treated under (A) LL and (B) HL conditions in 5-week-old CYS and WT <i>Arabidopsis</i> plants.	67
Figure 3.12: <i>Arabidopsis thaliana</i> lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4) and the WT were treated with lincomycin (LINCO) or norflurazon (NF).....	69
Figure 4.1: Agarose gel electrophoresis of DNA extracted from the leaves of the (WT) and three independent transgenic <i>A. thaliana</i> lines (PC2, PC7, and PC9).....	79
Figure 4.2: Expression analysis of OC-I in four-week-old <i>Arabidopsis</i> plants (PC lines and WT plants) compared with SAND as a housekeeping gene.....	80
Figure 4.3: Comparison of seed germination and seedling survival rate in transgenic lines overexpressing OC-I (PC2, PC7 and PC9) and the wild type (WT).....	81
Figure 4.4: A comparison of root architecture in wild-type (WT) <i>Arabidopsis thaliana</i> and transgenic lines expressing OC-I in the chloroplast in 10-day-old seedlings.....	83
Figure 4.5: A comparison of the representative phenotype of plants of the transgenic lines overexpressing OC-I in the chloroplast (PC2, PC7 and PC9) to that of WT.....	85
Figure 4.6: A comparison of shoot phenotype in the wild-type (WT) <i>Arabidopsis thaliana</i> and lines expressing OC-I in the chloroplast at 4, 6, 8, 10 and 12 weeks after sowing.....	86
Figure 4.7: A comparison of the amount of (A) chlorophyll, (B) carotenoid pigments and (C) protein in the leaves of wild-type (WT) <i>Arabidopsis thaliana</i> and lines expressing OC-I in the chloroplast (PC2, PC7 and PC9).	87
Figure 4.8: The effects of the expression of OC-I on photosynthetic CO ₂ assimilation rates in the leaves of PC lines and WT <i>Arabidopsis</i> plants grown under low light (LL) and high light (HL) conditions.....	88
Figure 4.9: (A) Total protease activities and (B) cysteine protease activity in the leaves of the PC lines and WT <i>Arabidopsis</i> plants grown under low light and high light conditions.....	89
Figure 4.10: Western-blot analysis of Rubisco large subunit, PSII reaction centre D1 protein and the phosphorylated form of the D1 protein extracted from leaves of 5-week-old plants of the transgenic PC lines and the WT grown under LL and then transferred to HL for 8 hours..	90
Figure 4.11: The effect of OC-I expression on the transcript levels of nuclear-encoded chloroplast and chloroplast-encoded photosynthesis genes treated under (A) LL and (B) HL conditions in 5-week-old PC and WT <i>Arabidopsis</i> plants.	92
Figure 4.12: <i>Arabidopsis thaliana</i> lines expressing OC-I in the chloroplast (PC2, PC7 and PC9) and the WT were treated with lincomycin (LINCO), or norflurazon (NF).....	94
Figure 5.1: The top ten producers of wheat globally (2019.2020). Measurement is in metric tonnes.	102
Figure 5.2: A schematic diagram of the overall procedure for the selection of T4 generation, showing selected lines in each generation.	106
Figure 5.3: Overview of the production of transgenic wheat plants showing three main stages: Growing, flowering and harvesting.....	107

Figure 5.4: An example showing kanamycin selection of transformed wheat seeds.....	108
Figure 5.5: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3-week-old T1 wheat plants.....	109
Figure 5.6: PCR amplification of the OC-I transgene (600 bp) from 3-week-old T2 wheat plants.....	113
Figure 5.7: PCR amplification of the OC-I transgene (600 bp) from 3-week-old T2 wheat plants.....	114
Figure 5.8: PCR amplification of the OC-I transgene (600 bp) from 3-week-old T2 wheat plants.	115
Figure 5.9: PCR amplification of the OC-I transgene (600 bp) from 3-week-old T2 wheat plants.	116
Figure 5.10: PCR amplification of the OC-I transgene (600 bp) from 3-week-old T2 wheat plants.....	117
Figure 5.11: Analysis of relative expression level of the OC-I transgene in individual transformed plant from each transgenic wheat lines using specific primers for OC-I from 3-week-old T2 wheat plants..	119
Figure 5.12: Analysis of relative expression level of the OC-I transgene in individual transformed plant from each transgenic wheat lines using specific primers for OC-I from 3-week-old T2 wheat plants.	120
Figure 5.13: A comparison of the expression level of the OC-I transgene in transgenic wheat lines as pooled using specific primers for OC-I (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants.....	121
Figure 5.14: PCR amplification of the OC-I transgene (600 bp) from 3 old- week- T3 wheat plants.	123
Figure 5.15: Analysis of relative expression level of the OC-I transgene using specific primers for OC-I from 3-week-old T3 wheat plants..	124
Figure 5.16: Relative expression of the OC-I transgene in transgenic wheat seeds (WOC1, WOC2 and WOC3) and WT.....	125
Figure 5.17: A) A visual comparison of the size of the WOC1,WOC2 and WOC3 seeds compared to the WT. B) PCR amplification of the OC-I transgene from T4 wheat seeds..	125
Figure 6.1: A comparison of the seed weight, total protein content and the protein composition of the seeds of wild type (WT) <i>Arabidopsis</i> and OC-I expressing lines.....	134
Figure 6.2: A comparison of seed morphology of soybean seeds of the WT and the three independent transformed lines expressing OC-I (SOC-1, SOC-2 and SOC-3).	136
Figure 6.3: Protein content and composition of transgenic soybean seeds.....	138
Figure 6.4: A comparison of the seed properties of the WT and three independent transformed lines expressing OC-I: WOC-1, WOC-2 and WOC-3.....	140
Figure 6.5: Protein content and composition of transgenic wheat seeds.....	142
Figure 6.6: Cluster heat map representing the relative abundance of storage proteins in both the WT and WOC lines.	146
Figure 6.7: Cluster heat map of the abundance of the different storage proteins in the WT and WOC lines that can cause wheat allergy and coeliac disease.....	147
Figure 6.8: Cluster heat map of the relative abundance of various storage proteins in the WOC lines and the WT that are important for wheat quality.....	148

Figure 7.1: Phylogenetic tree providing an overview of the cysteine protease family of Arabidopsis.....	157
Figure 7.2: Phylogenetic overview of the well-characterized cysteine proteases in Arabidopsis.....	159
Figure 7.3: Phylogenetic tree for 13 cysteine proteases selected from different subfamilies of C1A in Arabidopsis	161
Figure 7.4: Phylogenetic tree showing the relationships between 13 Arabidopsis cysteine proteases and HvPAP14.....	163
Figure 7.5: A comparison of conserved protein domains in HvPAP14 (A) and CEP1 (B).	164
Figure 7.6: Prediction of 3D structure of (A) HvPAP14 and (B) CEP1 protein	165
Figure 7.7: Phylogenetic tree providing an overview of the cysteine protease family in wheat genome.	167
Figure 8.1: Diagram showing the degradation pathways of chloroplast proteins with OC-I located to cytosol and chloroplast.....	176

Chapter 1 . Introduction

Agriculture is predicted to face multiple challenges in the coming decades, not least because of the negative impacts of climate change, including an increase in the frequency of extreme weather events such as flooding and drought. Moreover, global agricultural productivity must increase considerably in the coming years to feed a growing global population (Mahajan and Tuteja, 2005). Studies to increase our current understanding of how plants tolerate environmental stresses are therefore important to identify new markers for the selection of improved crop varieties that can sustain yield even when they experience multiple stresses simultaneously (Tuteja, 2007). In recent years there has been considerable focus on improving the efficiency of photosynthesis as a mechanism to increase carbon gain and hence crop yields, for example by accelerating recovery from photoprotection (Kromdijk et al., 2016). A different mechanism that might enhance photosynthesis is to limit the proteolysis of photosynthetic proteins (Van der Hoorn, Renier AL, 2008).

1.1 Photosynthesis

Photosynthesis uses sunlight to drive electron transport with the subsequent assimilation of CO₂ into carbohydrates (Tanaka and Makino, 2009). Photosynthesis also plays a major role in cellular redox metabolism and signaling (Hisabori et al., 2007), not least by generating significant amounts of reactive oxygen species (ROS). Much of the nitrogen assimilated by plants is stored in the leaves as photosynthetic proteins during vegetative growth. This nitrogen is then released by regulated proteolysis during leaf senescence to support grain filling (Tanaka and Makino, 2009; Diaz-Mendoza et al., 2016). Photosynthesis is inhibited by environmental stresses, including drought, salinity, high light and high or low temperatures. The stress-induced

inhibition of photosynthesis causes a significant reduction in carbon gain and hence crop productivity (Chaves et al., 2009; Tuteja, 2007).

Photosynthesis occurs in the chloroplasts of higher plants. For simplicity, this process is often described in terms of light-dependent and light-independent reactions. The light-dependent reactions occur in the thylakoid membranes, whereas the light-independent reactions are localized in the stroma. The light-dependent reactions in the thylakoid membranes are driven by two photosystems (PS): PSII, which performs the light-dependent splitting of water and the transfer of electrons to plastoquinone, and PSI, which drives the light-dependent reduction of NADP to NADPH (Figure 1.1) The electron transport chain drives photophosphorylation to produce ATP. ATP and NADPH/H⁺ are then used in the assimilation of carbon dioxide in the Calvin cycle.

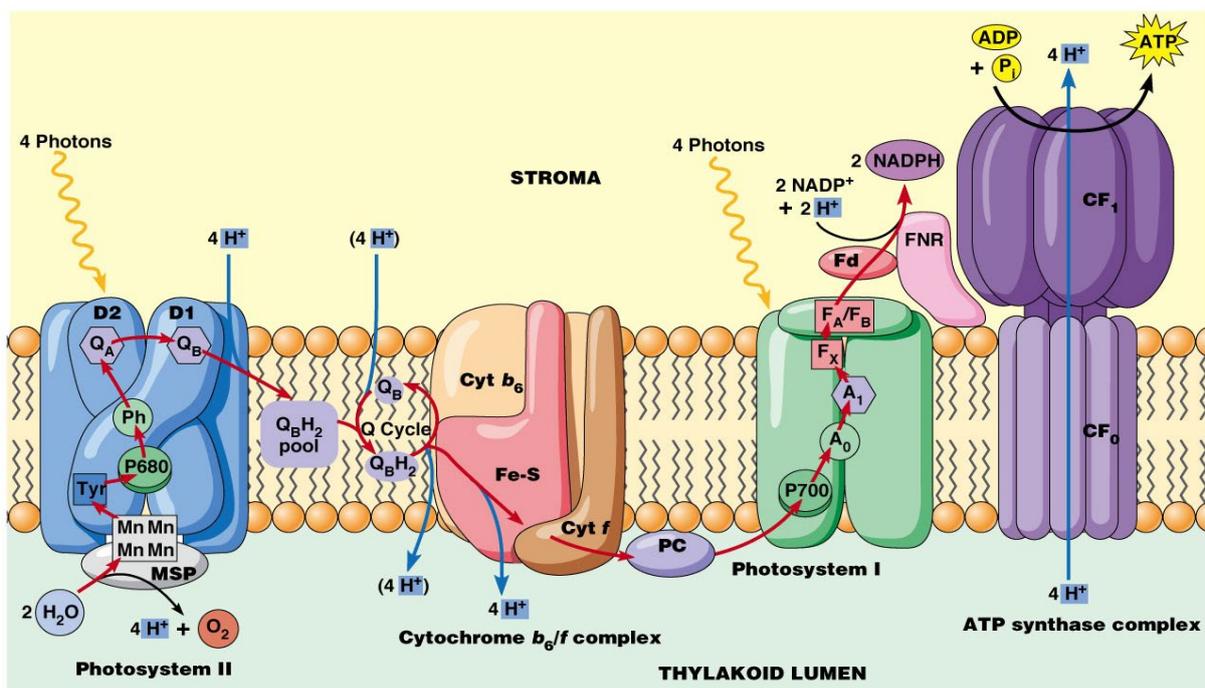


Figure 1.1: The diagram illustrates the chloroplast electron transport chain pathway and the interaction between photosystem I and photosystem II which occurs in the thylakoid membranes. The diagram is taken from http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-11/11_09.jpg

1.2 The impact of stress conditions in plant growth and development

The multifaceted challenges faced by global agriculture have been well documented in the literature, particularly in the context of the pressing need to expand food production capabilities in a sustainable manner (Fischer et al., 2014). Environmental stresses such as heat and drought limit crop productivity and threaten global food security (Lesk et al., 2016). Moreover, the predicted changes in weather patterns associated with climate change will only increase the frequency of extreme heat and drought events. These stresses decrease plant growth and induce changes in gene expression that facilitate survival during unfavourable environmental conditions (Duan et al., 2007). Developmental changes triggered by environmental stress include early leaf senescence, a process that requires extensive proteolysis, catalyzed by a range of proteases, including cysteine proteases (Carrión et al., 2013). However, increasing the breakdown and turnover of photosynthetic proteins will impair essential metabolic processes such as carbon assimilation and hence ultimately limit the productivity of the plants. High temperatures, for example, inactivate enzymes and increase protein degradation, as well as disrupting protein synthesis and membrane functions (Fahad et al., 2017).

The control of plant growth and development relies on an intricate balance between protein synthesis and protein degradation (Nelson and Millar, 2015). Regulated protein turnover is important at all developmental stages, including germination and senescence, as well as in plant defence responses such as programmed cell death (PCD) (Beers et al., 2000; Palma et al., 2002). Therefore, it is essential to obtain an in-depth understanding of cell maintenance processes, particularly elements such as protein dynamics, both to increase crop quality and to enhance the efficiency of crop production. Proteolysis is particularly important in plant responses to

environmental stress. A key challenge is therefore to develop a next generation of plants with modified protease activities that are better able to maintain yield over a wide range of environmental conditions.

1.3 Chloroplast protein degradation

Chloroplasts are complex organelles that fulfil crucial metabolic and signalling functions (Sadali et al., 2019). The photosynthetic proteins comprise 70% of the protein content of leaves. While a small number of proteins are encoded by the plastid genome and synthesised inside the chloroplasts, many photosynthetic proteins are encoded by nuclear genes and synthesized in the cytosol and then transported into the chloroplasts (Sakamoto et al., 2008). Chloroplast biogenesis is a complex process (Wang, F. et al., 2016). Regulated proteolysis is crucial to the maintenance of chloroplast quality control, and functions under both favourable and stress conditions (Wang, P. et al., 2018). Furthermore, the controlled degradation of chloroplasts is vital for the recycling of essential nutrients, such as the nitrogen stored in ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) (Wang, P. et al., 2018). During natural senescence, the number of chloroplasts is decreased by around 17% (Evans et al., 2010). Under stress conditions, about 28% of the chloroplast proteins is transported to the vacuole for degradation (Evans et al., 2010) through the process of autophagy (Otegui, 2018). Autophagy involves at least two specific pathways, which involve the production of Rubisco-containing bodies (RCB) and autophagy-dependent plastid bodies (PS). The degradation of chloroplast proteins can also occur by two other pathways that are independent of autophagy. In the first pathway, the process of chloroplast vesiculation (CV) liberates chloroplast proteins in vesicles that are targeted for degradation in the vacuole. In the second pathway, stromal proteins are degraded in senescence-associated vacuoles (SAVs) that contain cysteine proteases

(Schippers et al., 2015). However, the pathways of chloroplast stromal protein degradation remain poorly characterized, particularly during stress-induced senescence.

In contrast to the degradation of stromal proteins, the turnover of the PSII reaction center protein known as D1 is well characterized. D1 is highly susceptible to light-induced oxidative damage and has the highest rate of protein turnover of all the thylakoid proteins (Wang, F. et al., 2016). D1 protein degradation and repair are important in the prevention of the light-induced loss of photosynthetic functions (photoinhibition). Damaged D1 proteins must be removed from the membranes and replaced by a newly synthesized copy of D1, followed by re-assembly of the PSII complex. *In vitro* studies have suggested that various thylakoid proteases such as FtsH (a Zn²⁺ metallo-protease) and Deg2 (a serine protease) are involved in D1 protein degradation in higher plants and in cyanobacteria. In plants, these proteases are localized on the stromal side of the thylakoid membrane. Deg2 and other stromal peptidases are responsible for the cleavage of the large stromal DE-loop of the D1 protein, creating new termini that can be recognized by FtsH (Lindahl et al., 2000).

1.3.1 Rubisco degradation

Rubisco is perhaps the most abundant protein on earth (Raven, 2013). CO₂ is fixed by Rubisco catalysis, which uses ribulose-1,5-bisphosphate as the second substrate, to produce two molecules of 3-phosphoglycerate (3-PGA). However, Rubisco also catalyzes the oxygenation of ribulose-1,5-bisphosphate. When O₂ is fixed instead of CO₂, the reaction produces one molecule of 3-PGA and 2-phosphoglycolate (2PG). The production of 2PG marks the start of the pathway called photorespiration, in which 2PG is converted back to 3-PGA. This pathway involves multiple steps and releases previously fixed CO₂ and ammonia, and also uses ATP. The rate of the Rubisco

reaction is relatively slow, and so this enzyme is required in large amounts in order to drive the Calvin cycle. The large amounts of N stored in Rubisco are transferred to the seeds during leaf senescence. The expression of photosynthetic genes, including those that encode Rubisco and the light-harvesting chlorophyll *a/b*-binding protein (LHC), is decreased during leaf senescence and so they are classed as senescence down-regulated genes (Humbeck et al., 1996).

Rubisco is degraded both inside and outside of the chloroplasts (Irving and Robinson, 2006). However, key questions remain concerning the pathway of Rubisco degradation outside the chloroplasts and how it is controlled (Thoenen et al., 2007). The degradation of Rubisco involves the oxidation of specific cysteine residues, which alter the susceptibility of the enzyme to proteolysis by facilitating binding to the chloroplast envelope, marking it for degradation (Carrión et al., 2013). Rubisco is found outside the chloroplast and is localised in RCBs (also called Rubisco vesicular bodies (RVBs)) (Prins et al., 2008), as shown in Figure 1.2. These vacuole-like compartments are thought to contain the enzymes required for Rubisco degradation that are also found in vacuoles. This system functions alongside the 26S proteasome, which removes damaged and/or misfolded proteins, in controlled proteolysis (Vierstra, 2009). However, stress-induced oxidation can lead to carbonylation of the 26S proteasome, which favours the degradation pathway involving RCBs (Kurepa et al., 2009). The metabolite 2-carboxyarabinitol 1-phosphate (CA-1-P) is a competitive inhibitor of Rubisco, which protects the Rubisco protein from degradation inside the chloroplasts (Parry et al., 2008).

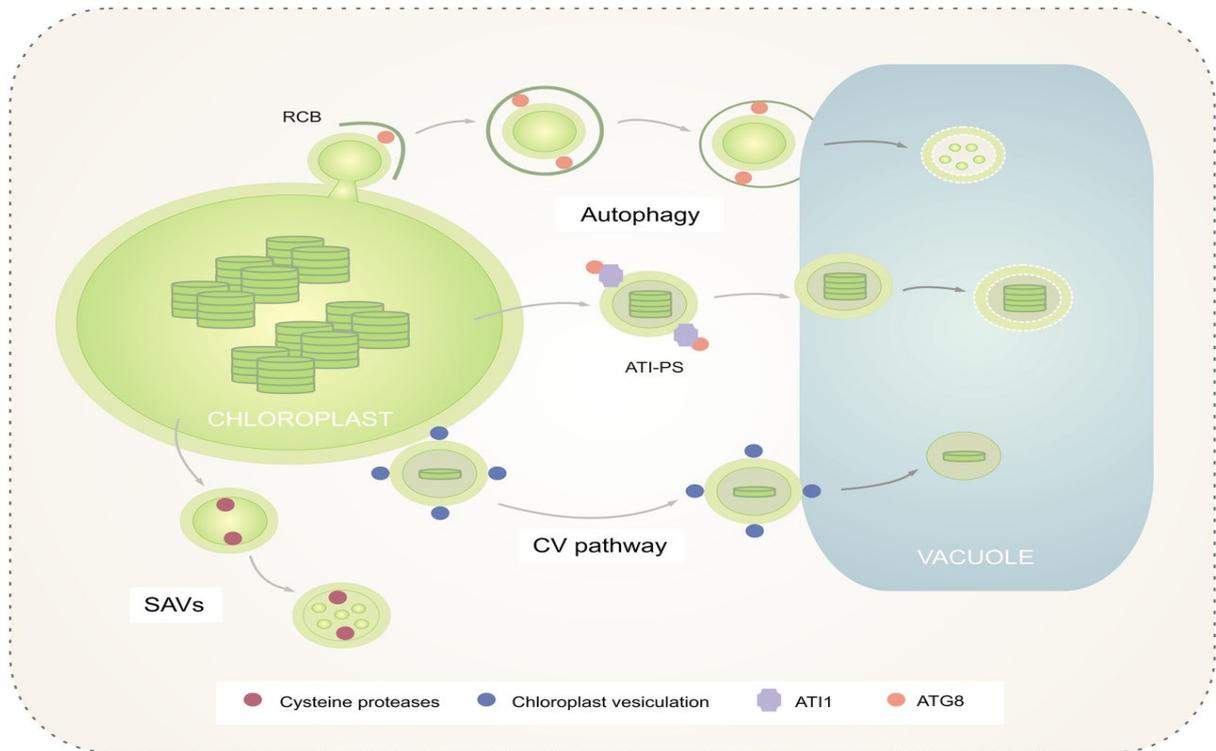


Figure 1.2: A scheme showing the degradation pathways of chloroplast proteins during senescence. Autophagy can occur through which involve the Rubisco-containing bodies (RCB) and autophagy-dependent plastid bodies (PS), respectively. Additionally, the degradation of chloroplast proteins is regulated by two other pathways independent of autophagy. In the first pathway, chloroplast vesiculation (CV) increases the production of vesicles that are liberated from the chloroplasts. These contain chloroplast proteins that are targeted for degradation in the vacuole. In the second pathway, stromal proteins are translocated into senescence-associated vacuoles (SAVs) containing cysteine proteases. Therefore the stromal proteins can be degraded inside the SAVs rather than being transported to the central vacuole for degradation (Schippers et al., 2015).

1.4 Proteases and their Classification

Proteolysis is essential for protein turnover, particularly during senescence (Carrión et al., 2013; Díaz-Mendoza et al., 2014). Proteases are widely distributed in nature. They are found in many different species, including mammals, plants, microorganisms and insects. These enzymes can be grouped according to the peptide bonds which are their substrates i.e. endopeptidases and exopeptidases. In exopeptidases, the bonds that are hydrolysed are located at the N- or C- terminals of the substrate proteins, while the endopeptidases target sites distant from the terminal sequences (Rawlings et al., 2011; Van der Hoorn, Renier AL, 2008). Proteases are divided into five families: Serine, Cysteine, Threonine, Aspartic and Metallo-proteases (Rustgi et al., 2018). Sequence and protein folding data have been used to further classify these proteases into subfamilies (Table 1.1) (Diaz-Mendoza et al., 2016). The catalytic residues of the active sites differ between these families. Serine proteases contain a catalytic triad of His-Asp-Ser, while threonine proteases have an active site comprising a threonine residue at the N-terminal (Rustgi et al., 2018). Aspartic proteases possess a dyad of two aspartates, and the metallo-proteases use a metal cation, usually Zn^{2+} in the active site (Rustgi et al., 2018). The peptidases from these groups are categorised into various families based on similarities in the sequences of amino acids. These families can be further subdivided into clans based on similarities in both the primary structure and the tertiary site, which indicate likely associations in evolutionary development (Kidrič et al., 2014).

Table 1-1: Plant Protease Classification.

Protease classification in plant	Identified proteases	References
Serine-proteinases	DegP protease (Family S1)	(Itzhaki et al., 1998)
	Subtilisins (Family S3)	(Figueiredo et al., 2014)
	Serine beta-lactamases (Family S12)	(Hall, B.G. and Barlow, 2004)
	Clp ATP-dependent proteases (Family S14)	(Yu and Houry, 2007)
	Lon protease (Family S16)	(Wickner and Maurizi, 1999)
Cysteine-proteinases	Caspase-like proteins	(Chichkova et al., 2004)
	Vacuolar-processing enzyme (VPE)	(Hara-Nishimura et al., 1998; Nakaune et al., 2005)
	Papain-like peptides	(Richau et al., 2012)
	Cathepsin-type proteases	(Gilroy et al., 2007)
	Asparaginyl endopeptidases	(Gruis, D. et al., 2004)
Aspartic-proteinases	Cardosins	(Ramalho-Santos et al., 1996)
Metallo-proteinases	Cathepsin D-like proteins	(Marttila et al., 1995)
	Metallo-proteinases FtSH	(Ito and Akiyama, 2005)
	Matrix-like enzymes	(Maidment et al., 1999)
Threonine proteases	The ubiquitin-proteasome 26S proteolytic system (family T1)	(Kurepa and Smalle, 2008)
	The ubiquitin-proteasome 20S	(Kurepa et al., 2009)

1.4.1 Intracellular localization of proteases

Plant proteases are localized in different cellular compartments, including the cytosol, nucleus, apoplast, vacuole, chloroplast, mitochondria and the Golgi apparatus (Figure 1.3). They are particularly abundant in seeds but they are also present in roots and leaves (Rogers et al., 1985), where they are regulated by proteinaceous inhibitors such as the Bowman-Birk inhibitors and phytocystatins, which are also found in all plant tissues (Rawlings et al., 2010a). Some proteases and protease inhibitors are specific to certain plant organs (Kidrič et al., 2014).

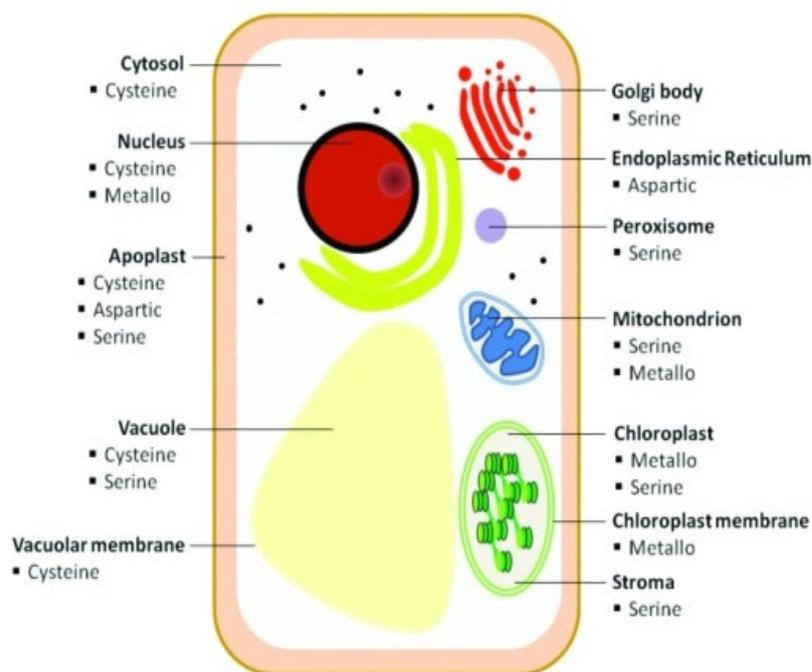


Figure 1.3: A scheme showing the localization of different protease types (cysteine, serine, metallo and aspartic proteases) in different subcellular compartments (Pillay, Priyen et al., 2014; Diaz-Mendoza et al., 2016).

1.4.2 Cysteine proteases (CPs)

Cysteine proteases (CPs) are considered to be the most plentiful type of proteases because they degrade and mobilize stored proteins (Martinez et al., 2009). These protein cleavage enzymes have a molecular mass of 21-30 and their activities are optimal in the pH range between 4.0-6.5 kDa (Grzonka et al., 2001). According to the MEROPS peptidase database, there are 140 cysteine proteases; these are classified into 15 families, which belong to five clans: cabain, caspase, clostripain, papain and streptococcal (Grudkowska and Zagdanska, 2004; Rawlings et al., 2011). CPs are responsible for between 30-90% of the maximal extractable proteolytic activity measured in plants that have been subjected to internal or external stimuli (Pernas et al., 2000; Wisniewski and Zagdanska, 2001; Sheokand et al., 2005). The localization of many CPs in the vacuoles is consistent with their acidic pH optima for activity (Callis, 1995). CPs have a cysteine residue which is activated by a histidine residue to begin a nucleophilic attack on the peptide bond of target proteins, as illustrated in Figure 1.4

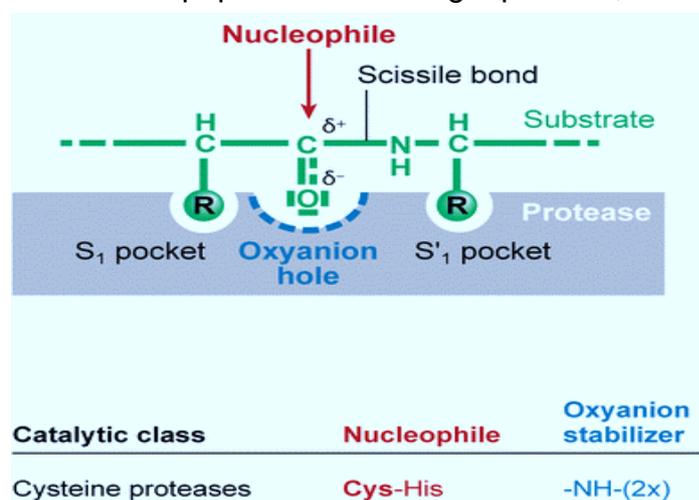


Figure 1.4: A scheme showing the mechanism of cysteine protease activity (adapted from van der Hoorn, 2008). The substrate (green) is held in place by its R groups in pockets of the CP (grey), with the substrate's carbonyl group in an oxyanion hole. A histidine residue accepts a proton from the cysteine residue, resulting in an anionic sulphur, and this attracts the carbon in the carbonyl group. A temporary dipole induced by the CP results in the carbon forming a bond with the sulphur, and the peptide bond is subsequently broken, with the donated proton binding with the nitrogen forming an amine group on one of the products. A water molecule is required to complete the carboxyl group on the other product (Van der Hoorn, Renier AL, 2008).

1.4.3 The roles of CPs in plant growth and development

CPs have been isolated and characterized from different organisms (Hall and Richards, 2013). Over 21 families of CPs have been described, with more than 50% of those found in viruses alone (Habib and Fazili, 2007). CPs hold a central position in the primary and secondary metabolism of plants. They are involved in the synthesis and metabolism of phytohormones such as ethylene (Romero et al., 2014). Ethylene influences seed germination and seedling growth, root hair development, fruit ripening and organ senescence (Bleecker and Kende, 2000; Mattoo and Handa, 2004); it is also involved in the plant's adaption to stress conditions (Ciardi et al., 2000). CPs are essential for the synthesis of S-adenosyl methionine (SAM) from methionine in the cytosol, which is then converted to 1-aminocyclopropane-1-carboxylic acid (ACC), the substrate for ethylene production (Bleecker and Kende, 2000; Iqbal et al., 2013).

CPs have been shown to have important roles at different stages of plant development, particularly in organ senescence and PCD (Solomon et al., 1999), as well as in protein storage and the regulation of protein and amino acid turnover (Grudkowska and Zagdanska, 2004; Grzonka et al., 2001). For example, vacuolar processing enzymes (VPE) are CPs that are found in a range of plant and animal organisms (Cai and Gallois, 2015; Sueldo and van der Hoorn, 2017). They were initially shown to function in the maturation of seed storage proteins plants (Hara-Nishimura et al., 1993; Hatsugai et al., 2004). VPEs activate protein precursors that operate in the vacuole (Hatsugai et al., 2006). The accumulation of two *Arabidopsis* CPs, RD21 and VPE γ , in ER bodies was shown to be involved in senescence-induced PCD (Rojo et al., 2003). While there is as yet no evidence to link autophagy to ER-body pathways, the CPs that are stored in ER-derived compartments of senescing tissues are able to reach the vacuole through the Golgi apparatus (Michaeli et al.,

2014). Although CP functions have been well characterized in model plant species such as *Arabidopsis*, there are relatively few studies in the literature on CP functions in cereals such as wheat.

CPs are also involved in the defence systems that protect plants against herbivores, insects (Konno et al., 2004) and fungal pathogens (Krüger et al., 2002), as well as abiotic stresses (Rabbani et al., 2003; Groten et al., 2006). For example, the levels of transcripts encoding the tobacco (*Nicotiana tabacum L.*) CP gene called NtCP2, a cysteine protease that is expressed in mature and senescing leaves, were increased when tobacco plants were exposed to drought (Beyene et al., 2006). Furthermore, vacuolar CP activity also plays a role in PCD (Beyene et al., 2006; Martínez, D.E. et al., 2007). Stress-induced protein degradation has a negative impact on plant growth, leading to reduced crop yields and quality (Fahad et al., 2017).

As early as the 1990s, it was recognised that CPs from differing families were expressed when plants were exposed to different stresses, including drought and low temperatures (Brzin and Kidrič, 1996; Ingram and Bartels, 1996). The expression of genes that encode for a variety of putative proteases was increased by drought in *Arabidopsis thaliana* and other species (Seki et al., 2002; Bartels and Sunkar, 2005). Moreover, the expression of numerous protease inhibitors in response to abiotic stresses have been reported. For example, the genes that encode a CP vacuolar processing enzyme were expressed in *A. thaliana* in response to heat shock (Li et al., 2012). Further analysis of how stress alters the activity and expression of CPs and their endogenous protease inhibitors is required to fully understand the stress-induced regulation of these proteins (Kidrič et al., 2014).

The activity of CPs is highly regulated in plant cells. In particular, the presence of tight binding protease inhibitors, known as called cystatins, is important in CP regulation. Cystatins, which block CP activity, have been identified and characterized in different plant species (Kunert et al., 2015). While transgenic plants expressing the rice cystatin OC-I show improved abiotic stress tolerance (Van der Vyver et al., 2003; Quain et al., 2014; Kunert et al., 2015), relatively little is known about the specificity of cystatins (Botha et al., 2017).

Three hundred and sixty-six cystatin-like sequences have been identified to date along with 951 C1 family sequences. These include the papain-like CPs in the *Viridiplantae* kingdom (<http://www.phytozome.net/>) (Kunert et al., 2015). The C1A family has been extensively studied in barley (*Hordeum vulgare* L.) because of the availability of genome information (Botha et al., 2017). The expression and function of individual C1 cysteine proteases has been investigated in barley (Velasco-Arroyo et al., 2016). Recently, the cysteine protease called HvPAP14 was shown to be localized in barley chloroplasts (Frank et al., 2019). This enzyme may be responsible for the partial degradation of Rubisco within the chloroplasts prior to autophagy (Xiong et al., 2007).

The CP family in wheat is much less well characterised than that in barley (Botha et al., 2017). The absence of extensive genome sequence information hinders the possibility of utilising technologies such as RNA-seq to characterise the expression of wheat CPs and cystatins. Extensive transcriptional changes have been reported in wheat leaves both in response to a range of environmental stresses and during the senescence process (Pearce et al., 2014). It would be useful to have better information on the wheat genes that encode specific proteases that are induced by environmental stresses and that are linked to leaf senescence (Botha et al., 2017).

1.5 Plant protease inhibitors

Protease inhibitors (PIs) have been mainly studied in three families of the plant kingdom, including Solanaceae, Gramineae, and Leguminosae (Pillay, P et al., 2012). PIs are small proteins which are largely located in storage tissues such as tubers and seeds, as well as in shoots and roots (Habib and Fazili, 2007; Diaz-Mendoza et al., 2016). PIs are found in all the plant organs, where they are located within the cell wall and apoplast as well as the internal cellular compartments. PIs are categorised based on their functions. They either block the protease active site (competitive) or they alter the tertiary structure (non-competitive). Both processes result in a loss of function. PIs are most commonly described in relation to their protease targets (Kidrič et al., 2014). They are often considered to be anti-metabolic proteins and they have proved to be an attractive target for the control of insect pests using transgenic plants because they interfere with insect digestion (Grzonka et al., 2001; Diaz-Mendoza et al., 2016). In addition, the induction of PIs in plants is a vital response to insect infestation or attack by pathogens (Grudkowska and Zagdanska, 2004). PIs also have important roles in plant responses to wounding, cold, drought and other abiotic stresses (Pernas et al., 2000).

There is a general consensus that CPs and PIs are important components of the network of regulated proteolysis pathways in plants, and that PIs prevent uncontrolled proteolysis (Kidrič et al., 2014). Environmental stresses elicit changes in the expression of genes that encode proteases and PIs, but relatively little mechanistic information is available about how they specifically regulate the protein content and composition of cells. Most data on proteolytic enzyme and PI functions is interpreted in terms of general physiological and developmental functions (Vierstra, 1996; Adam and Clarke, 2002; Palma et al., 2002; Schaller, 2004; Vierstra, 2009). For example, it

has been hypothesised that vacuolar proteases are specifically induced by developmental senescence and stress-induced senescence in a more general protein degradation pathway (Martínez, D.E. et al., 2007).

1.5.1 Phytocystatins

Cystatins or phytocystatins (PhyCys), which are proteins that inhibit plant cysteine proteases, are expressed in most if not all dicot and monocot cell types, where they have been detected in the vacuoles and in the cytoplasm, for example in potato (*Solanum tuberosum*) (Nissen et al., 2009) and tomato (*Solanum lycopersicum*) (Madureira et al., 2006). PhyCys contain a central Gln-Xaa-Val-Xaa-Gly motif in the protein sequence, where Xaa stands for any amino acid (Kunert et al., 2015). The C-terminal region contains either a Pro-Trp or Leu-Trp motif, whereas a conserved glycine residue is located in their N-terminal regions (Kunert et al., 2015). Phytocystatins and animal cystatins showed similarities in their conserved QxVxG motifs and in the mechanisms by which they inhibit the activity of their target enzymes, i.e. by interacting and blocking the active sites of their target proteins (Figure 1.5) (Christoff and Margis, 2014).

Several PhyCYSs have been characterized and shown to play important roles in processes such as in fruit development (Neuteboom et al., 2009), PCD (Belenghi et al., 2003) and seed germination (Hong et al., 2007; Hwang et al., 2010). PhyCys functions are most often described in the literature in terms of their functions as regulators of endogenous protein turnover and in the defence against herbivores and pathogens (Urwin et al., 2003; Outchkourov et al., 2004; Yang and Yeh, 2005; Christova et al., 2006; Martinez et al., 2009; Benchabane, M. et al., 2010). Their role as defence proteins is perhaps the best characterised. PhyCys were found to inhibit

insect and nematode digestive protease activities in *in vitro* experiments that employ artificial diets and in bioassays on plants that had been transformed to express different PhyCys genes (Atkinson et al., 2004; Carrillo et al., 2011). Furthermore, these inhibitors also show antipathogenic and antimite activities (Gutierrez-Campos et al., 1999; Martinez, M et al., 2003; Carrillo et al., 2011). However, little is known about the specific targets for PhyCys or the types of proteins that subsequently provide protection from degradation (Martínez, M. et al., 2012). PhyCYSs are also important in plant defences against abiotic stress responses but in most cases their precise functions remain poorly characterised (Gaddour et al., 2001; Zhang, X.M. et al., 2009; Zhang, X. et al., 2008; Diop et al., 2004).

A range of transgenic plants that constitutively express cystatins have been used to study their roles in abiotic stress responses (Van der Vyver et al., 2003). The assumption is that tolerance to stresses is achieved because PhyCys block or partially prevent stress-induced protein turnover (Van der Vyver et al., 2003; Zhang, X. et al., 2008). Hence, although such 'pleiotropic' effects are only usually thought to be unintentional metabolic interference, they can be useful in crop improvement (Benchabane, M. et al., 2010). The ectopic expression of protease inhibitors in leaves had negligible effects on plant growth and development (Benchabane, Meriem et al., 2008). Moreover, recombinant protease inhibitors that operate against specific endogenous proteases could be used to modulate *in situ* proteolytic activities (Faye et al., 2005).

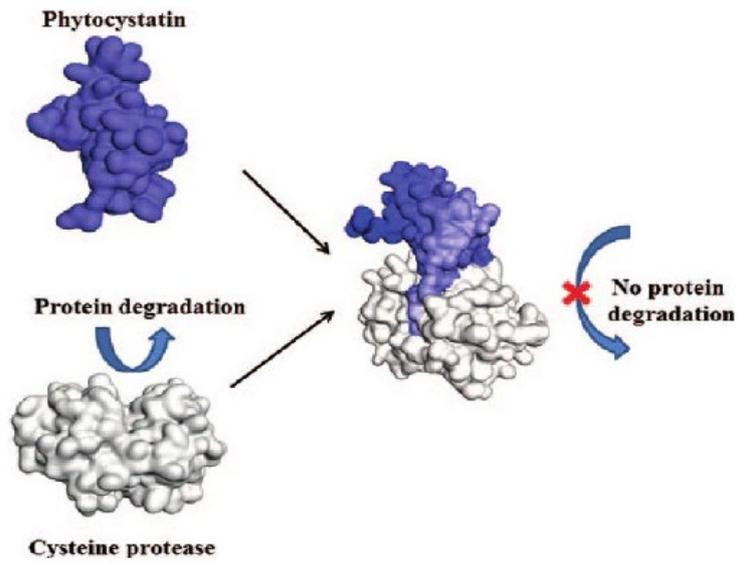


Figure 1.5: Scheme showing the mechanism of phytocystatin action. A cysteine protease-cystatin complex is formed in which the target protein is inhibited, preventing protein degradation (Kunert et al., 2015).

1.1.1.1 Rice cystatin, Oryzacystatin, OC-I

Oryzacystatin I (OC-I) was first identified in rice seeds (*Oryza sativa L. japonica*) (Abe, Keiko et al., 1987). This small protein of 120 amino acids has a molecular mass of 15,355 Daltons (Dou et al., 2011). The gene encoding OC-I is located on rice chromosome 1 (Dou et al., 2011). The three-dimensional structure contains one α -helix and a five-stranded β -sheet motif, but like other family-1 cystatins (stefins), OC-I lacks disulfide bonds (Figure 1.6). OC-I is a competitive inhibitor of papain-like CPs, as shown in Figure 1.7 (Benchabane, M. et al., 2010). Transgenic plants expressing OC-I were first produced to enhance resistance to weevils (*Sitophilus oryzae L.*) and the red flour beetle (*Tribolium castaneum H.*) (Lawrence and Koundal, 2002). The activities of papain-type enzymes such as oryzains α and β were inhibited by OC-I (Quain et al., 2015). In addition, OC-I could be useful in controlling nematodes and other pests in plants (Urwin et al., 2001).

Studies on OC-I expression in transgenic plants showed that abiotic stress tolerance and biotic stress tolerance were increased by the presence of the transgene (Van der Vyver et al., 2003; Prins et al., 2008). When transgenic OC-I-expressing tobacco plants were transiently transformed by an *E. coli* gene encoding glutathione reductase (GR), the resultant glutathione reductase activity was higher than when the wild type transformed to express GR, suggesting that OC-I can inhibit the activities of endogenous stress-induced cysteine proteases that target GR (Pillay, P et al., 2012). The expression of OC-I in tobacco altered shoot growth and development (Figure 1.8). Moreover, the transgenic tobacco plants expressing OC-I showed delayed leaf senescence, together with greater accumulation of leaf protein, delayed Rubisco degradation, and increased shoot biomass production and seed yields (Van der Vyver et al., 2003; Prins et al., 2008). Expression of a GFP recombinant OC-I protein in the transgenic tobacco plants showed that the protein was present in the leaf cytoplasm, chloroplasts and vacuoles (Prins et al., 2008). Transgenic soybean plants expressing OC-I had a greater number of root nodules, although the nodules were slightly smaller than those on wild type controls (Kunert et al., 2015; Quain et al., 2015). Moreover, *A. thaliana* plants expressing OC-I showed similar results to those obtained with tobacco expressing OC-I (Van der Vyver et al., 2003). The transgenic lines had a slower rate of growth at the early stages of development but a greater leaf area at flowering than the wild type (Quain et al., 2014). Cysteine proteases have not as yet been identified in *A. thaliana* or tobacco chloroplasts. A key question therefore concerns how OC-I expression can alter the abundance of chloroplast proteins and enhance the activity of Rubisco during age-related or stress-induced senescence (Prins et al., 2008). Transgenic *Arabidopsis* lines have been produced that express OC-I either in the

cytosol or the chloroplasts (Quain et al., 2014; Quain et al., 2015) to address the question of whether OC-I specifically targets chloroplast proteins.

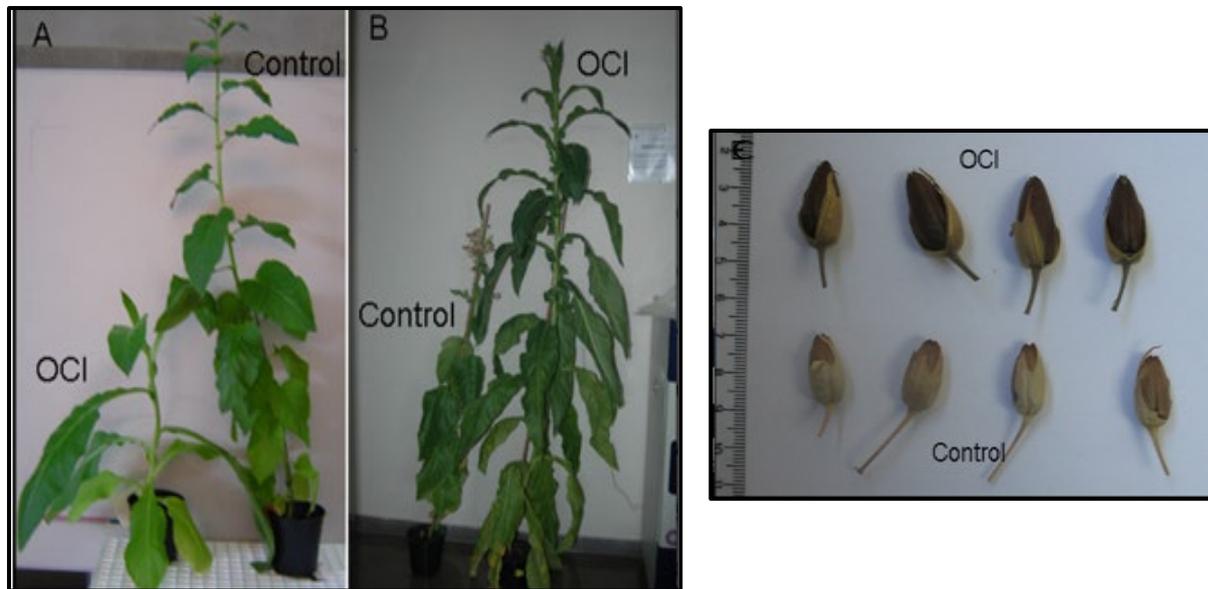


Figure 1.8: OC-I expression of tobacco delays senescence and increases biomass production and seed yields. Phenotypes of wild-type control and OC-I-expressing tobacco plants at 4 (A) and 8 (B) weeks, together with examples of seed pods (C) from control and OC-I-expressing tobacco plants, illustrating the larger pods in the latter (Van der Vyver et al., 2003).

1.6 Hypothesis and objectives

Previous studies have shown that OC-I expression resulted in the accumulation of chloroplast proteins, particularly in senescent leaves, and that photosynthesis in the transgenic plants was better protected against the inhibition caused by exposure to cold and drought stress (Quain et al., 2014; Quain et al., 2015). The hypothesis on which these studies are based is that CPs with relevant chloroplast functions are important in the control of plant growth and abiotic stress tolerance. It is proposed that developmentally regulated and stress-induced proteolysis can be limited by constitutive expression of OC-I. OC-I has previously been expressed in a range of plant species, including tobacco, soya and *A. thaliana*. In these studies, OC-I was expressed without a targeting sequence, and hence the product of the transgene was presumed to be located in the cytosol. In an earlier PhD study in the Foyer lab, transgenic *A. thaliana* plants were produced in which OC-I was specifically targeted to the chloroplasts (Quain et al., 2014). However, the mechanism of action of OC-I in *Arabidopsis* chloroplasts is unknown because CPs have, as yet, been identified only in barley chloroplasts.

The current research was undertaken to study the effects of OC-I expression in the chloroplasts or the cytosol on the growth and shoot phenotypes of *Arabidopsis* plants to determine whether OC-I influences processes within the chloroplasts as well as the cytosol. Recently, a cysteine protease inhibitor was described in barley chloroplasts (Frank et al., 2019). A key question concerns how OC-I expression alters the abundance of chloroplast proteins and enhances the Rubisco activity during age-related or stress-induced senescence.

The rice OC-I protein has been identified as having a role in both defence against pathogens and abiotic stress. This study will mainly focus on the effects of OC-I on

plant growth and development under optimal and stress conditions such as high light.

1.6.1 Research hypotheses:

The hypothesis of this thesis is that the expression of OC-I target to chloroplast as well as cytosol in *Arabidopsis* significantly influences plant growth and development under optimal and high light conditions (Chapter 3 and 4). Moreover, the expression of OC-I affects seed size and seed protein contents in all three of the studied plant species (*Arabidopsis*, wheat and soybean; Chapter 6). Hence, this study will help future researchers to identify any chloroplast proteins that are inhibited by OC-I and to characterize the possible roles of these proteins in important chloroplast functions such as protein turnover and Rubisco degradation.

The specific objectives of the study were:

1. To characterise the phenotypes of transgenic *Arabidopsis* plants expressing the rice cystatin, oryzacystatin-I (OC-I), in the cytosol and chloroplasts to that of the wild type (Chapter 3 and 4).
2. To study the effects of OC-I expression on gene expression and the abundance of the Rubisco and the photosystem II reaction centre D1 proteins in *A. thaliana* under high light stress (Chapter 3 and 4).
3. To determine the effects of OC-I expression on retrograde signalling pathways that regulate photosynthetic gene expression in *A. thaliana* (Chapter 3 and 4).
4. To select T2, T3 and T4 transgenic wheat lines overexpressing OC-I (Chapter 5).
5. To investigate the influence of OC-I on the size and properties of *Arabidopsis*, soybean and wheat seeds (Chapter 6).
6. To identify all papain-like cysteine proteases in both *Arabidopsis* and wheat, using public databases (Chapter 7).

Chapter 2 . Materials and Methods

2.1 *Oryzacystatin-I* (OC-I) Constructs

Various vectors that overexpress the rice (*Oryza sativa*) cysteine protease inhibitor I *oryzacystatin-I* (OC-I) were constructed in the Christine Foyer Lab, Faculty of Biological Sciences, University of Leeds for expression in different plant species (Quain et al., 2014) .

2.1.1 Overexpression of OC-I in *Arabidopsis* lines

Transgenic *Arabidopsis* lines expressing OC-I either without a targeting sequence (for expression in the cytosol) or with a chloroplast targeting sequence (for expression in the stroma) were prepared. The constructs designated as pLBRCys-1 and pLBRPRKCys-1, were prepared by Dr Eugene Makgopa, a former PhD student in the Christine Foyer group, Faculty of Biological Sciences, University of Leeds (Quain et al., 2014). Both constructs contained the gene encoding OC-I cystatin under the control of a double CaMV promoter, and a CaMV terminator sequence (Figure 7A and B). Additionally, the plasmids contained a gene (Bar) that confers resistance to the herbicide BASTA, also under the control of a double CaMV promoter, and a gene (aaDa) encoding spectinomycin resistance to bacterial selection (Quain et al., 2014). The pLBRCys-1 is responsible for the expression of OC-I to the cytosol (Figure 2.1A). The pLBRPRKCys-1 contains OC-I and the sequence encoding the signal peptide of phosphoribulokinase (PRK; Figure 2.1B) leading to targeting of the protein in the chloroplast stroma (Figure 2.2; Jonak, AIT Austrian Institute of Technology, unpublished data).

The transgenic plants were prepared by the floral dip method from the *Agrobacterium tumefaciens* strain GV3101 (Clough, S.J. and Bent, 1998). The *A. thaliana* ecotype Columbia (Col-0) was used as the wild-type (WT) in both cases (Quain et al., 2014).

Transgenic plants were selected using the leaf painting method (Paz et al., 2006). In brief, the glufosinate (an ammonium salt of phosphinothricin, the main active ingredient of the herbicide BASTA) was applied on the leaves. It may also be possible that high concentrations of glufosinate resulted in a high mortality rate in the transgenic plants. Therefore, PCR and qRT-PCR were also used to select successful transformed plants (Figure 2.3; (Quain et al., 2014)). The T4 generation of transgenic lines was analysed in the present studies. Lines that express OC-I in the cytosol are hereafter designated as CYS lines, while those that express OC-I in the chloroplasts are designated as PC lines.

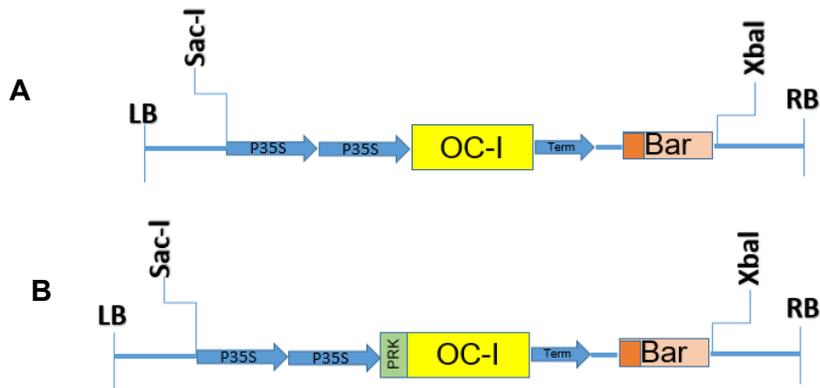


Figure 2.1: Expression cassettes in destination vectors used to target OC-I to **(A)** the cytosol (pLBRCys-1) and to **(B)** chloroplast (pLBRPRKCys-1).

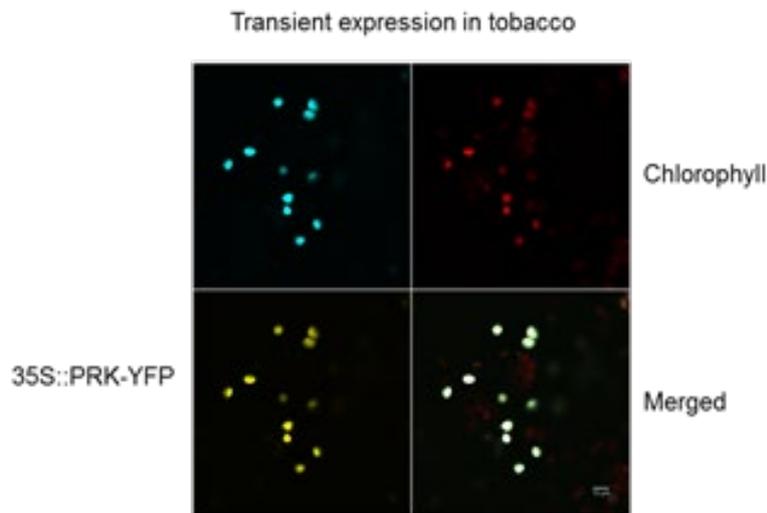


Figure 2.2: Transient expression of yellow fluorescent protein in chloroplasts, demonstrating the ability of the PRK signal peptide to target transgenic proteins to plastids.



Figure 2.3: Representative image of transgenic *Arabidopsis* plants selected using BASTA leaf painting. Resistant plants were observed to be green (1), partial resistance in plants were slightly yellowed with some partially resistant green patches (2 and 4) and susceptible plants were completely yellow (3).

2.1.2 Overexpression of OC-I in wheat lines

The OC-I construct used for expression in wheat plants was prepared by Gloria Comadira, a former PhD student in the Christine Foyer group, assisted by Dr Barry Causier, a member of the School of Biology, Leeds University (Comadira, 2015). The Gateway cloning system (Invitrogen) was used to produce the OC-I expression vector, starting with the removal of the OC-I gene (derived from cDNA) from a pLBR19-OC-I construct. In accordance with the Gateway technique, the OC-I sequence was cloned into a pENTR1A vector containing a kanamycin resistance gene (Comadira, 2015).

The pENTR1A-OC-I vector was sent to the National Institute of Agricultural Botany (NIAB, Park Farm, 1 Villa Rd, Impington, Histon, Cambridge, UK), where the transgenic wheat lines were generated. The intermediate cassette containing the OC-I coding sequence was recombined into binary vector pSc4Act-R1R2 to create pRMH052 using a Gateway LR Clonase II kit (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Following sequence verification, this plasmid was electro-transformed into the *A. tumefaciens* strain LBA4404 pSB1 (Hellens et al., 2000; Komari et al., 1996).

Hexaploid spring wheat cv. Fielder plants (USDA ARS) grown in controlled environment chambers at 20°C day/15°C night with a 16h-day photoperiod were used for the transformation. Transformation of immature embryos isolated for 14-20 days post-anthesis was carried out by co-cultivation with *Agrobacterium* containing pRMH052, the rice actin promoter was used in the final construct (Figure 2.4), for two days in the dark (Ishida et al., 2015). Subsequent removal of the embryonic axis and tissue culture was performed as previously described (Risacher et al., 2009). Thirty-seven regenerated wheat plants were confirmed as transformed by PCR amplification of the OsCrystatin transgene and the T-DNA copy number was determined by qPCR

assay (Milner et al., 2018). The T1 generation of transgenic wheat was produced and analysed by PCR, which was used to identify the number of independent insertions in the genome of 37 independently transformed lines, alongside 4 empty vector controls. Plants that contained a single insertion were selected for further analysis (Table 2.1).

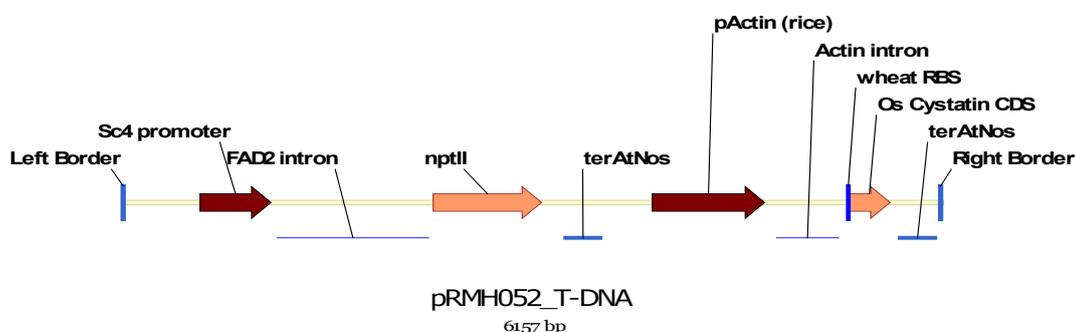


Figure 2.4: The destination vector, pSc4ActR1R2, used with the pENTR1A-OC-I to create the final binary expression vector pRMH052

Table 2-1: Transgenic wheat lines (T1) were provided by the National Institute of Agricultural Botany (NIBA), including 37 independent transformed lines together with 4 empty vector controls (E.con). Copy number estimation by qPCR and confirmation by PCR of the transgene are shown and gene of interest (GOI) PCR data confirms that the T-DNA has been completely inserted into lines. Blue labels indicate single insertion lines that are selected for T2 generation.

Plant Name	Copy Number	GOI PCR	Plant Name	Copy Number	GOI PCR
Line 1	1	+	Line 21	1	+
Line 2	2	+	Line 22	4+	+
Line 3	3	+	Line 23	1	+
Line 4	4+	+	Line 24	3	+
Line 5	4+	+	Line 25	2	+
Line 6	2	+	Line 26	1	+
Line 7	1	+	Line 27	2	+
Line 8	2	+	Line 28	1	+
Line 9	1	+	Line 29	1	+
Line 10	1	+	Line 30	2	+
Line 11	1	+	Line 31	4+	+
Line 12	1	+	Line 32	2	+
Line 13	2	+	Line 33	2	+
Line 14	3 or 4	+	Line 34	4+	+
Line 15	4+	+	Line 35	4+	+
Line 16	4+	+	Line 36	4+	+
Line 17	4+	+	Line 37	4+	+
Line 18	3 or 4	+	E. con 3	0	-
Line 19	4+	+	E. con 4	0	-
Line 20	4	+			
E. con 1	0	-			
E. con 2	0	-			

2.1.3 Overexpression of OC-I in soybean lines

Using constructs prepared by Dr Eugene Makgopa in the Christine Foyer Lab, Faculty of Biological Sciences at the University of Leeds independent transformed transgenic lines were produced by Dr Kan Wang at Iowa State University (Plant Transformation Facility, Iowa State University, USA) (Quain et al., 2014). Three independent lines expressing OC-I (SOC-1, SOC-2 and SOC-3) were selected for further analysis together with the soybean wild type (*Glycine max* cultivar Williams 82) (Makgopa, 2014).

2.2 Seed sterilization

The vapor-phase sterilization protocol (Clough, S. and Bent, 2000) was used to sterilize the seeds. Seeds of the *Arabidopsis* wild type (WT) and transgenic *Arabidopsis* lines were placed in separate opened Eppendorf tubes in a fume hood with a beaker containing 100 ml of domestic bleach and 3 ml of concentrated HCl (1M). Seeds were then exposed to chlorine gas for two hours. After sterilization, the seeds were used immediately for growth on agar plates.

2.3 Growth conditions

2.3.1 The growth of transgenic *Arabidopsis* lines on agar plates

Seeds of the transgenic *Arabidopsis* lines and WT were grown in agar on half-strength Murashige and Skoog (MS) media containing 2.2 g MS basal medium, 0.1 g L⁻¹ Myo-inositol, 10 g sucrose and 0.5 g L⁻¹ 2-(N-morpholino) ethane sulfonic acid (MES) buffer in 1 L dH₂O. The pH was adjusted to 5.7 with potassium hydroxide (0.2M: KOH) and 10 g Agar was then added into it. The medium was autoclaved at 121°C and then poured into square sterile plates. Sterilised seeds were grown in the plates and placed in the cold-dark room for three days before transfer to controlled environment

chambers with a 16/8 h day/night photoperiod at 20°C, 60% humidity and a light intensity of 400 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for 10 days.

2.3.2 The growth of transgenic *Arabidopsis* lines on soil

Seeds of the transgenic *Arabidopsis* lines and the WT were sown in pots (5 cm x 5 cm) containing compost and placed in a dark cold room overnight to encourage germination. The pots were then transferred to controlled environment chambers with 16/8 h day/night photoperiod at 20°C, 60% humidity and a light intensity of 400 $\mu\text{mol m}^{-2} \text{s}^{-1}$. After 10 days, the plants were transplanted into a new medium potting tray (William Sinclair Horticulture Ltd, UK) and grown under glasshouse conditions for a 12 weeks.

2.3.3 The growth of transgenic wheat lines on soil

Seeds of transgenic lines expressing OC-I and WT of wheat were sown in pots (5 cm x 5 cm) containing compost and placed in a controlled environment under glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Every two week, the plants were replaced it to a containers that are about 2 to 4 inches larger in diameter.

2.4 Phenotypic analysis

Differences between the transgenic *Arabidopsis* lines and the wild type will be explored under optimal conditions. Parameters related to growth and development, such as germination, root architecture, shoot biomass, time to flowering, number of leaves and rosette diameter will be measured. The transgenic lines and WT were grown simultaneously and their relative positions were randomised in the growth chamber. Each measurement in the phenotypic analysis involved 24 plants per line and three replicates were performed for each experiment.

2.4.1 Germination efficiency and seedling survival

To examine seed germination in the wild-type (WT) and transgenic lines (*Arabidopsis*, soybean and wheat), a number of seeds per line were sown in agar plates or soil and placed in a controlled environment chambers. The number of seeds that germinated (the appearance of radicles) was then recorded. The seedling survival rate also was analysed as the number of viable seedlings after ten days (expressed as a percentage as the number of seeds sowed).

2.4.2 Root architecture

A Canon digital camera (EOS 450D) was used to capture images of 10 day-old seedlings of the WT and the transgenic lines, grown on ½ MS agar media plates. The length of the primary roots was measured using the software ImageJ (<https://imagej.nih.gov/ij/download.html>). The number of lateral roots was recorded and lateral root density was calculated as the number of lateral roots divided by the length of the primary root (Placido et al., 2020).

2.4.3 Shoot growth analysis

Transgenic *Arabidopsis* lines and WT were collected and separated into shoots and roots. The fresh shoots were immediately weighed and then placed in an oven at 80°C for two days. The dry shoots were weighed again and the number of leaves (excluding cotyledons) per rosette was counted. The time of flowering, which is the age at which the plants began to produce flowering stems, also was recorded. In addition, the distance was measured between the tips of the largest opposite pairs of leaves on each plant to determine the rosette diameter using the software ImageJ.

2.5 Seed yield

To examine the seed yield of the transgenic *Arabidopsis*, wheat and soybean lines and the WT, plants were harvested by hand after maturity. The whole harvested plant was used to estimate seed size, the number of seeds per plant and seed weight. Images of seeds were captured using a Canon digital camera (EOS 450D). The number of seeds in each of six plants was counted and the average weight of 100 seeds was recorded.

2.6 Leaf pigments

Leaves (100 mg) from both the transgenic and WT *Arabidopsis* lines were collected and ground in liquid nitrogen using a mortar and pestle; 80% (V/V) acetone was then added to each tube. The homogenates were centrifuged (Centrifuge 5804R, Eppendorf, UK) at $20.817 \times g$ at 4°C for ten minutes until a white pellet was observed in the tubes. The pigment contents (Chl a, Chl b, Chl a + b, and Carotene) were measured using a cuvette placed in spectrophotometer at absorbance of 645 nm and 663 nm. Eighty percent (V/V) acetone was used as a blank. The following equations (Lichtenthaler, 1987) were used to determine the pigment concentrations:

$$\text{Chl a } (\mu\text{g/ml}) = 12.25A_{663.2} - 2.79A_{646.8}$$

$$\text{Chl b } (\mu\text{g/ml}) = 21.50A_{646.8} - 5.10A_{663.2}$$

$$\text{Total chlorophyll Chl a + b } (\mu\text{g/ml}) = 7.15A_{663.2} - 18.71A_{646.8}$$

$$\text{Carotene } (\mu\text{g/ml}) = (1000A_{470} - 1.82 C_a - 85.2C_b)/198.$$

2.7 Total protein content extracted from *Arabidopsis* leaves

Samples of transgenic *Arabidopsis* lines and WT (three fresh leaves from each lines of 4-week-old plants) were ground in liquid N₂ using a precooled mortar. One millilitre of protein extraction buffer (AS08 300, Agrisera) was added for every 100 mg of fresh

weight. Samples were centrifuged at $20.817 \times g$ for 10 minutes and then the supernatants were collected in new 2 ml Eppendorf tubes. Then, the protein concentrations (mg/g) were measured using the Pierce Microplate BCA Protein Assay Kit (Thermo Scientific). The calibration curve was made using BSA (1 mg/mL) as standard with dilutions to 0.5, 0.25, 0.125 and 0.06325 mg/mL. Nine microlitres of sample or standard was added to the wells of a microtiter plate and then 4 μ L of compatibility reagent solution was added to the sample in each well. Plate was covered and incubated at 37°C for 15 minutes. Then, 260 μ L of BCA Working Reagent (WR) to each well and plate was covered and incubated at 37°C for 30 minutes. The plate was kept for cooling at room temperature for five minutes. The absorbance was read at 562 nm using a FLUOstar Omega plate reader (BMG Labtech GmbH, Ortenberg, Germany). The final concentration was calculated in terms of mg protein/g FW.

2.8 Total protein content extracted from seeds

Total protein extracts were prepared from dried, mature seeds of the WT and the transgenic Arabidopsis, soybean and wheat lines. These seeds were ground in liquid nitrogen into a fine powder using a chilled mortar and pestle; 0.2 g was then weighed out and placed in a tube containing 20% isopropanol (500 ml). The samples were subjected to a tube rotator for one hour and then centrifuged at 20.817 g for 10 minutes at 4°C. The supernatants were collected in new tubes, and 10 ml of cold acetone was added; the tubes were then thoroughly vortexed. The extracts were incubated at -20°C overnight. Next, the samples were centrifuged for 30 minutes at 20.817 g at 4°C. The pellets were dried at room temperature for one hour and suspended in a protein extraction buffer (0.5 ml, Agrisera). This was followed by vortexing until the

pellet dissolved. The protein concentrations (mg/g) were measured as described in Section 2.8.

2.9 Production of soy flour and isolation of soy flour protein isolates (SPI)

The extraction of soy flour was carried out in Professor Brent S. Murray's Lab, School of Food Science and Nutrition, University of Leeds. The extraction was conducted under optimal conditions of flour-to-water ratio, temperature and pH etc. Briefly, the soybean seeds were ground to a fine powder (soy flour); the defatted soy flour was extracted using the solvent hexane (99%) at a ratio of 5:1. An alkali solution (pH 8.5) was then added to the defatted soybean flakes to remove carbohydrates and produce a soy protein concentrate (SPC). Next, acidic water (pH 4.5) was added and the samples were centrifuged to produce the soy protein isolate (SPI).

2.10 SDS gel electrophoresis and protein staining

Protein samples (10 µg) of transgenic Arabidopsis, soybean, wheat and WT plants were mixed with 4x Laemmli Sample Buffer (Bio-Rad, Herefordshire, UK) containing 0.1% β-mercaptoethanol and boiled at 80°C for ten minutes. Protein samples (20 µL) and PageRuler™ Pre-stained Protein Ladder (5 µL; Thermochemical, Paisley, UK) were loaded onto 4-20% Mini-PROTEAN® TGX™ precast gels (Bio-Rad, Herefordshire, UK). The gels were run using the Bio-Rad system (Bio-Rad, Herefordshire, UK) in an SDS running buffer at 120 V for one hour. After separation of proteins by SDS-PAGE, the gels were stained in a staining solution, Quick Coomassie Stain (Generon, 11 Whittle Pkwy, Slough, UK), for one hour to visualise protein bands, and a gel picture was taken using an INGENIUS gel imager (Syngene, Cambridge, UK).

2.11 Western Blotting

Western blotting was used to detect specific proteins in the protein mixtures, as described in Figure 2.5. The Trans-Blot Turbo System (Bio-Rad, UK) was used to transfer the proteins onto nitrocellulose membrane according to the manufacturer's guidelines. The nitrocellulose membrane was incubated in 5% skimmed milk powder (Marvel) prepared in TBS-T (25 mM Tris, 150 mM NaCl, 2 mM KCl, 0.1% Tween, pH 7.4) overnight on a rocking agitator at 4°C, to block non-specific proteins binding to the membrane. The membranes then were incubated with the following primary antibodies: RbcL; Rubisco large subunit, form I and form II (AS03 037, Agrisera, Sweden) PsbA; D1 protein of PSII (AS05 084, Agrisera, Sweden) and PsbA; D1 protein of PSII, phosphorylated (AS13 2669, Agrisera, Sweden) in 1:10000 dilution were then diluted in 5% skimmed milk powder in TBS-T. The nitrocellulose membranes were soaked in solution of primary antibodies and kept on a rocking table overnight at 4°C. Next, the membranes were washed in TBS-T three times for 15 minutes each time and then transferred to TBS-T buffer containing 5% skimmed milk powder with a 1:10000 dilution of horseradish peroxidase (HRP)-conjugated anti-rabbit secondary antibodies (AS09 602, Agrisera, Sweden). This was kept on a rocking table for two hours at room temperature. The membranes were then again washed in TBS-T three times for 15 minutes each time. Proteins bands were then visualised by washing the membranes for 5 minutes in Chemiluminescence substrate (SuperSignal™ West Pico PLUS, Thermo Scientific, Leicestershire, UK) and recorded using an INGENIUS gel imager (Syngene, Cambridge, UK).

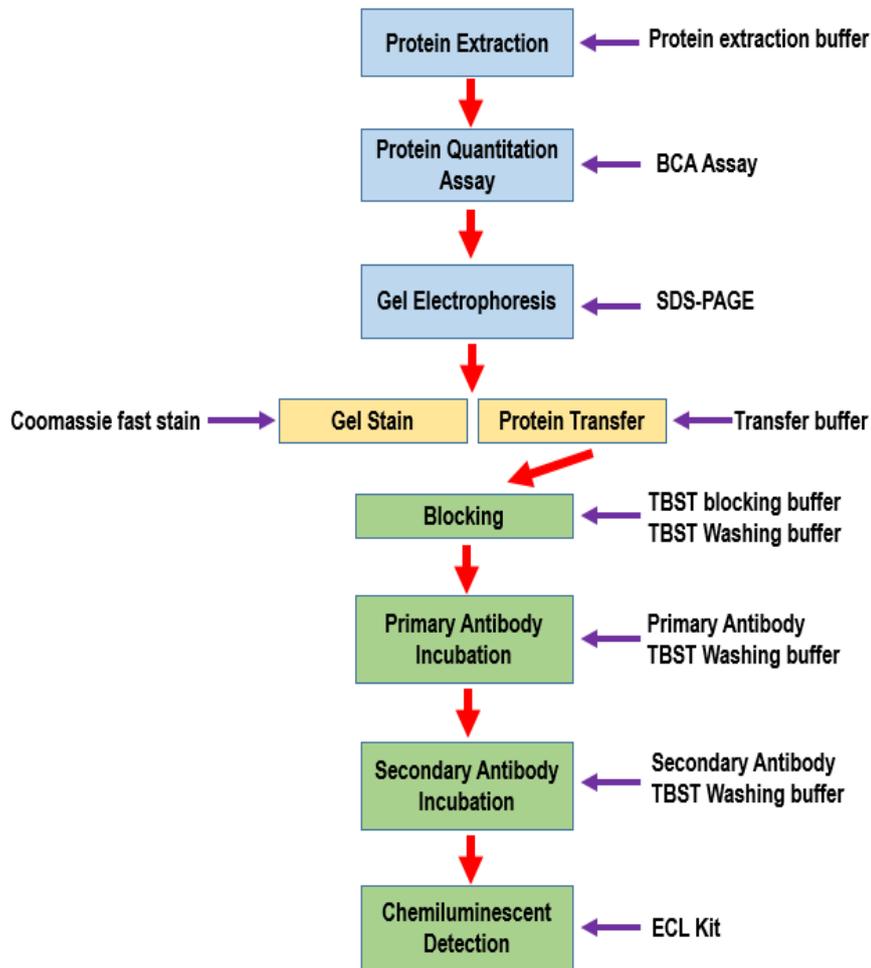


Figure 2.5: A flowchart of the steps in the protein extraction method and western blot procedure used in this study. Proteins were extracted and SDS-PAGE gels were prepared. After separation of proteins by SDS-PAGE, one gel was stained in staining solution Quick Coomassie Stain to visualise protein bands. Another gel was used for electrotransfer of the protein to nitrocellulose membrane.

2.12 Selection on kanamycin plates

The antibiotic kanamycin A was used as a selection marker to select the T2, T3 and T4 generations of transgenic wheat lines expressing OC-I. Transgenic lines and WT seeds were grown on ½ MS agar media sterile magenta GA-7 vessel containing kanamycin A (50mg/ml) (Pan et al., 2010). These plates were placed in a dark room at 20°C for five days and then transferred to controlled environment chambers (20/15°C day/night temp., 16hr day length, 350-500 $\mu\text{Em}^{-2}\text{s}^{-1}$) for six hours. The plates were then covered with aluminium foil and kept in the dark at 20°C for five days. Seedlings resistant to Kan were identified based on the survival phenotypes, i.e. the appearance of the coleoptile and the first leaf, and long roots that penetrated into the Kan selection media. These seedlings were then transplanted into soil to obtain T2 generation seeds. The procedures were repeated again thereafter to obtain T3 and T4 generations seeds.

2.13 High light treatments

Transgenic and WT *Arabidopsis* plants were grown in compost in controlled environment chambers at 20°C and 60% humidity with a 16/8 hour day/night photoperiod at low light intensity ($250 \mu\text{mol m}^{-2} \text{s}^{-1}$) for five weeks. The WT and transgenic lines were grown at the same time and their relative positions were randomised in the growth chamber. Thereafter, half of the plants were transferred to high light (HL) at ($800 \mu\text{mol m}^{-2}\text{s}^{-1}$) for eight hours. Six plants per line were involved in each experiment, each of which was repeated three times. Leaf samples were collected at each time point and placed in liquid nitrogen. Samples were stored in -08°C for further analysis.

2.14 The use of norflurazon and lincomycin to inhibit chloroplast processes

Norflurazon (NF), a herbicide that inhibits the carotene biosynthesis pathway (Burhans and Heintz, 2009), and lincomycin (LINCO), an antibiotic that inhibits protein synthesis in the chloroplasts (Yakandawala et al., 2003) were used to study the effects of OC-I on chloroplast to nucleus retrograde signalling. CYS, PC and WT seeds were sterilized (see Section 2.2) and grown on half-strength MS media (see Section 2.3.1) supplemented with NF (5 μM), LINCO (500 μM) and EtOH as a control. Plates were kept for five days in a controlled environment (described in Section 2.3.1). Five-day-old control and treatment seedlings were photographed with a Canon digital camera (EOS 450D). RNA was then extracted from both control and treatment seedlings following the NucleoSpin® RNA Plant kit protocol described in Section 2.17.3. The levels of *LHCA*, *LHCB1* and *LHCB2* transcripts were measured using quantitative real-time PCR (QPCR), as described in Section 2.17.3.2.

2.15 Photosynthetic carbon assimilation

Photosynthetic CO₂ assimilation rates were measured on fully expanded leaves of transgenic and WT *Arabidopsis* plants (three fresh leaves per genotype of 5-week-old plants) that had either been maintained for 8 hours under low light conditions or had been exposed for eight hours under high-light conditions (800 $\mu\text{mol m}^{-2} \text{s}^{-1}$) using a portable photosynthetic gas exchange system (Model LI-6400XT; LI-COR Biotechnology UK Ltd, St. John's Innovation Centre, Cambridge, UK), as described by Soares et al. (2008). The LI-6400XT delivered a light intensity of 800 $\mu\text{mol.m}^{-2}.\text{s}^{-1}$ and a CO₂ level of 400 $\mu\text{mol mol}^{-1}$. This infrared gas analyser system was set to a light intensity of 250 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and an atmospheric CO₂ level of 400 $\mu\text{mol m}^{-2} \text{s}^{-1}$ in the leaf chamber. One leaf of 4- week- old- plant was inserted into a 3x2 cm 6400-02B LED gas exchange chamber. Each measurement was performed at 20°C.

Photosynthesis was allowed to stabilise under these conditions in the leaf chamber for 15 minutes prior to measurement.

2.16 Total protease activity and cysteine protease activity

Protease activities were determined using the Abcam's Assay Kit (ab111750, Discovery Drive, Cambridge Biomedical Campus, Cambridge, UK). This quantitative method uses a highly quenched, fluorescein isothiocyanate (FITC)-labelled casein as a general protease substrate. The FITC-casein substrate is broken down into small peptides by proteases activities in the plant samples resulting in a decrease in fluorescence quenching. The fluorescence of peptide fragments was estimated at an excitation/emission (Ex/Em) wavelength of 485/530 nm. The assay was performed according to the manufacturer's instructions. In brief, leaf samples of transgenic *Arabidopsis* lines and WT (three fresh leaves from each lines of 5-week-old plants) were extracted using Assay Buffer (1:4 ratio) and placed into microcentrifuge tubes. The samples were then centrifuged (Centrifuge 5804R, Eppendorf, UK) at $20.817 \times g$ at 4°C for five minutes until the supernatants were clear. A standard curve was produced with FITC (as the protease substrate) at dilutions of 0, 0.05, 0.1, 0.15 0.2, and 0.25 nmol/well. Five microliter samples of control, leaf extracts or standard solution were added to each of the wells on a microplate; Assay Buffer was then added to a final volume of 100 μ L/well. Next, reaction mixture of Assay Buffer (48 μ L) and protease substrate solution (2 μ L) were added to the control and leaf extract wells. To determine cysteine protease activity, the specific cysteine protease inhibitor E-64 (5 mM; Sigma, Aldrich, Dorset, United Kingdom) was added to one set of samples. The Ex/Em was read at 485/530 nm using a FLUOstar Omega plate reader (BMG Labtech GmbH, Ortenberg, Germany). This measurement was designated reading 1 (R1) at time 1 (T1). The plate then was covered and incubated at room temperature for 30

minutes and the Ex/Em was read at 485/530 nm; this was designated Reading 2 (R2) at time T2. The unquenched FITC fluorescence generated by proteolytic digestion of the substrate is represented as $\Delta\text{RFU} = \text{R2} - \text{R1}$. The FITC standard curve was plotted (Figure 2.6A) to obtain the amount (B nmol) of FITC generated between T1 and T2 in the reaction wells. Protease activity could then be calculated using the equation shown in Figure 2.6B. Then, data were presented as $\mu\text{mol}/\text{min}/\text{mg}$.

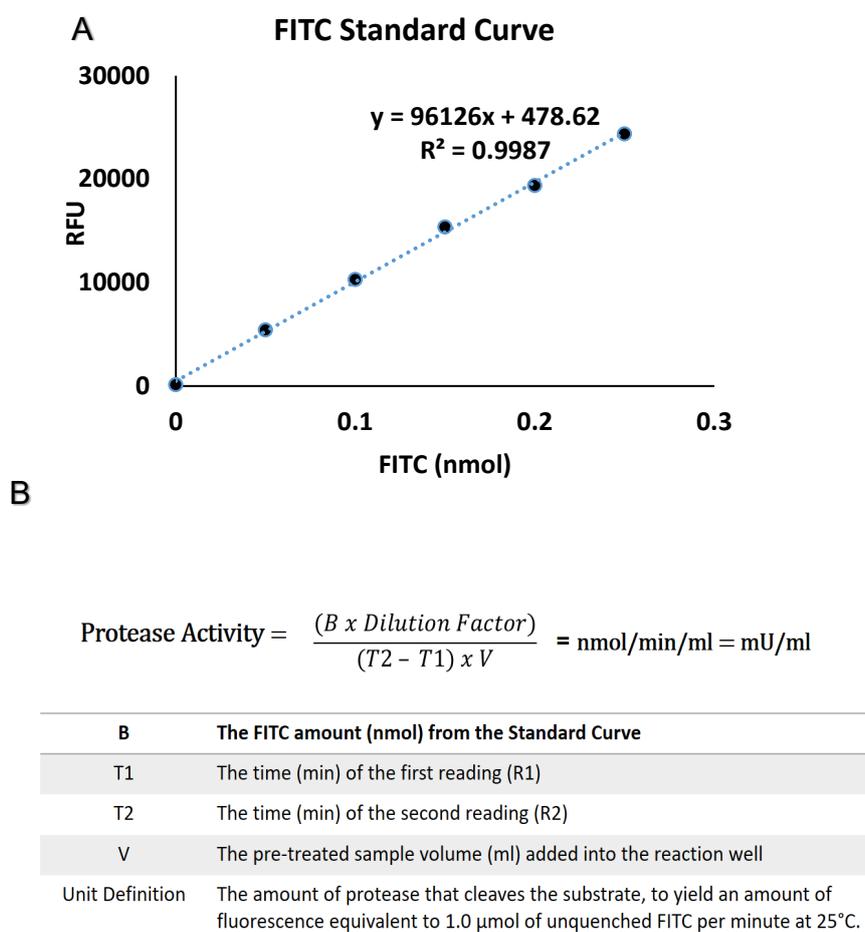


Figure 2.6: The FITC standard curve (A) and equation (B) used to calculate protease activity.

2.17 Nucleic acid extraction

2.17.1 DNA extraction from transgenic Arabidopsis expressing OC-I and WT

leaves

Leaf samples were collected from 4-week-old transgenic lines (CYS and PC) and WT plants for the extraction of DNA. In total three leaves were harvested from each transgenic plant and three leaves from the WT plants. Leaf samples were ground in liquid nitrogen. DNA extraction was performed according to the procedure described by Lu (2011). Frozen samples were transferred to 1.5 ml Eppendorf tubes and 200 μ L of extraction buffer (200 mM Tris buffer, 250 mM NaCl, 25 mM EDTA, 0.5% SDS, pH 7.5) was added to each tube. The tubes were vortexed for five seconds and were then centrifuged at 20,817g for two minutes. The supernatant of the samples was transferred to a fresh tube and 300 μ L isopropanol was added, and the mixtures were centrifuged at 20,817g for ten minutes. The pellets were washed with 70% ethanol (EtOH) and then centrifuged at 20,817g for 10 minutes. The supernatant was removed, then the pellets were re-suspended in 100 μ L H₂O. The DNA concentrations were quantified using the Nanodrop (NanoDrop ND-1000, Thermo) method and samples were measured at 260/280 nm.

2.17.2 DNA extraction from wheat leaves and seeds

DNA was extracted from wheat leaves and seeds according to the manufacturer's instructions (Qiagen DNeasy plant kit; Qiagen Ltd, Skelton House, Manchester, UK). For each sample, one hundred milligrams of tissue was ground in liquid nitrogen using a mortar and pestle. Homogenates were then transferred to microfuge tubes. Four hundred microliters of Buffer AP1 and 4 μ L RNase A were then added to each sample tube. Samples were mixed on a vortex and then incubated at 65°C for ten minutes. Next, 130 μ L of Buffer P3 was added to each of the tubes and then mixed. The tubes

were incubated on ice for five minutes and the lysates were centrifuged (Centrifuge 5804R, Eppendorf, UK) at $20.817 \times g$ for five minutes. The lysates were then transferred to QIAshredder spin columns, which were placed in 2 ml collection tubes and then centrifuged (Centrifuge 5804R, Eppendorf, UK) at $20.817 \times g$ for two minutes. The eluates were collected and transferred to 2 ml microfuge tubes and 1.5 μL of buffer AW1 were added to each. Six hundred and fifty microlitres of the mixture was transferred to a DNeasy spin column that was placed in a 2 ml collection tube and centrifuged at $20.817 \times g$ for one minute. The DNeasy spin column was placed in a new 2 ml collection tube to which 500 μl of buffer AW2 was added; the tube was then centrifuged at $20.817 \times g$ for two minutes. Next, the spin column was transferred to a new 2 ml microcentrifuge tube and 100 μl of Buffer AE was added to the membrane. The column was incubated for five minutes at room temperature, then centrifuged at $20.817 \times g$ for one minute. This final step was repeated to remove the column-bound DNA. The DNA concentrations were then quantified as described in Section 2.18.1.

2.17.2.1 The PCR reaction and program

PCR reactions were carried out in a total reaction mixture volume of 20 μl . Reaction mixtures were prepared using manufacturer's protocol and made up 50% of the reaction mixture. The PCR tubes were a master mix comprised of the following (per reaction): Biomix™ Red (10 μl ; Bioline, London, United Kingdom), 1 μM forward and reverse primers (2 μl) and sterile water (6 μl); the DNA sample was then added (2 μl ; 100 ng). Biomix™ Red was used according to the placed in a thermal cycler (Biorad, Hemel Hempstead, United Kingdom). The PCR program comprised a 5-second melting phase at 94°C, followed by 38 cycles (15 seconds at 94°C, 15 seconds at 60°C and 40 seconds at 72°C), with a final 5-minute extension step at 72°C. Products of PCR were analyzed using agarose gel electrophoresis.

2.17.2.2 Agarose gel electrophoresis

DNA molecules were amplified after PCR using 1.2% agarose gel electrophoresis. Agarose powder (1.2 g) was dissolved in Tris-acetate-EDTA (TAE) buffer (100 ml, pH 7.8) containing 40 mM Tris, 20 mM acetic acid and 1 mM EDTA. The solution was heated by microwave and Sybr®Safe (3 µL; Life Technologies, Paisley, United Kingdom) was added to the solution and left to set. Fifteen samples (10 µL) were loaded into each well and a 1 kb ladder (Fisher Scientific, Loughborough, UK) was used as a marker. Gel electrophoresis was run at 100V in 1x TAE buffer for 40 minutes. The bands were visualized under ultraviolet (UV) light and photographed by an INGENIUS gel imager (Syngene, Cambridge, UK).

2.17.3 RNA extraction

RNA was extracted from leaf samples (100 mg) of 4-week-old *A. thaliana* and control and treatment *A. thaliana* seedlings with inhibitors using the Spectrum™ plant total RNA kit (Sigma-Aldrich, Haverhill, UK). The samples were ground in liquid nitrogen using a mortar and pestle. Lysis Solution (500 µl) and β-mercaptoethanol (10 µl) were then added to the samples and the tubes were vortexed vigorously. The samples were incubated at 56°C for five minutes and then centrifuged (Centrifuge 5804R, Eppendorf, UK) at 20. 817 × g for five minutes. The clear lysate supernatants were filtered using NucleoSpin® filtration columns and placed into collection tubes that were then centrifuged (Centrifuge 5804R, Eppendorf, UK) at 20. 817 × g for one minute. The clarified flow-through lysate (200 µl) was collected and a binding solution (500 µl) was added and mixed by vortex. The mixture was transferred to binding columns and placed in collection tubes that were centrifuged (Centrifuge 5804R, Eppendorf, UK) at 20. 817 × g for one minute. A mixture of DNase digestion buffer (70 µl) and DNase I (10 µl) were added to the centre of binding columns filter and the samples were then

incubated at room temperature for 15 minutes. A first wash to the bound RNA was performed using wash solution 1 (500 μ l); the columns then were centrifuged (Centrifuge 5804R, Eppendorf, UK) at $20.817 \times g$ for one minute. A second wash was repeated twice using wash solution 2 (500 μ l) which was added to the columns before they were centrifuged (Centrifuge 5804R, Eppendorf, UK) at $20.817 \times g$ for 30 seconds. RNA was transferred to new columns to which elution solution (50 μ l) was added; the tubes were left to sit for one minute and then were centrifuged (Centrifuge 5804R, Eppendorf, UK) at $20.817 \times g$ for one minute. The concentration of RNA was measured using a Nanodrop (ND-1000 Spectrophotometer, Labtech International, UK), to determine the absorbance ratio of 260/280 nm (A_{260}/A_{280}), which is used to estimate RNA purity, and the absorbance ratio of 260/230 nm (A_{260}/A_{230}), which is used to estimate DNA quality. RNA samples with A_{260}/A_{280} ratios of 2.00 were selected as pure RNA and stored at -08°C for further analysis.

2.17.3.1 Synthesis of complementary DNA (cDNA)

A QuantiTect Reverse Transcription Kit (QIAGEN, Manchester, UK) was used to synthesise cDNA in accordance with the manufacturer's protocol. gDNAse Wipeout Buffer (2 μ L) was added to RNA samples (1 μ g) to remove genomic DNA then RNase-free water was added to make the final volume of 14 μ l. Sample tubes were then incubated for two minutes at 42°C . A mixture composed of 1X final concentration of Quantiscript RT Buffer (4 μ l), Quantiscript Reverse Transcriptase (1 μ l) and F/R primers mix (1 μ L)) was prepared and added to each sample. The same mixture was prepared without reverse transcriptase and was used as a negative reverse transcriptase control to check RNA contamination with DNA present in the sample. A thermal cycler (Biorad, Hemel Hempstead, UK) was used to run a PCR programme of

42°C and 95°C for 30 and three minutes, respectively. cDNA samples were sorted at -20°C.

2.17.4 Quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR)

PCR mixtures were prepared with 2x QuantiFast® SYBR® Green (10 µl; Qiagen, Manchester, UK), F/R primers mix (1 µM), cDNA (10 ng). The volume was made to 20 µl using RNase-free water. Reactions were performed on a low-profile 96-well plate (STARLAB, Milton Keynes, UK) and quantified using the Agilent Aria Mx Real-Time PCR system (Agilent Technologies LDA UK Limited Stockport, Cheshire, UK) and the plate was run according to the manufacturer's instructions. Three technical replicates were performed per sample and the same mixture was used without cDNA as a negative control. The PCR amplifications were programmed as follows: 95°C for five minutes followed by 95°C for ten seconds, then 60°C for 30 seconds. This cycle was repeated 40 times and a final extension was set at 60°C for 30 seconds. The melting curve was analysed to determine whether any mispriming had occurred. The data was analysed using the delta-delta CT method ($\Delta\Delta Ct$) (Livak and Schmittgen, 2001) and the expression ratio (R) was calculated using the equation below:

$$\text{ratio} = \frac{(E_{\text{target}})^{\Delta C_p \text{ target (control - sample)}}}{(E_{\text{Ref}})^{\Delta C_p \text{ Ref (control - sample)}}$$

E represents the amplification efficiency, C_p is the crossing point, *target* is a measure of gene expression and *Ref* is a measure of housekeeping gene, Actin and sand were used in this study.

2.17.5 Semi-quantitative reverse transcription-polymerase chain reaction (RT-PCR)

The semi-quantitative reverse transcription-polymerase chain reaction (RT-PCR) was used as described by Dallman and Porter (1994). Firstly, cDNA and a master mix were prepared as described in Sections 2.18 and 2.18.2.1, respectively. Each sample was divided into three tubes. The PCR program was run at 95°C for five minutes, followed by 38 cycles (10 seconds at 94°C, 30 seconds at 60°C, and 30 seconds at 60°C), with a final extension step of 60°C for 30 seconds. One set of tubes was removed after 20, 25 and 30 cycles. The PCR products were analyzed using agarose gel electrophoresis.

2.18 Determination of copy number in transformed wheat plants

The T-DNA copy number was determined by qPCR of the neomycin phosphotransferase II (nptII) copy number assay relative to a single copy of the wheat gene amplicon, GaMyb, normalised to a known single-copy wheat line (Milner et al., 2018). A TaqMan™ Copy Number Assay (Thermo Fisher Scientific Inc. Waltham, Massachusetts, USA) was used to estimate the copy number of transgenic wheat plants. In brief, genomic DNA (gDNA) was prepared as described in Section 2.18.2 and 5 ng/μL gDNA samples were diluted with 1× TE buffer (pH 8). The PCR mixture was then prepped to a final volume of 8.8 μL as follows: Master Mix (2X; 5.5 μL), TaqMan™ Copy Number Assay (20X; 0.55 μL), TaqMan™ Copy Number Reference Assay (20X; 0.55 μL) and nuclease-free water (2.2 μL). The mixture was added to a low-profile 96-well plate (STARLAB, Milton Keynes, UK) and gDNA was then added to the wells. The PCR amplification was quantified using the Mx3005P qPCR System (Agilent Technologies, LDA UK Limited, Stockport, Cheshire, UK). The PCR programmed was run as at 95°C for ten minutes followed by 95°C for 15 seconds,

then 60°C for 60 seconds. This cycle was repeated 40 times and a final extension was set at 60°C for 36 seconds. The data was analysed using $\Delta\Delta C_t$ method (see section 17.7.4).

2.19 PCR primers design

The Primer 3 website (<http://primer3.ut.ee/>) was used to design the primers used in this study (Table 2.2- 2.4). Primers had a length of between 20 to 25 nucleotides, with 50% GC content. The melting temperature (T_m) of the primers ranged between 60°C and 62°C with 1°C as difference between primers pair. Primer specificity was confirmed using BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>).

Table 2-2: List of forward and reverse primers used to amplify OC-I plasmid in *Arabidopsis* and wheat.

Genes	Primers Sequence 5'-3' (Forward / Reverse)
CYS. OC-I <i>Arabidopsis</i>	CATCGACAGGCTTGA ACTCC TCACCGAGCACAACAAGAAG
PC. OC-I <i>Arabidopsis</i>	TCA CCG AGC ACA ACA AGA AG AGC TCC TTG AAG TCC ATC CA
OC-I wheat	CGATCGGGTGAAATTCGGATCC GCTTCGTCAGGCTTAGATGT

Table 2-3: Lists of forward and reverse *Arabidopsis* primers used in qPCR and Semi-quantitative (RT-PCR)

Genes	Primers Sequence 5'-3' (Forward / Reverse)
Actin 2	GGCTCCTCTTAACCCAAAGG GAGAGAACAGCTTGGATGGC
OC-I	TCACCGAGCACAACAAGAAG CATCGACAGGCTTGA ACTCC
Sand family protein (<i>SAND</i>)	AATTAACAGTCCGCAACAGC GACCCAACAGAGTAGAACA
Photosystem I light harvesting complex A (<i>LHCA</i>)	TTGGCCATTGAGTTCTTAGCCA AAGCCGACTGTTGCACACAGA
Light-harvesting chlorophyll a/b-protein 1(<i>LHCB1</i>)	GGAACGGAGTCAAGTTTGGGA CAAATGCTCTGAGGAA
Light-harvesting chlorophyll a/b-protein 2(<i>LHCB2</i>)	AAGTCGTGAATGTACTTATTGGTG GGTGGTGTGGTTCATTAAGGT
Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (<i>rbcS</i>)	CCTCCGATTGGAAAGAAGAA TACACAAATCCGTGCTCCAA
photosystem II protein D1(<i>psbA</i>)	GTGGCTGCTCACGGTTATTT CCAAGCAGCCAAGAAGAAGT
Photosystem I reaction center subunit II (<i>psbD</i>)	CCGTCCCAAATCCCTCTCCTTC AGAAGACCACCGGTGCTTCCAG

Table 2-4: Lists of forward and reverse wheat primers used in qPCR and copy number.

Genes	Primers Sequence 5'-3' (Forward / Reverse)
Actin wheat	CTCTGACAATTTCCCGCTCA ACACGCTTCCTCATGCTATCC
OC-I	GGGAATGGGGCTCTCGGATGTA GGCATCCCCTTCCTTCACCTCA
Copy number QPCR	CTCCTGCCGAGAAAGTATCCA GCCGGATCAAGCGTATGC
Probe	[FAM]TGGCTGATGCAATGCGGCG[TAMRA]

2.20 Label-free quantitative proteomics analysis

A proteomics analysis was performed using mass spectrometry (MS) on protein extracts prepared from transgenic and WT wheat seeds as described previously (Min et al., 2016). The samples were prepared as described previously (Sections 2.9, 2.11 and 2.12). The samples (three biological replicates per line) were sent to the Advanced Mass Spectrometry Facility (School of Biosciences, University of Birmingham, Birmingham) for proteomic analysis. Trypsin digestion was performed using 10 μ L of protein spots excised from the gel to which 100mM ammonium bicarbonate (40 μ L: pH 8) was added. Then, 10 mM dithiothreitol (DTT: 50 μ L) was added to the samples that were incubated at 56°C for 30 minutes. Samples were then cooled to room temperature. Fifty millimetres of iodoacetamide (50 μ l) was added to each sample to alkylate cysteines; the samples were then incubated in the dark at room temperature for 30 minutes. Trypsin gold (Promega, Southampton, Hampshire, UK, 6 ng/ μ l) was added to the samples and incubated at 37°C overnight.

Following sample preparation, a full Fourier transform-based mass spectrometry (FT-MS) scan (m/z 360–1600) and subsequent higher-energy collisional dissociation (HCD) MS/MS scans of the 20 most abundant ions were performed with a dynamic exclusion setting of 15S. The MS and MS/MS scans were searched against Uniprot database using Thermo Scientific™ Proteome Discoverer™ version 2.2 software, SEQUEST-HT algorithm. Proteins having two or more high confidence unique peptides were accepted as a real hit.

Maxquant software (version 1.5.3.30) followed by Perseus software (version 1.5.8.5) were used to analyse MS/MS data. Maxquant software was used to analyse a large set of data provided from MS. Using Perseus made it possible to search the imputation of missing values of protein intensities from a normal distribution (with a width of 0.3

and a downshift of 1.8) and perform statistical analyses. Proteins with a ≥ 1.5 -fold change (FC) were considered significantly differentially abundant.

2.21 Bioinformatics analysis and construction of phylogenetic tree

The UniProt website (<http://www.uniprot.org/>) and Ensembl genomes database (<https://www.ensembl.org/index.html>) were used to search for cysteine protease sequences in *Arabidopsis* and wheat, respectively, in which a total of 280 and 431 cysteine protease sequences respectively were found. The sequence alignments were made using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). From each alignment, a circular phylogenetic tree was constructed using iTol (<http://itol.embl.de>). Other tools used in this study include Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) to predict 3D structure and conserved domains (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) to identify the protein domains.

2.22 Statistical analysis

The statistical analysis of the data was carried out using the one-way ANOVA and two-way ANOVA in order to examine whether there was a significant difference between the means of transgenic lines and the control (WT) under normal or stress conditions. Tukey's HSD (honestly significant difference) was used as a post-hoc test at a stringency level of $p < 0.05$. The asterisks indicate statistical significance as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$. Each measurement in the phenotypic analysis involved 24 plants per line and three replicates were performed for each experiment. All statistical analysis was performed using SPSS v.13 for Windows (Statistical Package for Social Sciences, Chicago). Perseus was used to search the imputation of missing values of protein intensities from a normal distribution (with a width of 0.3 and a downshift of 1.8) and to perform statistical analyses. Proteins with a ≥ 1.5 -fold change (FC) were considered significantly differentially abundant.

Chapter 3 : The characterisation of transgenic *Arabidopsis* plants expressing the rice cystatin, *oryzacystatin-I* (OC-I), in the cytosol

3.1 Introduction

Regulated proteolysis is an important cellular process. In this context, cysteine proteases (CPs) have been shown to fulfil crucial roles during seed germination (Toyooka et al., 2000), programmed cell death (PCD) (Solomon et al., 1999), root development (Quain et al., 2014), flowering (van der Hoorn, R. A. and Jones, 2004) and senescence (Belenghi et al., 2003; Beyene et al., 2006). The activities of CPs can be involved up- or down-stream regulation of plant responses to biotic and abiotic stresses. Cysteine protease inhibitors, which are called cystatins or phytocystatins, have been implicated in the control of stress tolerance and lifespan, but their precise functions remain poorly characterised. Exposure to abiotic stress during the development of transgenic *Arabidopsis* plants increased CP expression (Huang et al., 2007). Additionally, *Arabidopsis* plants that over-express two papain-like cysteine protease inhibitors, showed enhanced tolerance to salt, cold and drought (Zhang, X. et al., 2008).

Oryzacystatin I (OC-I), which was first identified in rice seeds (*Oryza sativa L. subsp. japonica*), is perhaps the best-characterised phytocystatin. Early studies on OC-I were performed to determine whether it could be used to enhance insect resistance (Benchabane, M. et al., 2010). Previous research in the Foyer lab has shown that OC-I expression in tobacco, soybean and other species has a marked effect on the shoot phenotype and increases stress tolerance (Prins et al., 2008; Quain et al., 2015). For example, transgenic tobacco plants expressing OC-I showed delayed leaf senescence and an accumulation of chloroplast proteins, particularly ribulose-1, 5-bisphosphate

carboxylase/oxygenase (Rubisco), together with increased biomass production and seed yield (Van der Vyver et al., 2003; Prins et al., 2008; Quain et al., 2015).

In previous studies OC-I was expressed in a range of plant species, including tobacco, soya and *Arabidopsis* without a specific targeting sequence. Transgenic *Arabidopsis* plants were produced that express OC-I in the cytosol (Quain et al., 2014). *Arabidopsis* plants were prepared by the floral dip method using the *Agrobacterium tumefaciens* strain GV3101 (Clough, S.J. and Bent, 1998) carrying the plasmid pTF101.1-Cys-I (see Section 2.1.1: Materials and Methods for details).

The studies reported in this chapter describe the characterisation of the transgenic *Arabidopsis* plants expressing OC-I in the cytosol. The shoot and root phenotypes of these plants were analysed to determine whether cytosolic CPs are important in the regulation of shoot and root development. In addition, experiments were performed to establish whether OC-I expression alters plant responses to abiotic stress conditions. The objectives of this studies reported in this chapter are:

1. To compare the root and shoot phenotypes of *Arabidopsis* plants with ectopic OCI expression in the cytosol (CYS1, CYS3 and CYS4) relative to the wild type (WT).
2. To determine both the effects of high light stress on gene expression and the abundance of the Rubisco and the photosystem II reaction centre D1 proteins in the wild type and transgenic plants.
3. To study the effects of the chloroplast inhibitors, norflurazon (NF) that inhibits carotenoid synthesis and Lincomycin that inhibits chloroplast translation, on the retrograde signalling pathways between the chloroplasts and nuclei that regulate photosynthetic gene expression.

3.2 Results

The present studies were performed on WT *Arabidopsis* and three independent transgenic lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4).

3.2.1 Confirmation of OC-I expression in *Arabidopsis* plants

The presence of OC-I in the leaves of 4-week-old transgenic *Arabidopsis* plants was confirmed by genomic DNA PCR analysis on a 1.2% agarose gel using the specific primers shown in Table 2.2 (see Section 2.19: Materials and Methods). A band with a size of 200 bp was present in the CYS lines. This band was absent in both the WT and the negative controls (Figure 3.1). Thus, the presence of the 200 bp band in the CYS lines but not in the WT confirms that all transgenic plants contained the OC-I coding sequence (Figure 3.1).

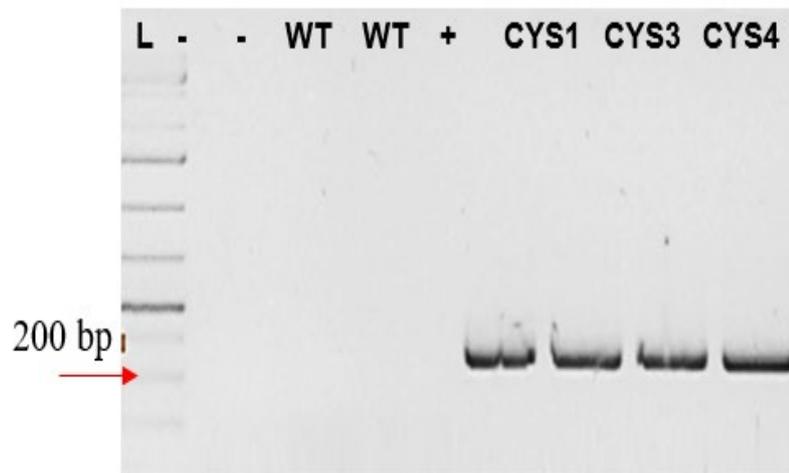


Figure 3.1: Agarose gel electrophoresis of DNA extracted from the leaves of wild type (WT) and 3 independent transgenic *A. thaliana* cell lines (CYS1, CYS3, and CYS4). Following PCR analysis, 10 μ l DNA (100ng) samples were loaded in the wells of 1.2% agarose gel and electrophoresis was performed at 100V for 40 min at room temperature. Lane L contained a ladder (1kb ladder). Lanes 1 and 2 were the negative controls, Lanes 3 and 4 are WT, Lane 5 contained the plasmid containing the 200bp segment of the OC-1 gene which was used to act as a positive control, and Lanes 6, 7 and 8 are CYS1, CYS3 and CYS4 respectively.

The levels of OC-I transcripts were determined using semi-quantitative RT-PCR and real-time PCR in the leaves of 4-week-old plants. The presence of OC-I transcripts was observed in all the CYS lines but not in the WT (Figure 3.2A). The levels of OC-I transcripts was similar in each individual transgenic line (Figure 3.2B).

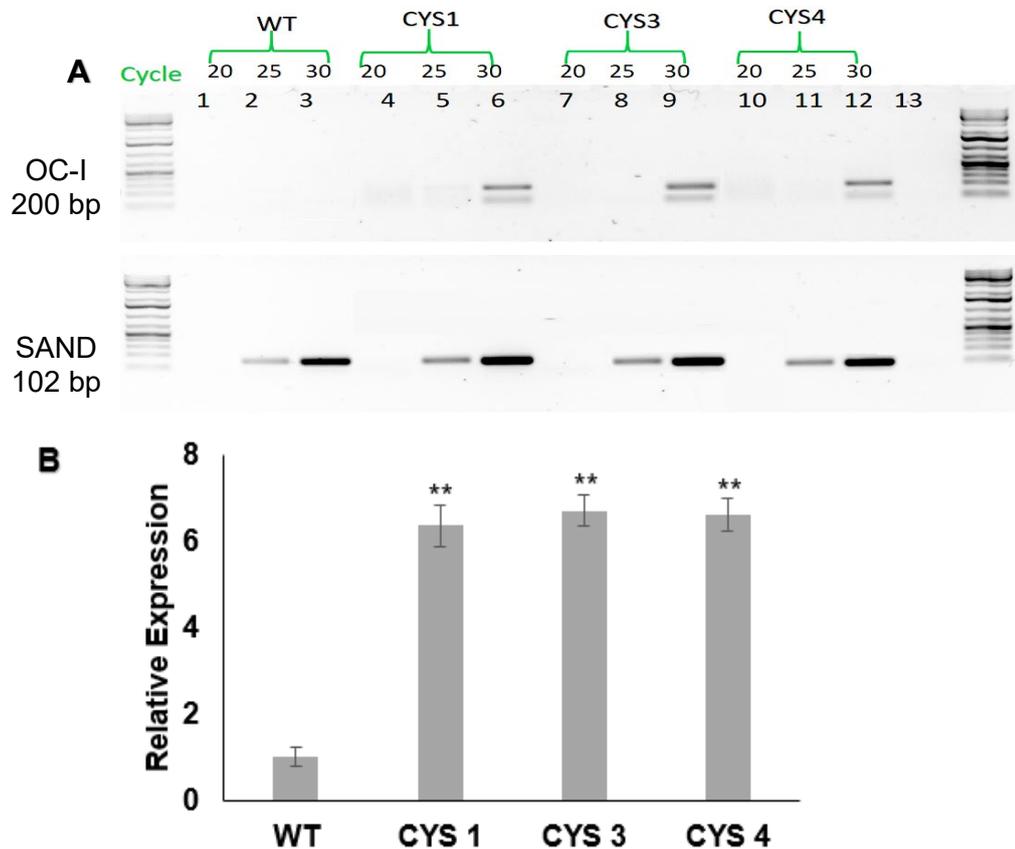


Figure 3.2: Expression analysis of OC-I in four-week-old *Arabidopsis* plants (CYS lines and WT plants) compared with SAND as a housekeeping gene. **(A)** Semi-quantitative PCR products were analysed on 1.2% agarose gel, with the expected PCR band of the OC-I gene (200 bp) and the expected PCR band of the SAND gene (102bp) as controls. **(B)** Relative expression of the OC-I transgene in three transgenic *Arabidopsis* plants expressing OC-I in the cytosol. WT was assigned a value of 1. The data was normalised to the *Arabidopsis* SAND gene. Means \pm SD for the plants in each line are indicated by bars. The asterisks indicate significant differences to WT plants (**p-value < 0.01, ANOVA).

3.2.2 Seed germination

Seed germination is a crucial stage of plant development. Germination is complete when the radicle protrudes through the seed coat. The level of seed germination was determined in 100 CYS1, CYS3, CYS4 and WT seeds 5 days after sowing on agar plates containing half-strength Murashige and Skoog (MS) media. The germination of CYS seeds was indistinguishable from that of WT (Figure 3.3A). However, the germination rate (the appearance of the radicle) was slightly delayed in the transgenic OC-I-expressing seeds compared to the WT.

The survival rate of the seedling also was analysed. The number of viable seedlings after 10 days (expressed as a percentage as the number of seeds sown) was the same in CYS lines and the WT. There were no significant differences in the seedling survival rates (Figure 3.3B).

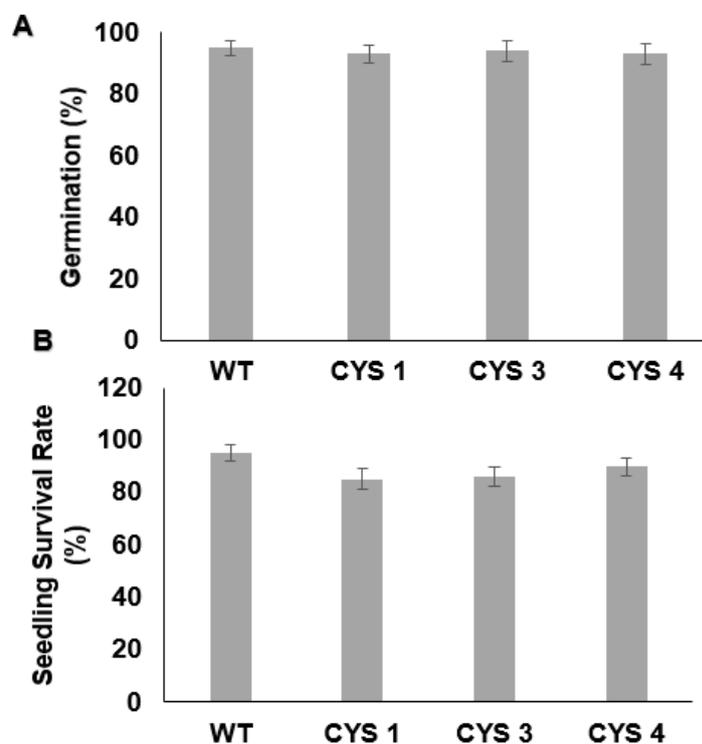


Figure 3.3: Comparison of (A) seed germination and (B) seedling survival rate in transgenic lines overexpressing OC-I (CYS1, CYS3, and CYS4) and wild type (WT). Bars represent means \pm SD (n=100 seeds).

3.2.3 Root architecture of transgenic line seedlings

A previous report on the lines studied here showed that the length of the primary roots of one-week-old seedlings was changed by OC-I expression in either the cytosol or the chloroplasts (Comadira, 2015). The length of the primary roots of 10-day-old seedlings grown on MS medium was therefore measured using ImageJ analysis of photographs (Figure 3.4A). The primary roots of the WT seedlings were approximately 2.5 cm in length at this stage whereas the length of the roots of all CYS lines was significantly longer (Figure 3.4B). Furthermore, the number of lateral roots and the lateral root densities were significantly higher in the CYS lines than the WT (Figure 3.4C and D). The CYS lines had an average of 7 lateral roots per seedling, whereas the WT had an average of 4 (Figure 3.4C). The density of the lateral roots was calculated as the number of lateral roots divided by the primary root length.

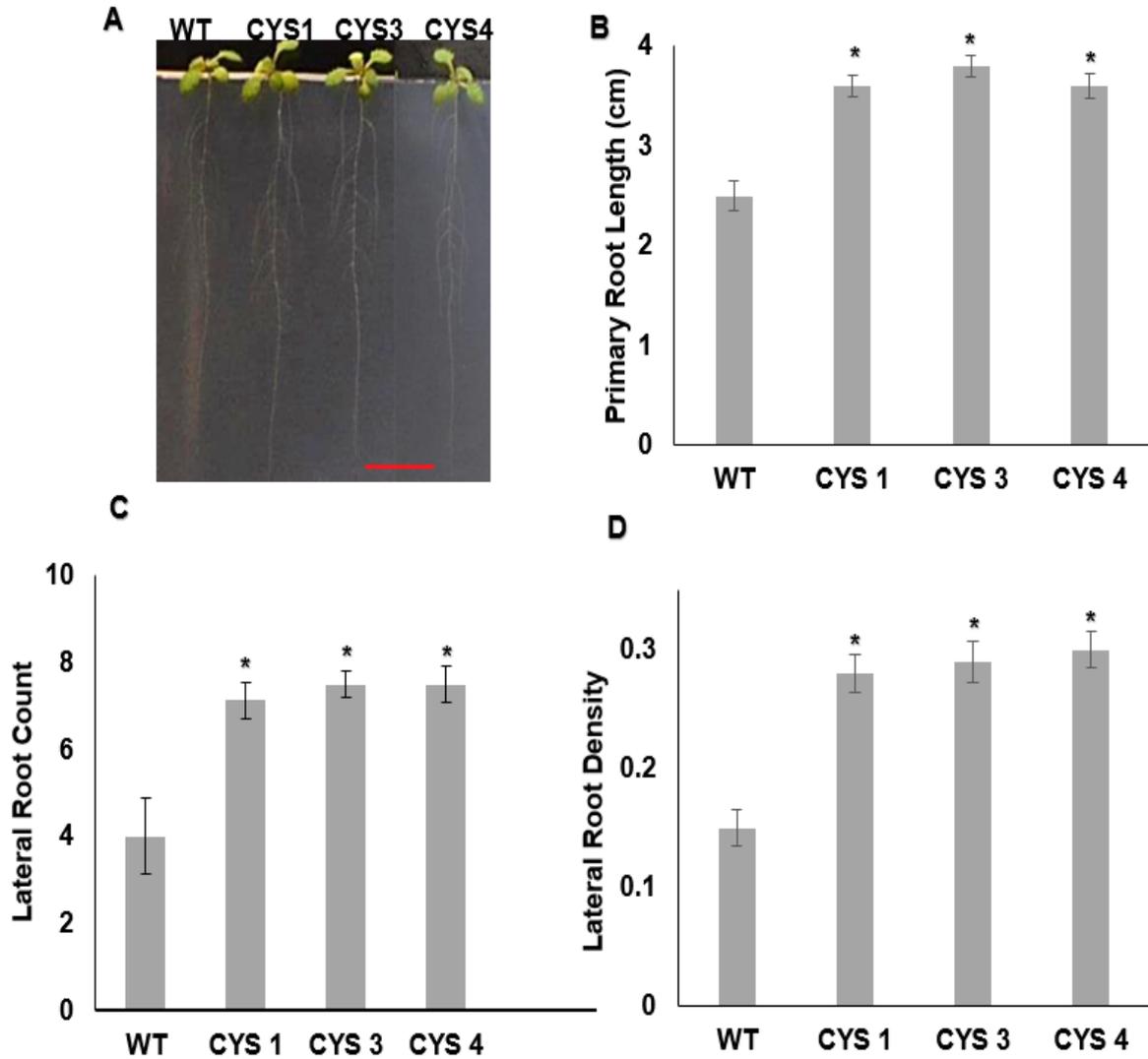


Figure 3.4: A comparison of root architecture in wild-type (WT) *Arabidopsis thaliana* and transgenic lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4) in 10-day-old seedlings: **(A)** photographs of 10-day-old seedlings grown on MS medium; **(B)** primary root length; **(C)** number of lateral roots; and **(D)** lateral root density. Mean values \pm SD ($n = 50$) are shown. The asterisks indicate significant differences to WT plants (* p -value ≤ 0.05 , ANOVA). Scale bar 3 mm.

3.2.4 Rosette morphology

Shoot phenotypes were characterised throughout development, using a range of parameters that were measured in weeks 4, 6, 8, 10 and 12.

3.2.4.1 Shoot phenotype, flowering, and biomass production

Representative images of the plants at 4 and 6 weeks after sowing are shown in Figure 3.5A and B. There were no visible differences in the vegetative development of the CYS lines and the WT at 4 weeks. However, there were clear differences in the time to flowering (Figure 3.5B). Whereas 70% of the WT plants flowered at week 5, very fewer CYS plants had flowered at this time point (Figure 3.6A). Flowering was clearly delayed in CYS lines (Figure 3.6A).

The biomass of the CYS1, CYS3, and CYS4 rosettes was significantly lower than that of the WT at 8 weeks after sowing (Figure 3.6B). However, from week 8 onwards, the shoot biomass significantly increased relative to the WT in all the transgenic lines expressing OC-I. The number of leaves and diameter of the rosettes were measured in all the lines at 4, 6, 8, 10 and 12 weeks after sowing (Figure 3.6C and D). The CYS lines tended to have fewer leaves and smaller rosette diameters than WT plants at all stages of shoot development (Figure 3.6C and D). However, the differences between the number of leaves and the rosette area in the CYS lines and the WT plants were significant only at weeks 10 and 12 (Figure 3.6C and 3.6D).

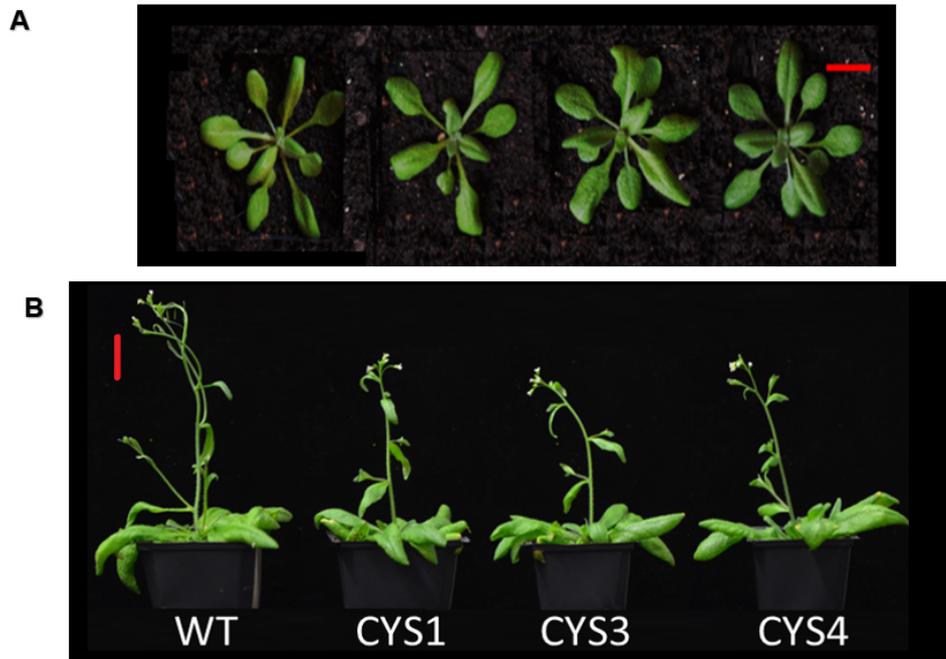


Figure 3.5: A comparison of the representative phenotype of plants of the transgenic lines overexpressing OC-I in the cytosol (CYS1, CYS3 and CYS4) to that of WT. The phenotype of **(A)** four-week-old and **(B)** six-week-old plants showed delayed flowering in the CYS lines grown on soil in a long day photoperiod (LD with 16 h day/8 h night) at 20°C, 60% humidity, at 400 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity. Scale bar 3 cm.

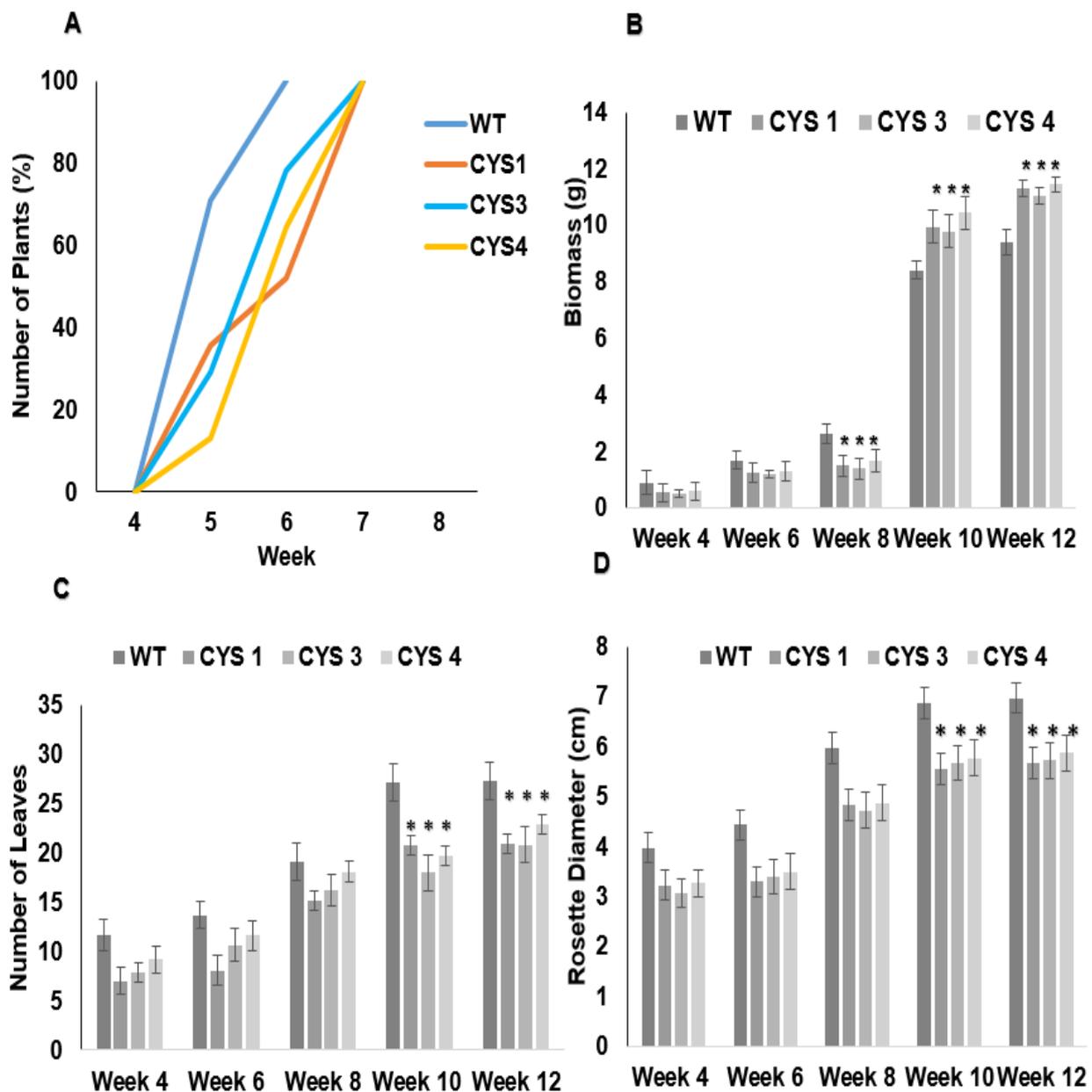


Figure 3.3: A comparison of shoot phenotype in the wild-type (WT) *Arabidopsis thaliana* and lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4) compared to the WT plants at 4, 6, 8, 10 and 12 weeks after sowing: **(A)** flowering time; **(B)** biomass; **(C)** number of leaves; **(D)** rosette diameter. Bars show the means \pm SD (n=24 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

3.2.4.2 Leaf pigments and protein contents

The leaf contents of chlorophyll, carotenoid pigments, and protein were compared in the WT, CYS1, CYS3, and CYS4 lines. The CYS lines had significantly less leaf chlorophyll and carotenoid pigments than the WT, particularly at the later stages of development (i.e. 10 and 12 weeks after sowing, Figure 3.7A and B). Leaf protein contents were similar in all the lines at 4 and 6 weeks after sowing. At weeks 8, 10 and 12, however, there was more protein in the leaves of the CYS lines than the WT (Figure 3.7C).

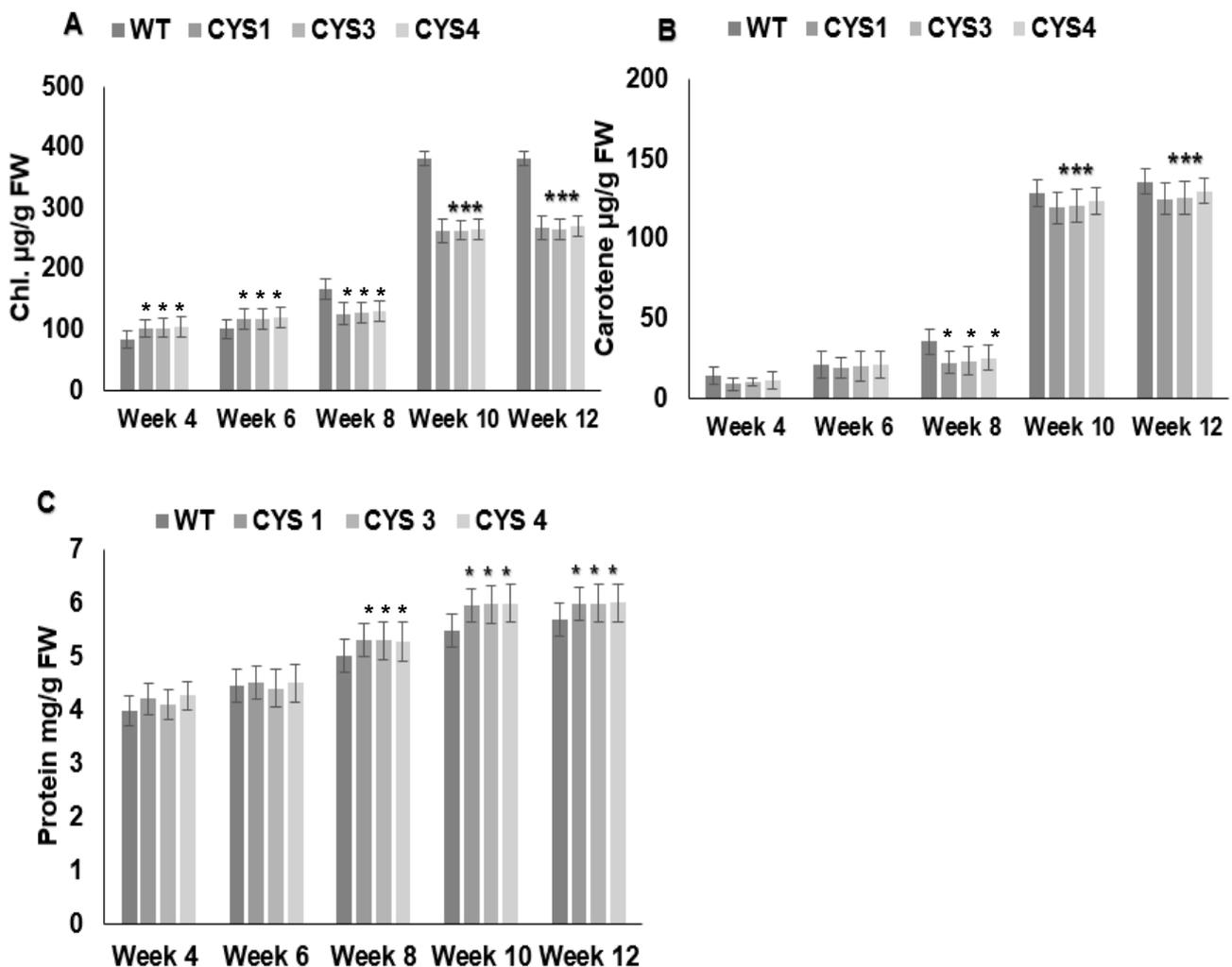


Figure 3.4: A comparison of the amount of (A) chlorophyll, and (B) carotenoid pigments and (C) protein in the leaves of the wild-type (WT) *Arabidopsis thaliana* and lines expressing *OC-I* in the cytosol (CYS1, CYS3 and CYS4). Bars represent means \pm SD (n=24 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

3.2.5 The role of OC-I in tolerance to high light stress

The effect of OC-I on plant responses to HL stress has not been studied previously. The effects of HL on photosynthesis rates, leaf protease activities and the abundance of proteins involved in photosynthesis were determined in the CYS lines and the WT. The levels of the D1 reaction centre protein of PSII and the large subunit of Rubisco were measured by Western blot. The CYS lines and the WT plants were grown in compost in controlled environment chambers at low light intensity (LL: $250 \mu\text{mol m}^{-2} \text{s}^{-1}$) for five weeks prior to exposure to HL stress (HL: $800 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 8 hours (see section 2.13; Material and Methods). The rates of photosynthetic carbon assimilation rates were similar in the CYS lines and the WT plants grown under LL conditions. Photosynthetic carbon assimilation rates were significantly decreased in the WT under HL conditions. The CYS lines had significantly higher rates of photosynthetic CO₂ assimilation than the WT under HL conditions (Figure 3.8).

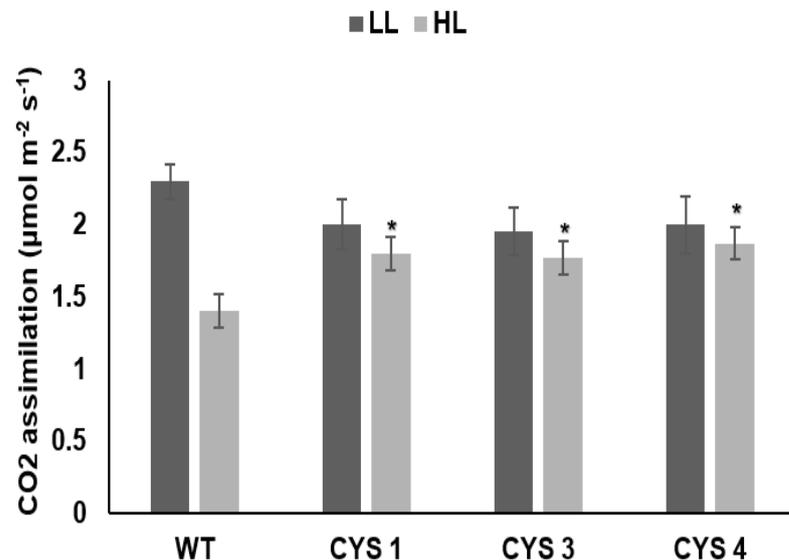


Figure 3.5: The effects of the expression of OC-I on photosynthetic CO₂ assimilation rates in the leaves of CYS lines and WT *Arabidopsis* plants grown under low light (LL) and high light (HL) conditions. Photosynthetic CO₂ assimilation in all leaves of the rosette was measured for CYS lines and WT plants that had been grown under low light ($250 \mu\text{mol m}^{-2} \text{s}^{-1}$) and then were transferred to a high-light environment ($800 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 24 hours. Bars represent means \pm SD (n=3 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

Total protease activities were measured in 5-week-old CYS and WT plants under LL and HL conditions using Abcam's Protease Activity Assay Kit (see Section 2.16: Material and Methods). The CYS lines had lower total protease activities than the WT under LL conditions. Exposure to HL increased the total protease activities of both CYS and WT lines (Figure 3.9A). However, the HL-induced increase of total protease activity in the CYS lines was less than observed in the WT (Figure 3.9A). The cysteine protease activity was compared in the CYS lines and the WT. The cysteine protease activity of the CYS lines significantly lower than the WT under LL conditions (Figure 3.9B). Exposure to HL increased the cysteine protease activity of the CYS lines but this increase was much less marked than that observed in the WT plants.

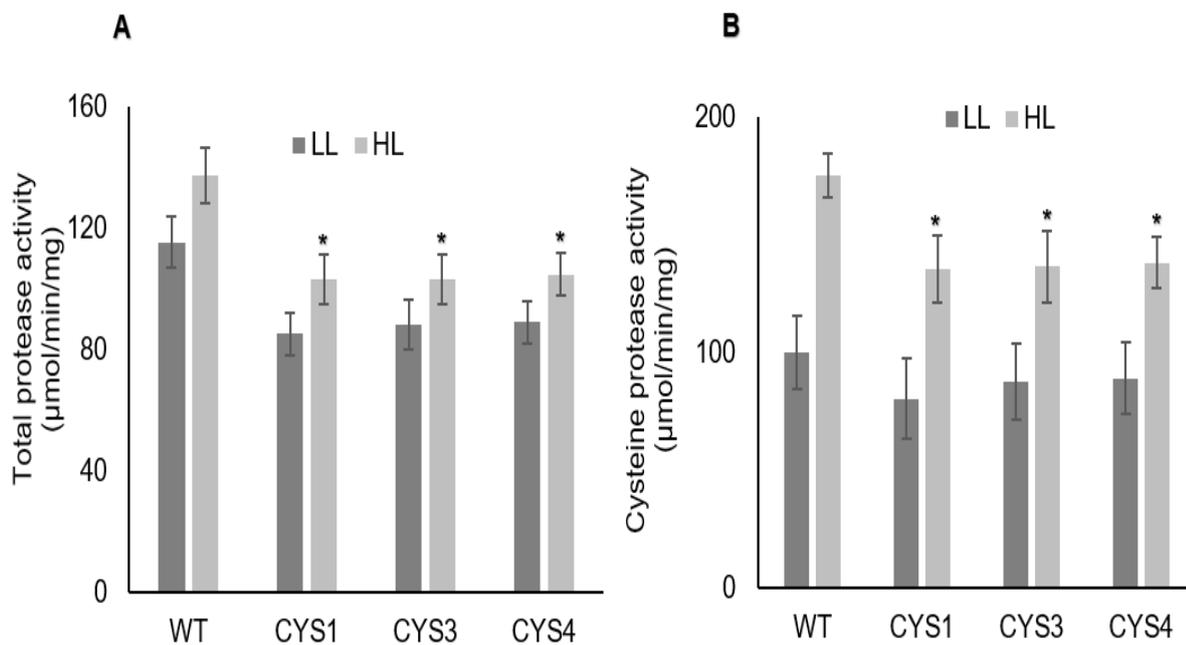


Figure 3.6: (A) Total protease activities and (B) cysteine protease activity in the leaves of CYS lines and WT *Arabidopsis* plants grown under low light (LL) and high light (HL) conditions. Bars represent means \pm SD (n=3). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

Western-blot analysis using specific antibodies was used to determine the abundance of the Rubisco large subunit, the D1 protein and the phosphorylated form of the D1, protein in the CYS and WT lines under LL and HL conditions. For these studies, leaf samples were harvested after 6h exposure to HL. The relative amounts of Rubisco, the D1 protein and the phosphorylated form of the D1 protein were lower in the CYS lines than the WT under LL conditions (Figure 3.10A). In contrast, the relative abundance of these proteins was much higher in the leaves of the CYS lines than those of the WT under HL stress (Figure 3.10B).

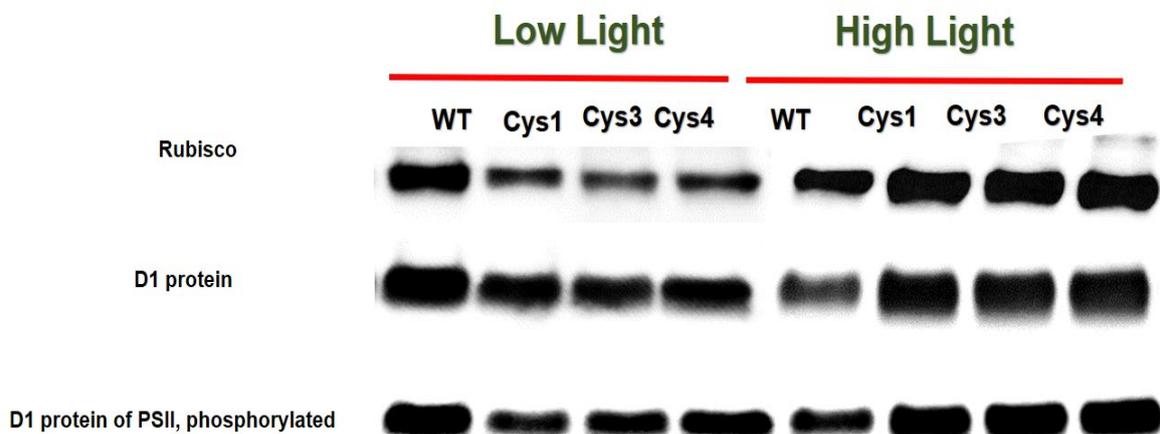


Figure 3.10: Western-blot analysis of Rubisco, D1 and D1 phosphorylated PSII proteins extracted from 5-week-old plants of the CYS lines and the WT grown under LL and then transferred to HL for 8 hours. Protein was extracted from leaves (10 μ g) using PAGE and then transferred to nitrocellulose membrane and analysed with primary antibodies: RbcL | Rubisco large subunit, anti-PsbA and anti-phosphorylated *PsbA* respectively.

Photosynthetic gene expression was compared in the CYS and WT lines. For these studies nuclear-encoded photosynthesis genes and chloroplast-encoded photosynthesis genes were selected for analysis. Plants were grown under LL and HL condition and samples were harvested for analysis by quantitative real-time PCR after 6h exposure to HL (Figure 3.11A and B). The levels of transcripts encoding the light harvesting chlorophyll a-b binding protein (*LHCA*), the light harvesting chlorophyll a-b binding protein 2 (*LHCB1*), the light harvesting chlorophyll a-b binding protein 2 (*LHCB2*), the small subunit of ribulose biphosphate carboxylase (*rbcS*) and the photosystem II D1 protein (*psbA*) were significantly higher in the leaves of the CYS lines than the WT under LL conditions (Figure 3.11A). However, the levels of transcripts encoding the Photosystem II D2 protein (*psbD*) were not significant different in the CYS lines and the WT under LL conditions. Under HL conditions, the levels of *LHCA*, *LHCB1*, *LHCB2*, *rbcS*, *psdA* and *psbD* transcripts in CYS lines were increased further relative to the WT than under LL conditions (Figure 3.11B). However, the levels of *psdA* transcripts were lower under HL than LL (Figure 3.11A and 3.11B).

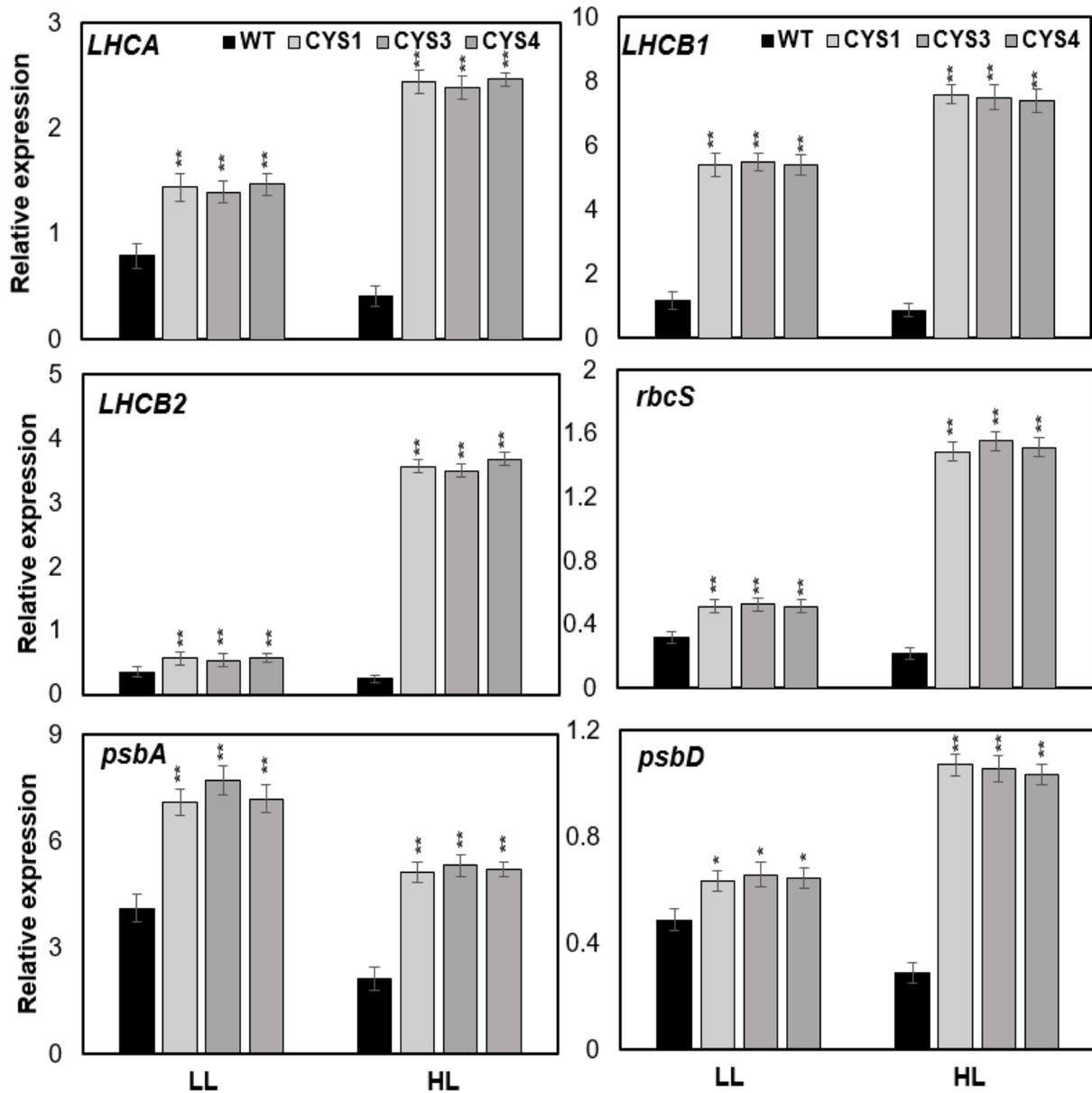


Figure 3.7: The effect of OC-I expression on the transcripts level of nuclear-encoded chloroplast and chloroplast-encoded photosynthesis genes treated under LL and HL conditions in 5-week-old CYS and WT *Arabidopsis* plants. LHCA: light-harvesting chlorophyll a binding protein; LHCB1: light-harvesting chlorophyll a binding B1; LHCB2: chlorophyll a-b binding protein 2 Ribulose biphosphate; *rbcS*: carboxylase small chain; *psbA*: photosystem II D1 protein; *psbD*: photosystem II D2 protein. The data were normalized to the actin gene. Bars represent means \pm SD (n=3 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, and p-value \leq 0.01 (**), ANOVA).

To analysis the effects of OC-I expression on chloroplast to nucleus retrograde signaling pathways that regulate photosynthetic gene expression the CYS and WT lines were grown for 7 days on agar plates containing ½ MS media with ethanol as a control (CONT) or media containing either Lincomycin (LINCO; 500 µM), or Norflurazon (NF;5 µM) (see Section 2.14: Materials and Methods). The seedlings of the CYS and WT lines showed a similar absence of chlorophyll in the presence of LINCO and NF compared to the control conditions (Figure 3.12A). However, whereas the WT seedlings did not have fully developed hypocotyls and cotyledons after 7 days to of LINCO treatment, the CYS lines had expanded hypocotyls and cotyledons in the presence of this inhibitor of chloroplast translation (Figure 3.12A).

The levels of *LHCA*, *LHCB1* and *LHCB2* transcripts were measured in the seedlings grown either in the absence or in the presence of inhibitors. The abundance of *LHCA* transcripts was lower in the CYS lines in the absence of inhibitors than the WT. In contrast, *LHCA* expression was significantly higher in the CYS lines than the WT in the presence of the inhibitors (Figure 3.12B). In addition, *LHCB1* transcripts were lower in the CYS lines than the WT in the absence of the inhibitors. However, the expression of *LHCB1* was higher in the CYS lines than the WT in the presence of LINCO or NF (Figure 3.12C). Moreover, the levels of *LHCB2* transcripts were lower in the CYS lines than the WT in the absence of the inhibitors. (Figure 3.12D). However, the LINCO and NF treatments significantly increased the abundance of *LHCB2* transcripts in the CYS lines compared to the WT (Figure 3.12D).

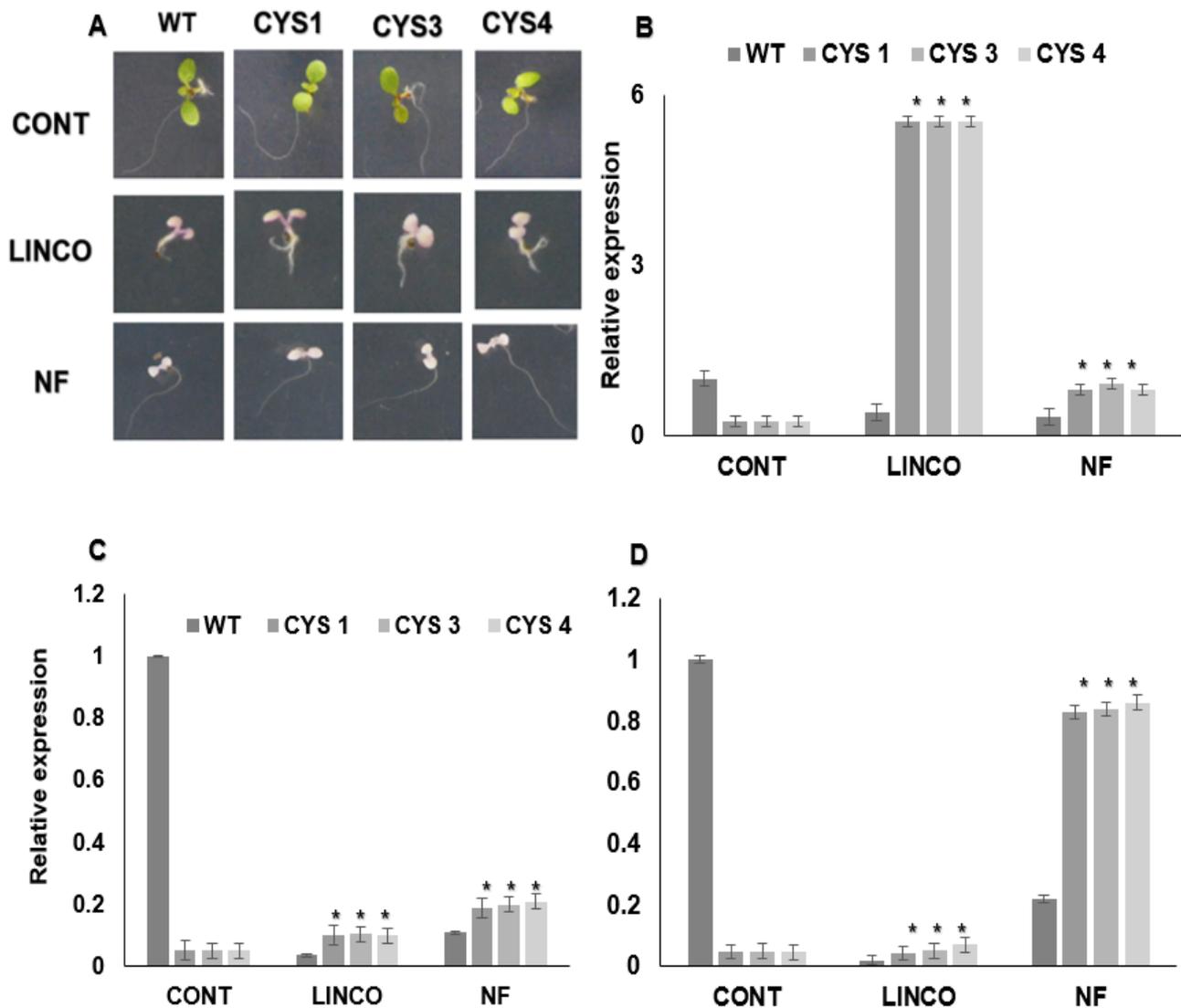


Figure 3.8: *Arabidopsis thaliana* lines expressing OC-I in the cytosol (CYS1, CYS3 and CYS4) and the WT were treated with lincomycin (LINCO), a chloroplast protein synthesis inhibitor, or norflurazon (NF; 5 μ M), an inhibitor that inhibits carotenoid synthesis. (A) Comparison of the representative phenotype of seedlings of CYS1, CYS3 and CYS4 and WT grown on MS-agar media containing either LINCO (500 μ M), or NF (5 μ M) for 7 days. The effect of LINC and NF on transcripts level of (B) LHCA, (C) LHCb1 and (D) LHCb2 in CYS lines and WT. Bars represent means \pm SD (n=24 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

3.3 Discussion

Phytocystatins have been implicated in the control of many important plant processes including stress tolerance, but their precise functions remain poorly characterized. The data presented in this chapter concerns the growth and flowering of T4 generation transgenic *Arabidopsis thaliana* lines, which express the OC-I protein in the cytosol (CYS), together with various aspects of leaf physiology and photosynthesis. There were clear differences in the growth and development of the CYS lines compared to the WT. The growth differences reported here are consistent with those in previous reports showing that cysteine proteases fulfil important roles in root and shoot development (Abe, K et al., 1987; Prins et al., 2008; Quain et al., 2014).

OC-I expression resulted in significantly slower seed germination, as defined by the appearance of radicle, in the CYS lines compared to the WT. This finding confirms previous observations in the Foyer lab (Roa-Roberts, 2014). The overexpression of AtCYS6, a phytocystatin that regulates seed germination, resulted in and lower CP activities and a slower germination rate in the transgenic *Arabidopsis* seeds (Hwang et al., 2009). In addition, the overexpression of *Brassica rapa* phytocystatin 1 (BrCYS1) delayed the germination of transgenic *Arabidopsis* seeds (Hong et al., 2007). Similar effects were found in potato tubers with ectopic expression of cereal cystatins (Munger et al., 2015).

OC-I expression increased the growth of the roots of the transgenic CYS lines relative to the WT controls. In addition, the CYS lines had more lateral roots and greater lateral root densities compared to the WT. It has previously been suggested that OC-I expression may influence shoot and root growth through alterations in strigolactone synthesis and signalling (Quain et al., 2014). Strigolactones are plant hormones that inhibit the branching of shoots but promote root growth.

The data presented in this chapter show that OC-I expression in the cytosol changes shoot development. The three independent CYS lines had smaller rosette diameters than the WT plants during vegetative growth up to the very later stages of development. The CYS rosettes accumulated less biomass than the WT during vegetative growth but had significantly more biomass than the WT at the later (reproductive) stages of development. Furthermore, the CYS lines had significantly fewer leaves than the WT throughout rosette development. In line with this observation, the time of flowering was delayed in all CYS lines compared to the WT. Decreases in cysteine protease activity have previously been associated with delayed flowering and delayed leaf senescence (Benchabane, M. et al., 2010). Papain-like cysteine proteases (PLCPs) are required for the development of anthers in *Arabidopsis* (Zhang, X.M. et al., 2009; Benchabane, M. et al., 2010). OC-I expression in tobacco was shown to delay flowering (Van der Vyver et al., 2003).

The effects of OC-I expression on shoot biomass accumulation were particularly evident in 10- and 12-week-old plants. At this stage, the transgenic plants were visibly larger and had significantly more biomass than the WT. These results may suggest that OC-I-dependent inhibition of CPs leads to improved biomass accumulation, as has been shown in previous studies on OC-I expression in the cytosol of tobacco (Prins et al., 2008). OC-I-dependent inhibition of cysteine protease activity was shown to limit protein degradation in the OC-I expressing tobacco leaves, suggesting that nitrogen remobilisation was changed. The effects of OC-I on nitrogen remobilisation may be responsible for the observed delay in flowering in the OC-I expressing tobacco leaves (Van der Vyver et al., 2003). However, the restriction of nitrogen remobilisation untimely has a beneficial influence on plant growth and biomass accumulation at the later stages of development. OC-I expression was shown to delay leaf senescence in

soybean, leading to much greater biomass accumulation, as was observed in the transgenic tobacco lines with OC-I expression (Quain et al., 2015). The CYS lines had lower leaf chlorophyll levels and lower carotenoid pigment contents than the wild type, particularly at the later stages of development. However, the levels of leaf protein were significantly higher in the CYS plants than the WT at weeks 8, 10 and 12. This finding is consistent with previous studies in other species. For example, OC-I expression in soybean led to an accumulation of leaf proteins (Quain et al., 2015). Taken together, these findings suggest that the expression of OC-I has marked effects on plant growth and development.

Previous studies of the role of OC-I in plant responses to environmental stresses have focused on drought (Quain et al., 2014), low nitrogen-induced senescence (Quain et al., 2015) and dark chilling (Cooper, 2016). Light is one of the most important environmental factors for plant growth, sunlight providing the energy that drives photosynthesis and so powers plant growth and development (Jiao et al., 2007; Kaiserli et al., 2015). However, exposure to excessive high light (HL) constitutes a stress that inhibits photosynthetic capacity (Mishra et al., 2012). The effects of HL stress on the photosynthetic CO₂ assimilation rates were therefore compared in the CYS and WT lines. Data are presented showing that photosynthetic CO₂ assimilation rates were higher in the CYS lines than the WT after exposure to HL stress. Moreover, the CYS lines retained higher levels of photosynthetic pigments than the WT following exposure to HL stress. These data suggest that the CYS lines are better protected against high light stress than the WT. Moreover, these findings confirm the results obtained in other species demonstrating that CYS lines in other species such as tobacco have a higher level of tolerance to different abiotic stress conditions (Van der Vyver et al., 2003; Prins et al., 2008). Cysteine proteases are activated by numerous

abiotic stresses, leading to premature senescence and PCD (Belenghi et al., 2003; Quain et al., 2014). OC-I expression in the CYS lines may prevent the stress-induced increases in cysteine protease activity that occur under stresses such as HL. The responses of leaf proteins to HL stress were different in the WT and CYS lines. The CYS lines accumulated more Rubisco and D1 proteins than the WT under HL stress. In addition, the levels of *RBCS* and *PSBA* transcripts were lower in all lines following exposure to HL stress. These data suggest that OC-I may prevent the degradation of the Rubisco and D1 proteins under HL stress because the observed increases in the abundance of these proteins was not a consequence of increased *RBCS* and *PSBA* expression.

The term “retrograde signalling”, is used to describe the pathways that transmit information from the chloroplasts and/or mitochondria to the nuclei in order to modify nuclear gene expression (NGE) (Leister, 2012). Changes in environmental conditions that alter photosynthesis or the metabolic state of the chloroplasts or mitochondria can influence NGE (Leister, 2012). Chloroplast to nucleus retrograde signalling pathways are particularly important in the responses of leaves to changes in irradiance because they allow efficient regulation of photosynthetic proteins in response to changing light levels. It is possible that signals from the organelles to the nucleus pass through the cytosol or they may be transferred through direct organelle/nucleus contact sites. A wide range of primary and secondary metabolites are thought to be involved in retrograde signalling. For example, peptides produced by the breakdown of damaged proteins inside the organelles may contribute to retrograde signalling (Møller and Sweetlove, 2010). The Genomes Uncoupled “Gun”, retrograde signalling pathway has been extensively characterised. For example, the *Arabidopsis gun* mutants, which were isolated using the chloroplast inhibitors lincomycin (LINCO), a chloroplast protein

synthesis inhibitor, and norflurazon, an inhibitor (NF) that inhibits carotenoid synthesis, have been extremely useful in the characterisation of retrograde signalling pathways (Wu et al., 2019). The LINCO and NF mutant screens involve the expression of photosynthesis-associated nuclear genes (PhANGs), which is decreased in the presence of these inhibitors (Woodson et al., 2013; Song, L. et al., 2018). The data presented here shows that the expression of OC-I in the cytosol exerts a strong influence on PhANG expression and hence chloroplast to nucleus retrograde signalling pathways. For example, the expression of the *LHCA* and *LHCB* genes, which were used as marker genes for the analysis of retrograde signalling in this study was significantly changed by OC-I expression. The levels of *LHCA*, *LHCB1* and *LHCB2* transcripts were greatly decreased in the WT plants in the presence of the LINCO and NF inhibitors, in agreement with previous studies (Karpinska et al., 2017). The data presented in this chapter shows that LINCO and NF treatments significantly increased the abundance of *LHCA*, *LHCB1* and *LHCB2* transcripts in the CYS lines compared to the WT. The *gun2*, *gun4* and *gun5* mutants accumulate high levels of *LHCB1* transcripts compared to the WT in the presence of NF (Voigt et al., 2010). The data presented here shows that the expression of photosynthesis-associated nuclear genes is upregulated when chloroplast biogenesis is blocked by LINCO and NF in the OC-I background, suggesting that cysteine protease-associated protein breakdown, which is blocked by OC-I, plays a key role in chloroplast to nucleus signalling. Further studies are required to determine the mechanisms by which OC-I expression in the cytosol regulates the retrograde signalling pathways.

In conclusion, the data presented in this chapter demonstrate that the cytosolic expression of OC-I not only influences plant growth and development but it also influences photosynthetic gene expression and leaf responses to HL. The findings

presented in this chapter provide new insights into the functions of cysteine proteases in physiological processes as well as stress tolerance. In this research and in previous studies, OC-I was expressed in transgenic plants without a targeting sequence. Hence, we assume that the gene product is located in the cytosol. The next chapter presents the findings of studies on transgenic lines in which OC-I is targeted to the chloroplasts. This will allow an examination of the effects of OC-I on processes within the chloroplasts as well as the cytosol.

Chapter 4 . The characterisation of transgenic *Arabidopsis* plants expressing the rice cystatin, *oryzacystatin-I* (OC-I) in the chloroplasts

4.1 Introduction

Chloroplasts house the photosynthetic processes that power plant growth and metabolism. They also play an important role in plant development (see section 1. 3). The degradation of chloroplast proteins in older leaves is crucial in the remobilization of resources to sustain the growth of younger leaves and grain filling. The mechanisms by which chloroplast proteins are degraded is not fully understood but this process can take place in vesicles that are associated with the vacuole as well as in the chloroplasts. Cysteine proteases and senescence-associated vacuoles (SAVs) are important in the degradation of chloroplast proteins during leaf senescence (see Section 1.3). However, the specific proteases and proteolytic mechanisms that are responsible for the degradation of chloroplast proteins such as Rubisco remain largely uncharacterised. Cysteine proteases are absent from *A. thaliana* chloroplasts (Majsec et al., 2017). However, a cysteine protease of the ovarian tumour-like cysteine protease family was detected in a proteomic study of pea chloroplasts by Makarova et al. (2000). The targeting of this cysteine protease to chloroplasts was confirmed by Bayer et al. (2011) using a transient transformation yellow fluorescent protein fusion construct. Earlier studies had shown that cysteine protease activity was present in the lumen of spinach chloroplasts (LP27) (Sokolenko et al., 1997). The 32 kDa form of the *Hordeum vulgare* cysteine protease (HvPAP14) was detected in the barley thylakoid lumen, indicating that the inhibitory pro-domain cleavage of this protein occurs in the lumen, where there is a low pH level that allows the recombinant enzyme to be activated (Frank et al., 2019). Furthermore, HvPAP14 activity may be localised at the

thylakoid membrane (Kramer et al., 1999; Järvi et al., 2013). It is generally thought that the degradation of Rubisco by cysteine proteases occurs in the vacuole (Thoenen et al., 2007). Rubisco levels were higher in the chloroplasts of cystatin-overexpressing tobacco plants (Prins et al., 2008). The discovery of HvPAP14 in barley chloroplasts may help to explain the results obtained with tobacco plants that overexpress cystatin. A partial degradation of Rubisco by HvPAP14 in the chloroplasts may precede the autophagic processes (Xiong et al., 2007) and the formation of the Rubisco-containing vesicles (Prins et al., 2008). Recently, HvPAP14 was confirmed to be a chloroplast protease and to have three forms with 40, 32, and 26 kDa proteins, which were detected in all stages of barley leaf development (Frank et al., 2019). Cysteine proteases are involved in chloroplast protein degradation (Thoenen et al., 2007; Carrión et al., 2013). The expression of phytocystatins may therefore effectively decrease protein degradation (Buet et al., 2019).

Oryzacystatin I (OC-I), which was identified in rice seeds (*Oryza sativa L. japonica*), is perhaps the best-characterized phytocystatin. This protein has been expressed in a range of transgenic plants, which were used to study the functions of cysteine proteases. Soybean plants that overexpress *oryzacystatin I* (OC-I) show enhanced branching and delayed senescence. Over-expression of OC-I slows vegetative growth in tobacco, *Arabidopsis* and soybean (Prins et al., 2008; Quain et al., 2014). In these studies OC-I was not expressed with a targeting sequence and it is hence assumed that the produce of the transgene is located in the cytosol. Despite the location outside the chloroplasts, Rubisco degradation was delayed in OC-expressing plants. Since phytocystatin has not previously been targeted to these organelles, the present study focussed on transgenic *Arabidopsis* plants, in which OC-I was targeted to chloroplasts. The production of the transgenic *Arabidopsis* plants expressing OC-I in the

chloroplasts was described previously (see Section 2.1.1: Materials and Methods). The following studies therefore address the question of whether OC-I expression in the chloroplasts alters shoot growth and development. A key question concerns whether OC-I expression in the chloroplasts has similar or different effects on the shoot phenotype to when expression is not targeted, and whether the abundance of chloroplast proteins is changed to enhances Rubisco protein accumulation particularly during age-related or stress-induced senescence.

The present study has two main aims. The first is to determine the effects of OC-I expression in the chloroplasts on the growth and shoot phenotypes of transgenic *Arabidopsis* plants. This analysis should allow the determination of whether OC-I influences processes directly within the chloroplasts as well as in the cytosol (Chapter 3). The second aim was to establish how OC-I expression in the chloroplast alters the plant response to stress.

The specific objectives of this chapter are:

1. To characterize the *Arabidopsis* plants with ectopic OC-I expression in the chloroplast (PC2, PC7 and PC 9) compared to the wild type (WT).
2. To determine the effects of high light stress on the abundance of the Rubisco and D1 proteins and transcripts in the transgenic plants.
3. To study the effects of the chloroplast inhibitors, norflurazon and Lincomycin, on retrograde signalling in the transgenic plants as well the effects of the presence of OC-I on gene expression.

4.2 Results

The present study, which was conducted in order to determine whether chloroplast cysteine proteases regulate shoot and root development in *Arabidopsis*, was performed on three independent transgenic lines expressing OC-I in chloroplast (PC2, PC7 and PC 9) as well as the WT plants.

4.2.1 Confirmation of OC-I expression in *Arabidopsis* plants

The OC-I protein was targeted to plastids using the, phosphoribulokinase (PRK) signal peptide (see Section 2.1.1: Materials and Methods). The presence of OC-I was confirmed in the leaves of 4-week-old transgenic *Arabidopsis* plants using PCR. Genomic DNA was analysed on a 1.2% agarose gel using the PRK forward primers and OC-I reverse primers shown in Table 2.2 (see Section 2.19: Materials and Methods). The presence of a 300 bp band in the PC lines provided evidence that all the transgenic plants contained the OC-I coding sequence (Figure 4.1).

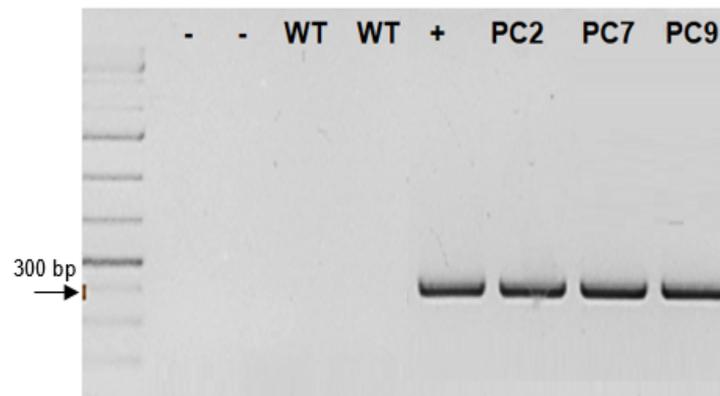


Figure 4.1: Agarose gel electrophoresis of DNA extracted from the leaves of the wild type (WT) and three independent transgenic *A. thaliana* lines (PC2, PC7, and PC9). Following PCR analysis, 10 μ l DNA (100ng) samples were loaded in the wells of a 1.2% agarose gel and electrophoresis was performed at 100V for 40 min at room temperature. Lane L contained a ladder (1kb ladder). Lanes 1 and 2 are the negative controls, Lanes 3 and 4 are WT, Lane 5 is a positive control amplified from a plasmid (pLBRPRKCys-I) containing a 300bp segment of the OC-1 gene and coding sequence of phosphoribulokinase (PRK) gene (+), and Lanes 6, 7 and 8 are PC2, PC7, and PC9 respectively.

Semi-quantitative RT-PCR and qRT-PCR using OC-I primers were performed to characterise OC-I expression in the leaves of 4-week-old transgenic plants. OC-I transcripts were detected in all the PC lines (Figure 4.2A), the PC7 and PC9 lines having the highest transcript levels (Figure 4.2B). Furthermore, the product amplified using SAND primers was detected in all samples, indicating that the RT-PCR cDNA template was present in the samples of both the WT and transgenic lines.

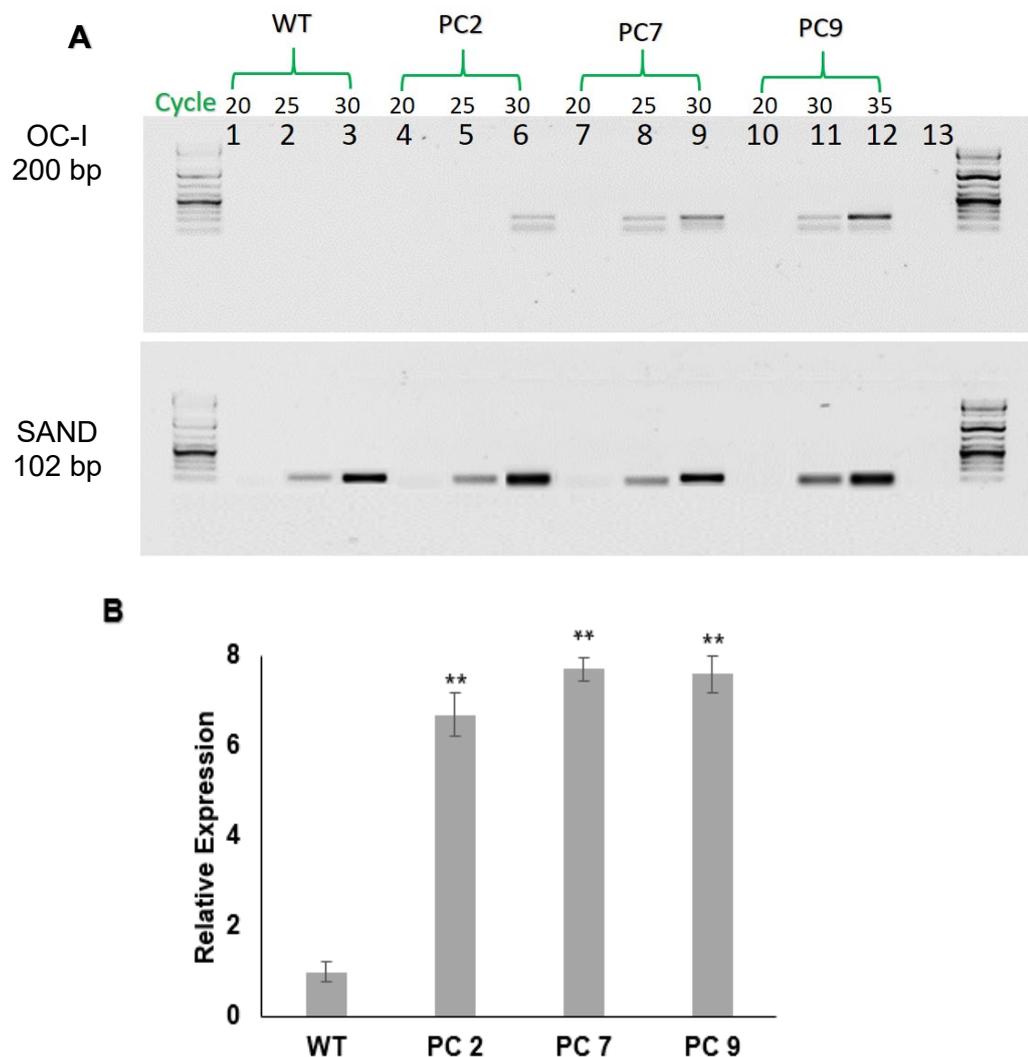


Figure 4.1: Expression analysis of OC-I in four-week-old *Arabidopsis* plants (PC lines and WT plants) compared with SAND as a housekeeping gene. **(A)** Semi-quantitative PCR products were analysed on a 1.2% agarose gel, with the expected PCR band of the OC-I gene (200 bp) and the expected PCR band of the SAND gene (102 bp) as controls. **(B)** Relative expression of the OC-I transgene in three transgenic *Arabidopsis* plants expressing OC-I in the chloroplast. WT was assigned a value of 1. The data was normalised to the *Arabidopsis* SAND gene. Means \pm SD for the plants in each line are indicated by bars. The asterisks indicate significant differences to WT plants (**p-value < 0.01, ANOVA).

4.2.2 Germination efficiency and plant development after germination

Seed germination was assessed five days after sowing. In these experiments 100 seeds of each of the WT and PC lines were germinated on agar plates containing half-strength Murashige and Skoog (MS) media. The germination of the PC lines showed a similar pattern to that of the WT, with no significant differences. However, the appearance of the radicle was slower in PC lines than the WT (Figure 4.3A). The number of viable seedlings after ten days was measured to assess the survival rates of the seedlings on the MS media. Seedling survival rates were similar in all lines (Figure 4.3B).

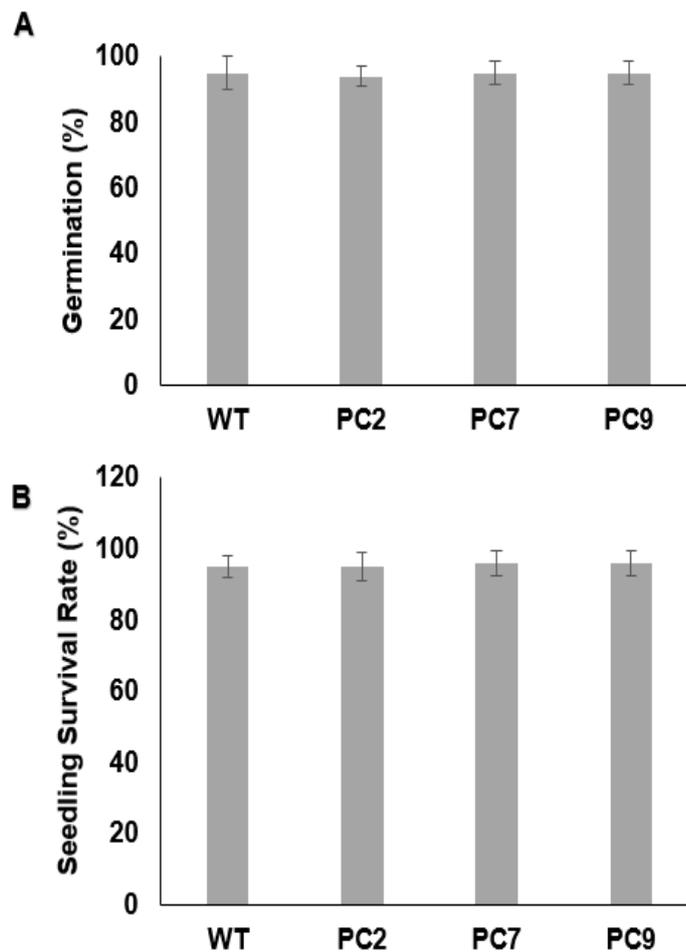


Figure 4.3: Comparison of (A) seed germination and (B) seedling survival rate in transgenic lines overexpressing OC-I (PC2, PC7 and PC9) and the wild type (WT). Bars represent means \pm SD (n=100 seeds).

4.2.3 Root architecture of transgenic line seedlings

Photographs of the PC and WT roots were analysed using ImageJ ten days after germination (Figure 4.4A). The length of the primary roots of the WT seedlings was approximately 2.5 cm at this stage, whereas the primary root length of PC seedlings was 1.5 cm in PC2 and 2.0 cm in PC7 and PC9 (Figure 4.4A). In addition, the PC lines had a much greater number of lateral roots than the WT (Figure 4.4C). Hence, lateral root densities were significantly higher in the PC lines than the WT (Figure 4.4D).

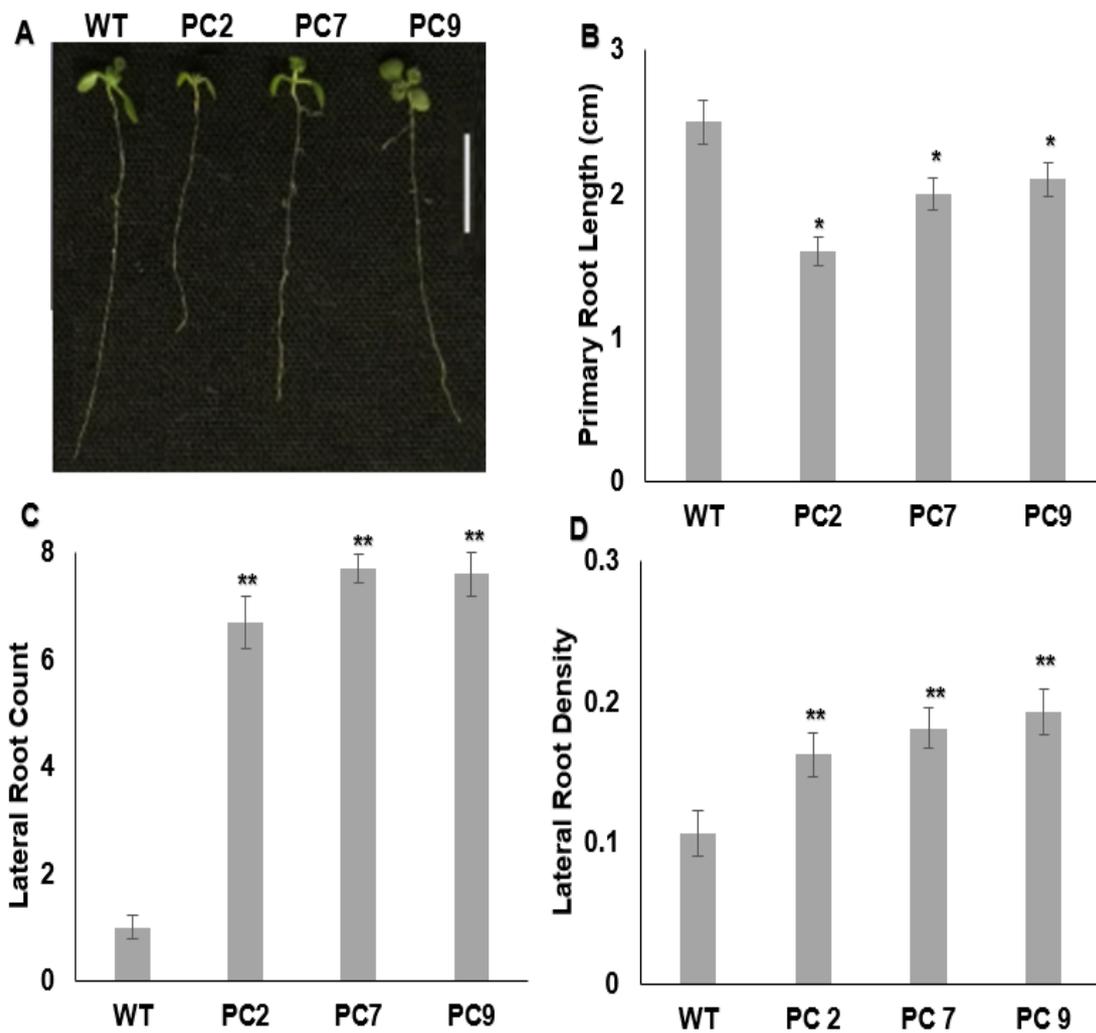


Figure 4.2: A comparison of root architecture in wild-type (WT) *Arabidopsis thaliana* and transgenic lines expressing OC-I in the chloroplast (PC2, PC7 and PC9) in 10-day-old seedlings: **(A)** photographs of 2-week-old seedlings grown on MS medium; **(B)** primary root length; **(C)** number of lateral roots; and **(D)** lateral root density. Mean values \pm SD (n=50) are shown. The asterisks indicate significant differences to WT plants (*p-value \leq 0.05 and p-value < 0.01, ANOVA). Scale bar: 3 mm.

4.2.4 Rosette morphology

Shoot phenotypes, flowering time and rosette biomass were measured throughout plant development. A range of shoot parameters were measured at weeks 4, 6, 8, 10 and 12 after germination.

4.2.4.1 Shoot phenotype, flowering, and biomass production

Representative images of the rosettes of the WT and PC lines at 4 and 6 weeks after sowing are shown in Figure 4.5A and B. The PC2 line and the WT had similar shoot phenotypes at week 4. The rosettes of two of the PC lines (PC7 and PC9) were visibly smaller than the other lines at this stage. In addition, the time to flowering was different in the PC lines and the WT (Figure 4.5B). Whereas 70% of the WT plants had flowered at week 5, none of the PC lines showed flowers at this point. Flowering was therefore delayed in PC lines compared to the WT (Figure 4.6A). Shoot biomass was significantly lower in the PC lines (Figure 4.6B) than the WT at 4, 6 and 8 weeks after sowing (Figure 4.6B). At subsequent development stages, however, shoot biomass was significantly increased in the PC lines relative to the WT. The number of leaves and the diameter of the rosettes were measured in all PC lines and the WT at 4, 6, 8, 10 and 12 weeks after sowing (Figure 4.6C and D). The PC lines tended to have fewer leaves and significantly smaller leaf diameters than the WT plants at all shoot development stages (Figure 4.6C and D).



Figure 4.3: A comparison of the representative phenotype of plants of the transgenic lines overexpressing OC-I in the chloroplast (PC2, PC7 and PC9) to that of WT. The phenotype of **(A)** four-week-old and **(B)** six-week-old plants showed delayed flowering in the PC lines grown on soil in a long-day photoperiod (LD with 16 h day/8 h night) at 20°C, 60% humidity, at 400 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity. Scale bar: 3 cm.

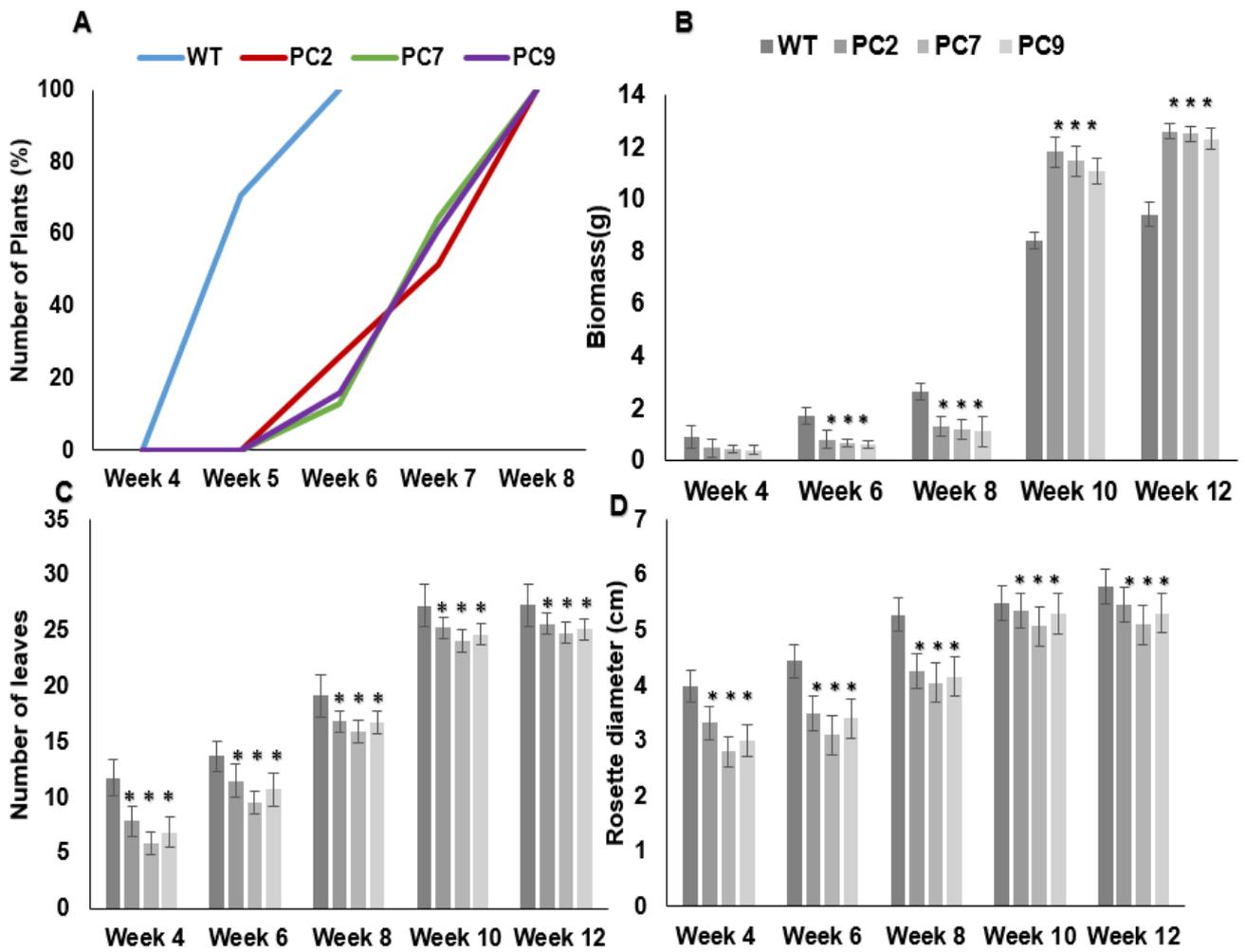


Figure 4.4: A comparison of shoot phenotype in the wild-type (WT) *Arabidopsis thaliana* and lines expressing OC-I in the chloroplast (PC2, PC7 and PC9) at 4, 6, 8, 10 and 12 weeks after sowing: **(A)** flowering time; **(B)** biomass; **(C)** number of leaves; **(D)** rosette diameter. Bars show the means \pm SD (n=24 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

4.2.4.2 Leaf pigments and protein contents

The amounts of chlorophyll, carotenoid pigments and protein were compared in the leaves of the WT and PC lines. The amounts of chlorophyll and carotenoid pigments were consistently greater in the rosette leaves of the PC lines at all stages of development than the WT (Figure 4.7A and B). There were no differences in leaf protein contents in the 4, 6 and 8-week-old PC plants compared to WT plants. However, the protein content at 10 weeks and at later stages of development was consistently higher in the leaves of the PC lines than the WT (Figure 4.7C).

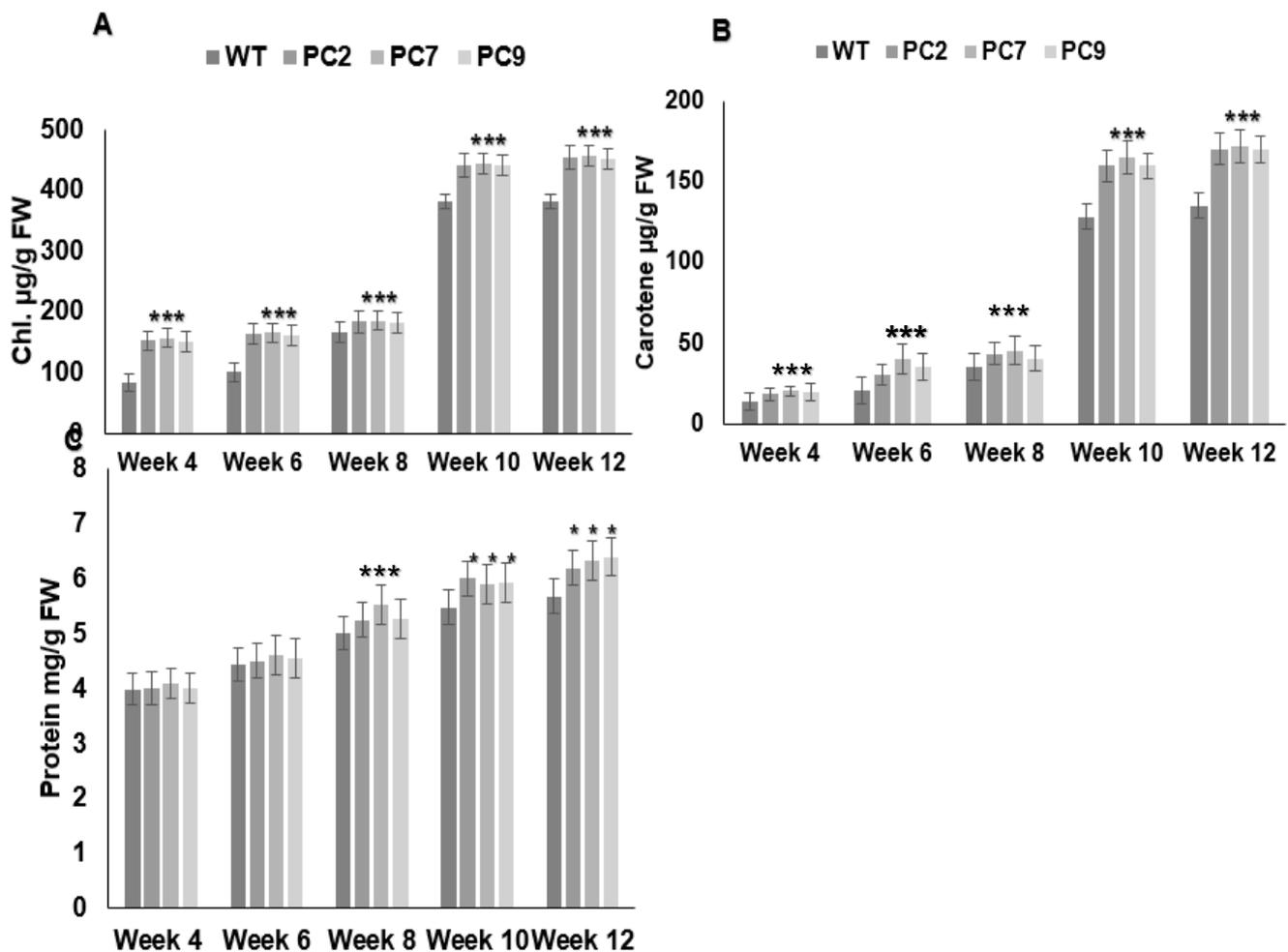


Figure 4.5: A comparison of the amount of (A) chlorophyll, (B) carotenoid pigments and (C) protein in the leaves of wild-type (WT) *Arabidopsis thaliana* and lines expressing OC-I in the chloroplast (PC2, PC7 and PC9). Bars represent means \pm SD (n=24 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

4.2.5 A role for OC-I in tolerance to high light stress

The effects of HL on photosynthesis rates, leaf protease activities and the abundance of photosynthetic proteins were determined in the WT and the PC lines. All plants were grown in compost in controlled environment chambers at low light intensity (LL: 250 $\mu\text{mol m}^{-2} \text{s}^{-1}$) for five weeks prior to exposure to HL stress (HL: 800 $\mu\text{mol m}^{-2} \text{s}^{-1}$) for 8 hours (see section 2.13; Material and Methods). Photosynthetic CO₂ assimilation was measured in fully expanded leaves immediately prior to exposure to HL. The rates of photosynthetic CO₂ assimilation were similar in all transgenic lines and WT under LL conditions. However, photosynthetic carbon assimilation rates were significantly decreased in the WT under HL conditions. The PC lines had significantly higher rates of photosynthetic CO₂ assimilation than the WT under these conditions (Figure 4.8).

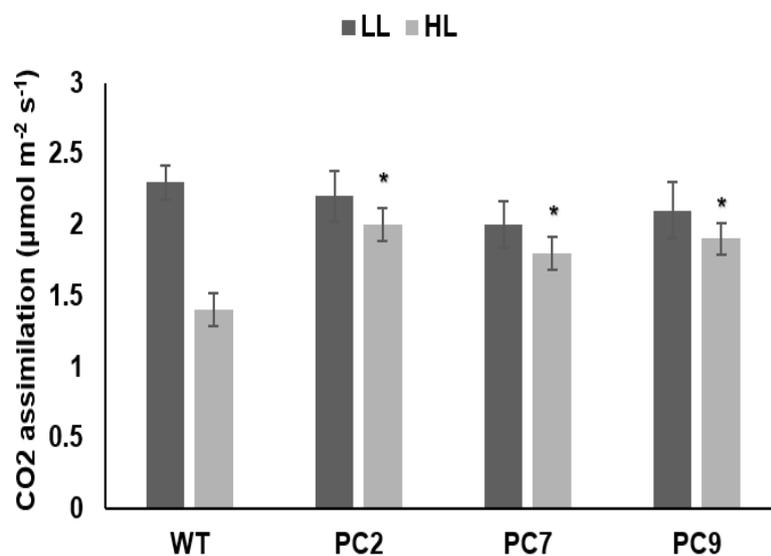


Figure 4.6: The effects of the expression of OC-I on photosynthetic CO₂ assimilation rates in the leaves of PC lines and WT *Arabidopsis* plants grown under low light (LL) and high light (HL) conditions. Photosynthetic CO₂ assimilation in all leaves of the rosette was measured for PC lines and WT plants that had been grown under low light (250 $\mu\text{mol m}^{-2} \text{s}^{-1}$) and then were transferred to a high-light environment (800 $\mu\text{mol m}^{-2} \text{s}^{-1}$) for 24 hours. Bars represent means \pm SD (n=3 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

Total protease activities were measured in 5-week-old PC and WT plants under LL and HL conditions using Abcam's Protease Activity Assay Kit (see Section 2.16: Materials and Methods). Total leaf protease activity was similar in both PC lines and WT under LL conditions. The total protease activities of the leaves increased when the plants were exposed to HL (Figure 4.9A). However, the HL-induced increase in total protease activity was much less marked in the PC lines than the WT (Figure 4.9A). Leaf cysteine protease activities were also measured in the PC lines and the WT. No differences in leaf cysteine protease activity were observed between PC lines and WT under LL conditions (Figure 4.9B). However, the leaf cysteine protease activity were significantly increased in the WT under HL conditions. In contrast, leaf cysteine protease activities decreased in the PC lines than in the WT under HL conditions (Figure 4.9B).

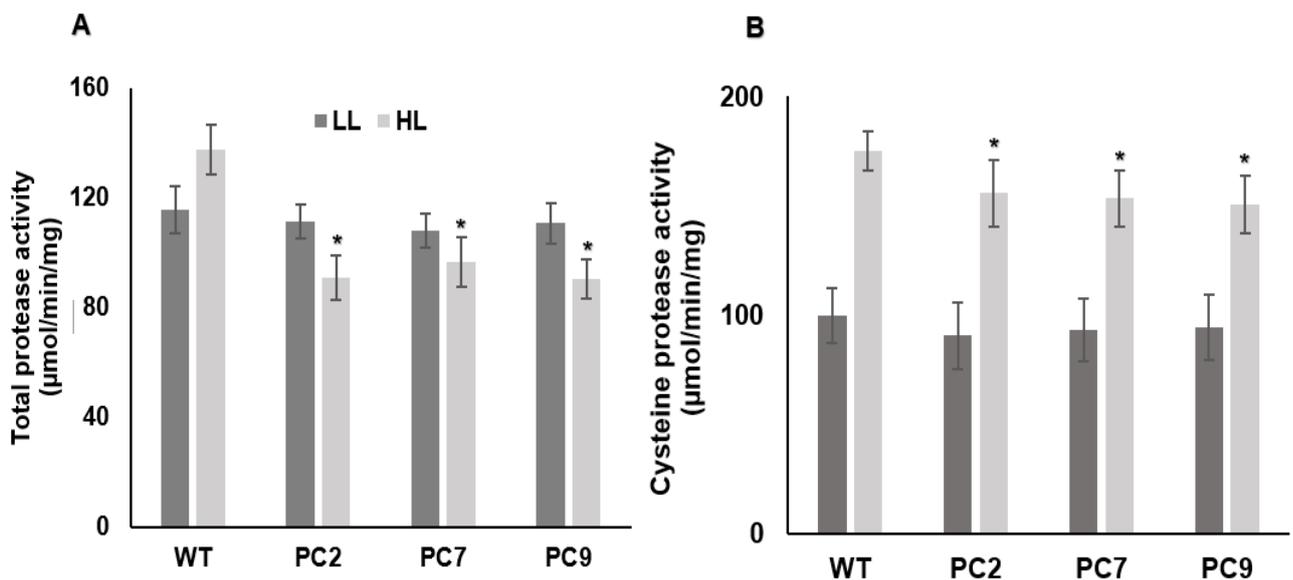


Figure 4.7: (A) Total protease activities and (B) cysteine protease activity in the leaves of the PC lines and WT *Arabidopsis* plants grown under low light (LL) and high light (HL) conditions. Bars represent means \pm SD (n=3). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

Western blot analysis using specific antibodies was used to determine the accumulation of the Rubisco large subunit, PSII reaction centre D1 protein and the phosphorylated form of the D1 protein in the leaves of the PC lines and WT plants under LL and HL conditions. There were no differences in the abundance of the Rubisco large subunit protein, the D1 protein and the phosphorylated form of the D1 protein between the lines under LL conditions (Figure 4.10A). However, the amount of these proteins increased markedly in the leaves of the PC lines relative to those of the WT after six hours of exposure to HL (Figure 4.10B).

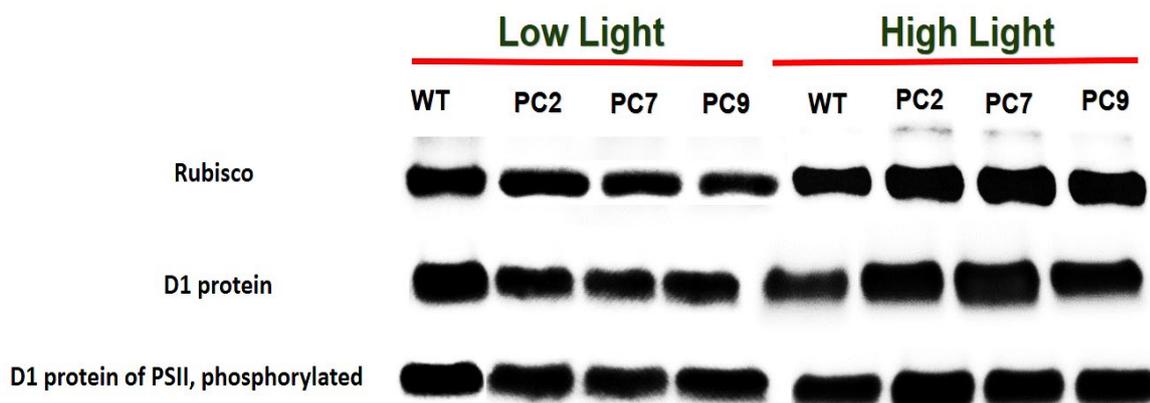


Figure 4.10: Western-blot analysis of Rubisco large subunit, PSII reaction centre D1 protein and the phosphorylated form of the D1 protein extracted from leaves of 5-week-old plants of the transgenic PC lines and the WT grown under LL and then transferred to HL for 8 hours. Protein was extracted from leaves (10 µg) using PAGE and then transferred to nitrocellulose membrane and analysed with primary antibodies RbcL | Rubisco large subunit, anti-PsbA and anti-phosphorylated PsbA respectively.

The expression of nuclear-encoded photosynthesis genes and of chloroplast-encoded photosynthesis genes was compared in the PC lines and the WT. Quantitative real-time PCR was performed on leaf samples were harvested from plants were grown under LL and after six hours of exposure to HL (Figure 4.11A and B). The levels of transcripts encoding the light-harvesting chlorophyll a-b binding protein (*LHCA*), the light-harvesting chlorophyll a-b binding protein 1 (*LHCB1*), the light-harvesting chlorophyll a-b binding protein 2 (*LHCB2*), the small subunit of ribulose biphosphate carboxylase (*rbcS*), the photosystem II D1 protein (*psbA*) and the photosystem II D2 protein (*psbD*) were significantly higher in the leaves of the PC lines than the WT under LL conditions (Figure 4.11A). The levels of all transcripts were lower under HL than LL (Figure 4.11A and 4.11B). While the levels of *LHCA*, *LHCB1*, *rbcS*, *psdA* and *psbD* transcripts were significantly higher in the PC lines than the WT than under HL conditions (Figure 4.11B), the levels of *LHCB2* transcripts were similar in all lines (Figure 4.11B).

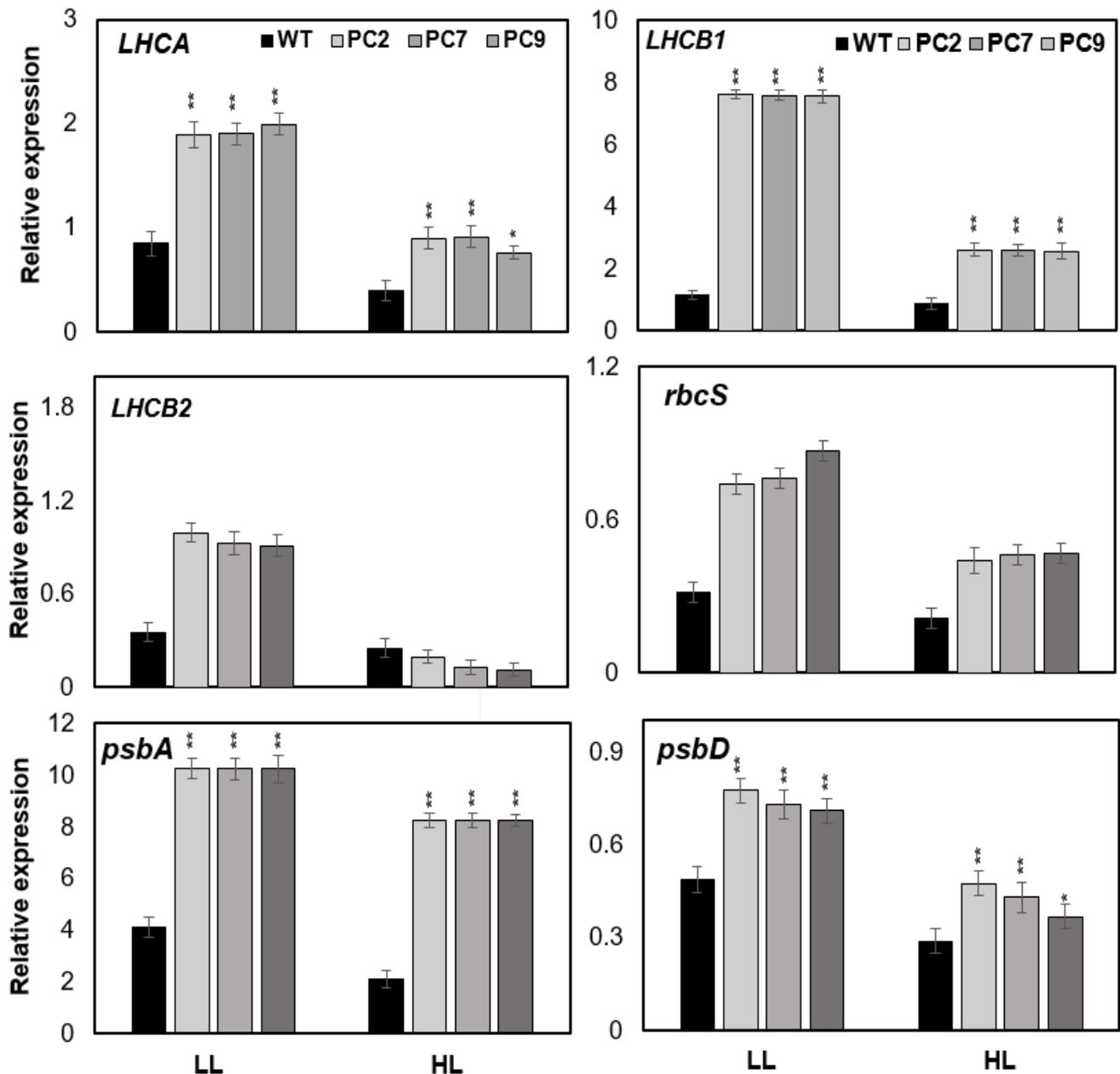


Figure 4.8: The effect of OC-I expression on the transcript levels of nuclear-encoded chloroplast and chloroplast-encoded photosynthesis genes treated under LL and HL conditions in 5-week-old PC and WT *Arabidopsis* plants. *LHCA*: light-harvesting chlorophyll a binding protein; *LHCB1*: light-harvesting chlorophyll a binding B1; *LHCB2*: chlorophyll a-b binding protein 2; ribulose biphosphate; *rbcS*: carboxylase small chain; *psbA*: photosystem II D1 protein; *psbD*: photosystem II D2 protein. The data were normalized to the actin gene. Bars represent means \pm SD (n=3 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, and **p-value < 0.01, ANOVA).

To analyse the effects of OC-I expression on chloroplast-to-nucleus retrograde signalling pathways that regulate photosynthetic gene expression, the PC lines and WT were grown for seven days on agar plates containing ½ MS media with ethanol as a control (CONT) or media containing either Lincomycin (LINCO; 500 µM), or Norflurazon (NF; 5 µM) to inhibit photosynthesis (see Section 2.14: Materials and Methods). The PC and WT seedlings showed were pale and lacked chlorophyll in the presence of LINCO and NF compared to seedlings grown in the absence of these inhibitors (Figure 4.12A). In contrast to the control seedlings, none of the lines had fully developed hypocotyls and cotyledons after seven days in the presence of inhibitors (Figure 4.12A).

The levels of *LHCA*, *LHCB1* and *LHCB2* transcripts were measured in the seedlings grown in the absence and presence of these inhibitors. The abundance of *LHCA* transcripts was decreased in PC lines compared to the WT in the absence of inhibitors (Figure 4.12B). However, the abundance of *LHCA* transcripts was significantly increased in PC lines in the presence of both inhibitors compared to the WT (Figure 4.12B). The expression of *LHCB1* and *LHCB2* was lower in the PC lines than the WT in the absence of inhibitors (Figure 4.12B-C). In contrast, the levels of *LHCB1* transcripts were lower in the PC lines compared to the WT in the presence of inhibitors (Figure 4.12B). Similarly, the abundance of *LHCB2* transcripts was significantly lower in the PC lines than the WT in the presence of inhibitors (Figure 4.12D). These results show that the presence of OC-I in the chloroplasts increased the expression of the *LHCA* gene in the presence of LINCO and NF, whereas the expression of *LHCBs* genes was repressed in these conditions.

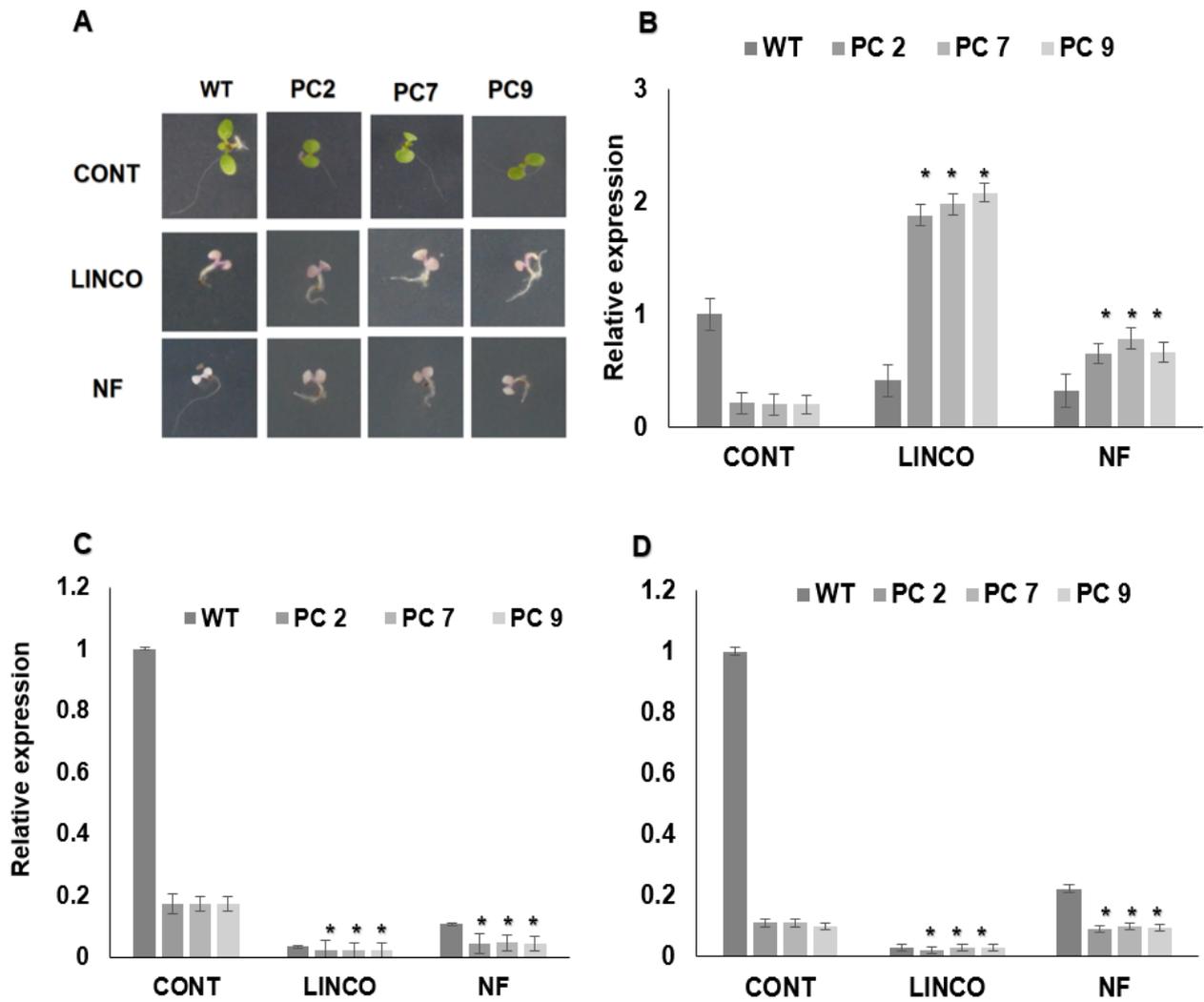


Figure 4.9: *Arabidopsis thaliana* lines expressing OC-I in the chloroplast (PC2, PC7 and PC9) and the WT were treated with lincomycin (LINCO), a chloroplast protein synthesis inhibitor, or norflurazon (NF; 5 μ M), an inhibitor that inhibits carotenoid synthesis. **(A)** Comparison of the representative phenotype of seedlings of PC2, PC7 and PC9 and WT grown on MS-agar media containing either LINCO (500 μ M), or NF (5 μ M) for seven days. The effect of LINC and NF on transcript levels of **(B)** *LHCA*, **(C)** *LHCB1* and **(D)** *LHCB2* in PC lines and the WT is also shown. Bars represent means \pm SD (n=24 plants). The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

4.3 Discussion

Cysteine protease inhibitors are involved in the control of protein turnover but their precise functions remain poorly characterized. Previous reports on the effects of constitutive expression of OC-I in transgenic plants have shown that plant growth and stress tolerance traits are changed in the presence of the cysteine protease inhibitor. The results presented in Chapter 3 showed that the expression of OC-I in the cytosol of transgenic *Arabidopsis* lines has a marked effect on plant growth and development. Moreover, OC-I expression exerted a strong influence over photosynthetic gene expression and leaf responses to HL. The results presented in this chapter concern the phenotypes and gene expression of plants where OC-I expression is targeted to the chloroplasts using an appropriate peptide targeting signal. These studies were performed on T4 generation transgenic *A. thaliana* lines, which express OC-I protein in the chloroplasts (PC; see Section 2.1.1: Materials and Methods for details).

Cysteine proteases are likely to fulfil important roles at every stage of plant development (Abe, K et al., 1987). OC-I expression in the chloroplasts led to a slower germination rate in the PC lines than the WT and the PC plants were smaller and accumulated less biomass throughout vegetative development. The phytocystatins called AtCYS6 (Hwang et al., 2009) and BrCYS1 (Hong et al., 2007) regulate seed germination. In addition, the germination-specific cysteine protease 1 (GCP1; At4g36880) is involved in the initial phase of germination in *A. thaliana* (Tsuji et al., 2013). OC-I expression may affect germination rate as a result of inhibition of these cysteine proteases, as suggested by earlier studies in the Foyer lab (Roa-Roberts, 2014). The OC-I-dependent inhibition of germination demonstrated here suggests that chloroplast-localized cysteine proteases are involved in the mobilisation of seed reserves. This intriguing observation is hard to explain at a mechanistic level. Plastid

starch degradation is important in driving germination and hence cysteine proteases that target starch mobilization enzymes may have a role in this process (Allorent et al., 2013).

OC-I expression in the roots of the PC seedling plastids decreased the length of the primary root relative to the WT controls. However, the PC lines had more lateral roots and greater lateral root density than the WT. This finding suggests that OC-I expression in the root plastids alters the metabolism or signalling of hormones such as auxin that regulate root architecture. The findings reported here are consistent with earlier studies in the Foyer lab showing that OC-I protein was expressed in root plastids altered root growth and development (Roa-Roberts, 2014). The senescence-associated gene 12 (SAG12) is a cysteine protease, which plays important roles in the low nitrogen (N)-induced remobilization of resources during leaf senescence in *Brassica napus* L. (Desclos et al., 2008) and *A. thaliana* (Poret et al., 2016). SAG12 is present in *A. thaliana* roots and its expression was observed when plants were cultivated under either high nitrogen or low nitrogen conditions (James et al., 2019). In addition, the process of primary root elongation involves the CEP2 protein. Primary roots were shorter in the absence of CEP2 because of a reduction in the length of trichoblasts and other epidermal cells (Höwing et al., 2018). Precisely how OC-I expression in plastids is able to inhibit the activities of these CPs is unknown.

OC-I expression in the chloroplasts delayed the vegetative development of the rosettes. The PC lines had a slower phenotype growth throughout vegetative development. However, after flowering had started all the PC lines accumulated much greater amounts of shoot biomass than the WT, particularly at the later stages of shoot development. The expression of OC-I was slightly higher in the PC7 and PC9 lines than PC2. This might explain the differences in shoot growth between the PC lines.

Flowering was delayed in all PC transgenic lines compared to the WT suggesting that reproductive development was delayed in the OC-I expressing lines because vegetative growth was slower. Delayed leaf senescence have been associated with a decrease in cysteine protease activity (Benchabane, M. et al., 2010). Moreover, OC-I expression in tobacco leaves was shown to delay flowering as well as leaf senescence (Van der Vyver et al., 2003). The small ubiquitin-like modifier (SUMO) proteases called SPF1 and SPF2 regulate fertility. SPF1 and SPF2 are cysteine proteases that regulate flowering time (Rawlings et al., 2006; Morrell and Sadanandom, 2019). Taken together, these findings suggest that OC-I expression in plastids may inhibit the activity of these proteases, resulting in the delayed rosette growth and flowering in the PC lines compared to the WT.

Although vegetative development was slower in the transgenic lines than the WT, the OC-I-dependent inhibition of CPs eventually enabled the transgenic plants to accumulate more biomass than the WT, as has been shown in previous studies on tobacco (Prins et al., 2008). The trend towards higher biomass accumulation was particularly evident in plants grown for 10 and 12 weeks. At this stage, the transgenic plants were visibly larger and had significantly more biomass than the WT. The cysteine protease responsive-to-dehydration-21 (RD21) is found in roots, leaves and flowers (Liu, Y. et al., 2020). It may be that OC-I expression slows down the activity of RD21 in order to regulate growth. Taken together, the findings support the conclusion that the expression of OC-I in chloroplasts and other types of plastid has a marked effect on plant growth and development.

The PC leaves in which OC-I was targeted to the chloroplasts showed a greater accumulation of leaf chlorophyll and carotenoids than the WT. These interesting findings are interesting, particularly because OC-I expression in the cytosol does not

greatly alter leaf pigment contents, as has been shown for example in soybean (Quain et al., 2014). The levels of leaf protein were significantly higher in the PC plants than in the WT at weeks 6, 8, 10 and 12, suggesting that OC-I expression in chloroplasts delays the turnover of chloroplast proteins such as Rubisco. OC-I expression in the cytosol has previously been shown to lead to an accumulation of chloroplast proteins (Prins et al., 2008; Quain et al., 2015). The senescence induced decreases in leaf protein and chlorophyll is related to the increased activities of cysteine and serine proteases (Poret et al., 2017). Taken together, the findings presented in this chapter show that the expression of OC-I in chloroplasts leads to an increased accumulation of leaf chlorophyll, carotenoids and proteins, indicating that OC-I has a marked effect on the turnover of the components in the leaves.

Proteases are found in the chloroplast stroma, the thylakoid lumen, the thylakoid membranes and the chloroplast envelope (Adam et al., 2006; Kato and Sakamoto, 2010). Exposure to stress increases the activities of the major chloroplast proteases, promoting the degradation of chloroplast proteins outside of the plastid (Mamaeva et al., 2020). The effect of OC-I expression in the chloroplasts on the abundance of chloroplast proteins, particularly Rubisco was examined in the PC lines under LL and HL. The levels of the Rubisco large subunit protein were changed in response to exposure to HL in the WT and PC lines. However, the HL-dependent decreases in the abundance of chloroplast proteins observed in the WT in response HL stress were absent from the PC lines, which accumulated chloroplast proteins under these conditions. The HL response in the WT is the result of increased chloroplast protein degradation as less light-harvesting and other photosynthetic proteins are required in plants exposed to HL. However, the PC lines had higher rates of photosynthetic CO₂ assimilation than the WT under HL conditions, indicating that OC-I expression

prevents the reduction in photosynthetic capacity that occurs in the WT in response to HL stress. This finding confirms those of previous studies showing that OC-I protect photosynthesis from stress induced decreases in capacity (Van der Vyver et al., 2003; Prins et al., 2008). Cysteine proteases are activated by numerous abiotic stresses (Belenghi et al., 2003; Quain et al., 2014). The HL-induced increases in cysteine protease activity observed in the WT under HL conditions were absent from the PC lines. This finding is in agreement with the higher levels of accumulation of the Rubisco and D1 proteins that was observed in the PC lines under HL compared to the WT. The higher accumulation of chloroplast proteins was correlated with higher *rbcS* and *psbA* transcripts in the PC lines than the WT following exposure to HL stress. To date, the only proteases that have been reported to contribute to the degradation of Rubisco inside chloroplasts are metallo- and aspartic proteases (Kato et al., 2004; Roberts et al., 2012). However, because stromule and vesicle formation, and autophagy pathways are considered to play a key role in chloroplast protein degradation, there is the intriguing possibility that OC-I expression exerts its effects in chloroplasts through interaction with the cellular vesicle trafficking system. Eight putative chloroplast-localized homologs of known protein components of the COPII cytosolic vesicle transport system have been identified in *Arabidopsis* (Khan et al., 2013). The data presented in this chapter suggests that OC-I may prevent the degradation of the Rubisco and D1 proteins under HL stress, leading to improved photosynthetic capacity. Precisely how OC-I might influence chloroplast gene expression to increase the abundance of *rbcS* and *psbA* is unknown.

Plant responses to HL and other stresses are controlled by retrograde signalling pathways (Wagner et al., 2004; Rossel et al., 2007; Xiao et al., 2012). The findings presented in this chapter show that OC-I expression affects chloroplast-to-nucleus

retrograde signalling pathways under both of the light levels used in these studies. LINCO and NF inhibitors are often used in experiments that are designed to explore the chloroplast-to-nucleus signalling pathways that are required to sustain functional chloroplasts through changes in nuclear gene expression (Terry and Smith, 2013). Part of the studies reported in this chapter were designed to investigate whether OC-I protein expressed in chloroplasts affects these retrograde signalling pathways. The levels of *LHCA*, *LHCB1* and *LHCB2* transcripts were greatly decreased in the WT plants in the presence of these inhibitors, as observed in previous studies (Karpinska et al., 2017). The presence of OC-I in the chloroplasts decreased the levels of *LHCB* transcripts but enhanced the abundance of *LHCA* transcripts in the presence of LINCO and NF. A lower level of inhibition of *LHCB* expression was observed in the genomes uncoupled (*gun 1*) mutants after LINCO and NF treatments (Inaba, 2010). Gun proteins are involved in retrograde signalling pathways (Wu et al., 2019). The *gun* mutants accumulate lower levels of singlet oxygen than the WT after LINCO and NF treatments, and so have lower levels of oxidative stress. This may also be the case when OC-I is expressed in the chloroplasts. In the majority of *gun* mutants, there are no significant changes in leaf chlorophyll contents or the expression of nuclear photosynthesis-related genes when the mutants are grown in the absence of inhibitors (Susek et al., 1993; Mochizuki et al., 2001). The *gun* phenotype only becomes apparent when treated the mutants are treated with norflurazon or other chloroplast inhibitors. Taken together, these studies suggest that the expression of OC-I in chloroplasts regulates *LHCB* expression as it does in the *gun 1* mutants. However further studies and additional experimental data are required to determine how OC-I expression influences retrograde signalling.

Taken together, the findings reported in this chapter show that chloroplast targeted OC-I-expression has significant effects on plant growth and development. These results are difficult to explain because OC-I inhibits papain-like cysteine proteases, which are not generally localized in plastids. However, the targeting of OC-I to chloroplasts, suggest that papain-like cysteine proteases play a significant role in chloroplast processes that exert a wide range of effects on shoot development. The data presented here indicates that chloroplast/plastid processes are particularly important in this regulation of plant growth and development. In comparison to the data presented in Chapter 3, the data presented here demonstrate that OC-I effects on plant processes differ depending on whether the protease inhibitor is expressed in the chloroplasts or in the cytosol. In the next chapter, the expression of OC-I in one of the most consumed cereals in the modern world, hexaploid wheat (*Triticum aestivum* L.), is described.

Chapter 5 . Selection of T4 generation transformed wheat plants over-expressing the rice cystatin, *oryzacystatin-I* (OC-I)

5.1 Introduction

Considered to be one of the ‘big three’ global grains along with maize (*Zea mays*) and rice (*Oryza sativa*) (Hinchliffe and Harwood, 2019), hexaploid wheat (*Triticum aestivum* L.) is amongst the most widely consumed cereals in the modern world. Of all staple food crops, wheat is the most the extensively grown (FAO, 2017), providing as much as 20% of human dietary protein and calories worldwide (Alaux et al., 2018). The UK and mainland Europe provide excellent environmental conditions for the cultivation of wheat (Reynolds et al., 2012) and, consequently, the European Union produced around 151.6 million metric tonnes of wheat in 2019/2020 (Figure 5.1).

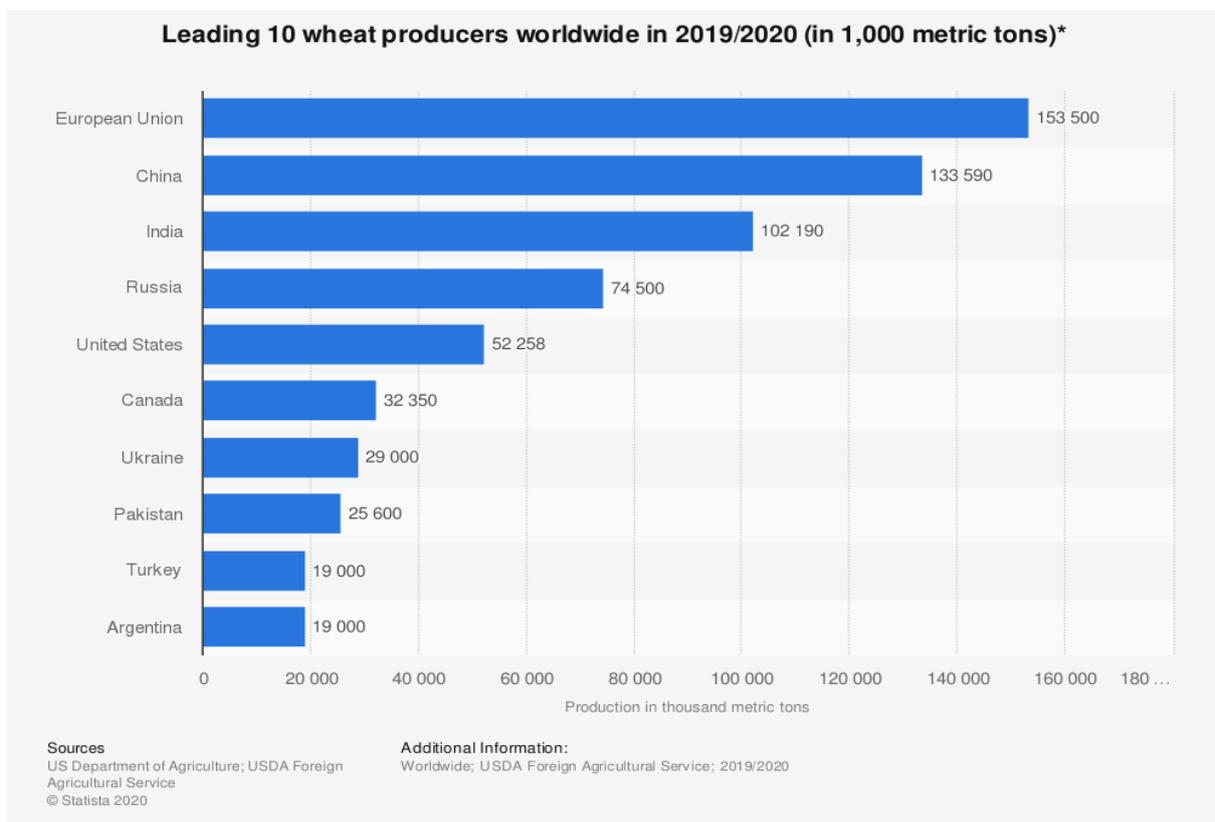


Figure 5.1: The top ten producers of wheat globally (2019.2020). Measurement is in metric tonnes. Source: <https://www.statista.com/statistics/237912/global-top-wheat-producing-countries/>.

However, the wide variety of biotic and abiotic stresses from which the crop suffers leads to extensive economic losses. Innovative technologies are therefore required to reduce the risks posed by these stresses and increase crop yields (Kashyap et al., 2020). There is therefore a direct link between the ability of wheat to survive these stresses to provide food security.

Genetic engineering has been proposed as a way to improve the characteristics of wheat that are related to its economic efficiency (Tester and Langridge, 2010; He et al., 2011). It is anticipated that in the future transgenic crop varieties will provide an increasing amount of high-quality agricultural produce, as it has already been shown that crops expressing certain transgenes suffer less damage caused by biotic and abiotic stresses (Chen, H. and Lin, 2013; Parisi et al., 2016; Briefs, 2017). Several studies have found that the transgenic suppression or overexpression of certain endogenous genes can enhance the agronomic traits of wheat (Fu et al., 2007; Gil-Humanes et al., 2010; Altenbach et al., 2014; Chen, D. et al., 2018; Mega et al., 2019).

However, to date there have been no studies evaluating wheat lines overexpressing cysteine protease inhibitors. The research hypothesis is based on the notion that the expression of OC-I, an inhibitor of papain-like cysteine proteinases might improve wheat yields and grain properties. In the following study, transgenic wheat plants that overexpress OC-I were produced and characterised to determine whether this transgene could improve wheat yields and seed properties.

For the experiments reported in this chapter, transgenic wheat lines overexpressing OC-I were first produced in the National Institute of Agricultural Botany (NIAB) and T1 generation transformed plants were supplied for further analysis. The following actions are described in this chapter:

1. The selection of T2, T3 and T4 transgenic plants using kanamycin.
2. Confirmation of the presence of the OC-I gene in both T2, T3 and T4 generation plants using PCR.
3. Analysis of OC-I transcript levels using qRT-PCR.

5.2 Results

The OC-I construct that was prepared in the Foyer lab was sent to National Institute of Agricultural Botany (NIAB), where it was used to transform wheat. T1 selection plants were selected as in the procedures described in section 2.1.2.: T1 generation seeds were provided by NIAB (Table 2.1: Materials and Methods). Of these, 37 regenerated wheat plants were confirmed to be transformed by PCR amplification of the OC-I coding sequence (Table 2.1: Materials and Methods).

5.2.1 The overall procedure for generation selection

The results presented in this section describe the procedures used to select transformed plants of the T2, T3 and T4 generations on media containing kanamycin, the confirmation of the presence of the OC-I transgene using PCR and finally the level of OC-I expression using RT-PCR. In these studies, T4 transgenic wheat seeds were selected over a period of two years (Figure 5.2) and more than 600 seeds in total were sown on media containing kanamycin. The kanamycin resistant transgenic seedlings were transplanted into soil and grown under glass house conditions with a long-day photoperiod (16-hour day/8-hour night) at an irradiance of $1000 \mu\text{mol m}^{-2}\cdot\text{s}^{-1}$ with 30/25°C day/night temperatures and 60% humidity (Figure 5.3). The production of transgenic wheat plants showing three main stages: growing, flowering and harvesting. The principal stages are germination (Figure 5.3 A) and seedling growth (Figure 5.3 B), tillering (Figure 5.3 C), stem elongation, booting and heading (Figure 5.3 D), milk and dough development (Figure 5.3 E) and ripening (Figure 5.3 F). Each growth cycle took about five months from germination to harvesting the seeds in the Plant Growth Facilities at the University of Leeds (Figure 5.3 G).

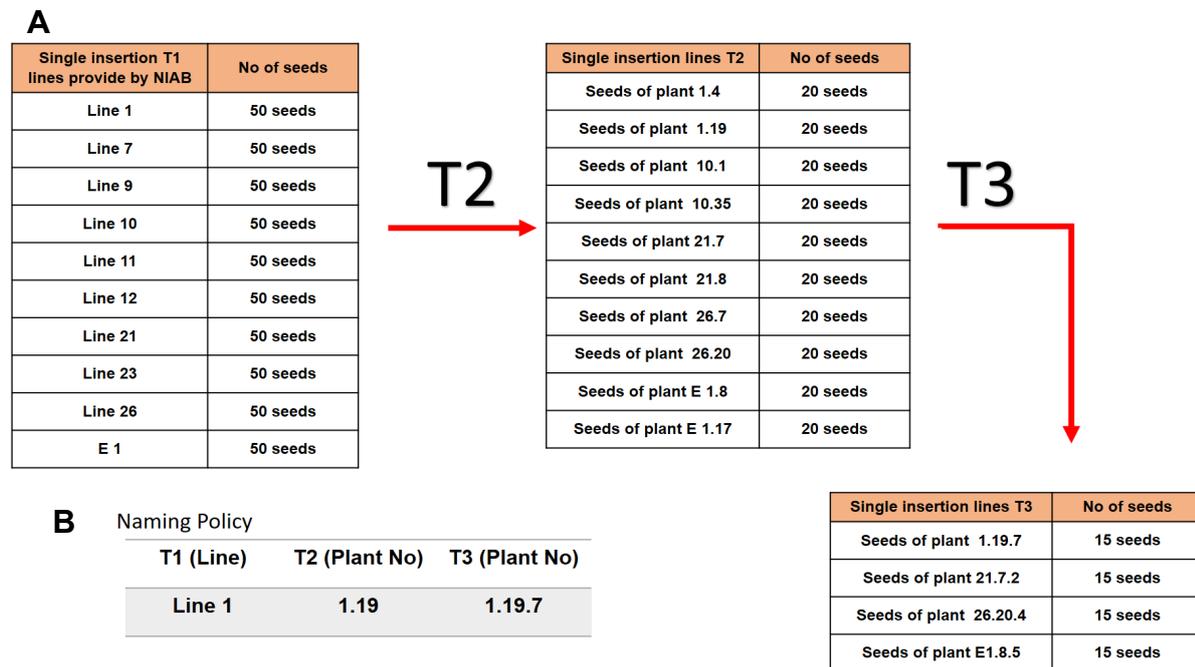


Figure 5.2: A schematic diagram of the overall procedure for the selection of T4 generation, showing selected lines in each generation. A) 10 single insertion (T1) lines provided by the NIAB were chosen to select T2 generation. 50 seeds per line (transformed seed) and E 1 (control) grown on media containing 10mg/ml kanamycin. Seedlings resistant to Kan were transplanted into soil to obtain T2 generation seeds. Seeds from plants 1.4, 1.19, 10.1, 10.35, 21.7, 21.8, 26.7, 20.20, E1.8 and E1.17 were chosen based on their PCRs and copy number to obtain T3 seeds. Seeds from plants 1.19.7, 21.7.2, 26.20.4 and E1.8.5 were chosen to obtain T4 seeds. B) An example showing the naming policy of the selected transformed wheat plants through T1, T2 and T3 generation.



Figure 5.3: Overview of the production of transgenic wheat plants showing three main stages: Growing, flowering and harvesting. The principal stages are germination and seedling growth (A and B), tillering (C), stem elongation, booting and heading, (D) milk and dough development (E), and Ripening (F). Plants were grown in the Plant Growth Facilities at the University of Leeds under glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of $400 \mu\text{Em}^{-2}\text{s}^{-1}$. (G). Each growth cycle, from germination to harvesting the seeds, took about four to five months.

5.2.2 Selection of kanamycin resistant T2 seeds from T1 seeds

Seeds of 10 single insertion independently transformed lines (T1; Table 2.1) were selected for analysis in the T2 generation (Figure 5.2). A total of 50 seeds per line as shown in Figure 5.2 was sown onto $\frac{1}{2}$ MS agar media containing 10mg/ml kanamycin (Kan). The construct contained the kanamycin resistance gene Neomycin phosphotransferase II (nptII), which was used to identify the transformed plants (see Section 2.6: Materials and Methods). Three percent of the germinated T2 seedlings were found to be resistant to Kan based on the survival phenotypes i.e. the appearance of the coleoptile and first leaf, and long roots that penetrated into the Kan selection media (Figure 5.4). These seedlings were then transplanted into soil to obtain T2 generation seeds.

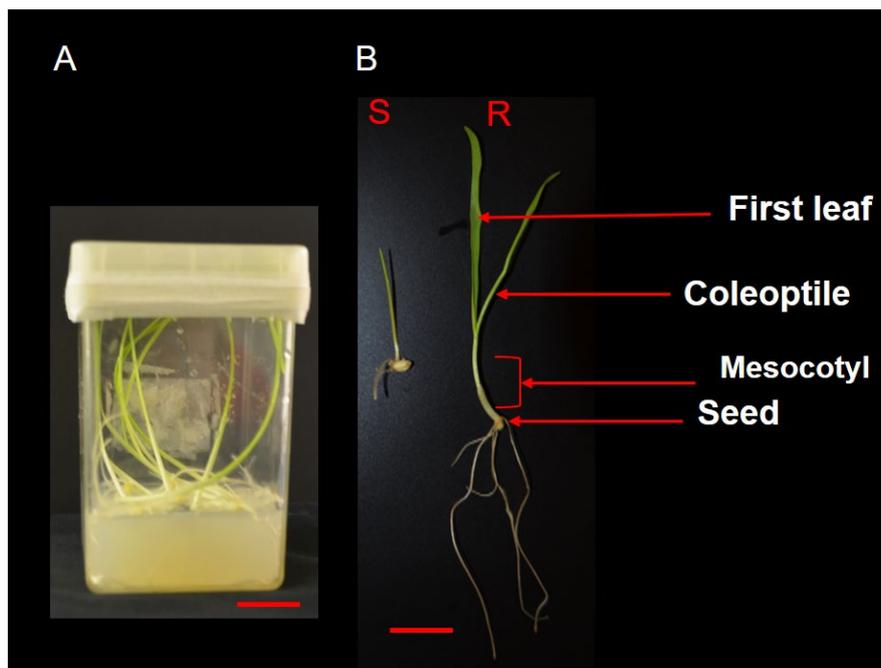


Figure 5.4: An example showing kanamycin selection of transformed wheat seeds. A) Magenta GA-7 Vessel containing kanamycin media. B) The phenotypes of sensitive (S) and resistant (R) seeds after 7 days of treatment with kanamycin.

PCR analysis was performed to detect the presence of OC-I transgene using primers spanning OC-I (forward primer specific to actin promoter and reverse primer specific to Nos terminal; Table 2.2; Materials and Methods). DNA were extracted at the three leaf stage. A 600-bp band was present in the samples from the transgenic plants (Figure 5.5). In addition, T-DNA copy number was determined in transformed lines by TaqMan assay (see section 2.18; Materials and Methods). 8 lines out of 19 transgenic lines were showed to be single insertion (Table 5.1). The T2 seeds were harvested and eight transgenic (1.4, 1.19, 10.1, 10.35, 21.7, 21.8, 26.7, and 20.20) with wild-type (E1.8 and E1.17) plants were chosen for subsequent generation depended on their PCRs and copy number as shown in Table 5.2.

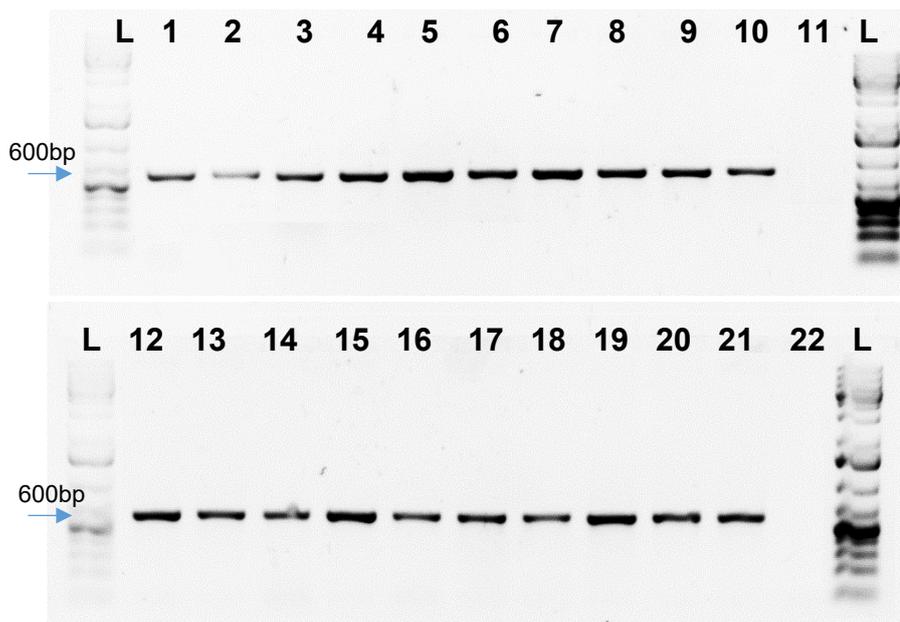


Figure 5.5: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3-week-old T1 wheat plants. DNA were extracted at the three leaf stage. Plants were grown under a under glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Lanes: L shows DNA marker, lanes 1-22 show DNA extracted from leaves of lines; L.1.4, L.1.19, L.7.5, L.7.40, L.9.11, L.9.2, L.10.1, L.10.35, L.11.44, E1.8, L.11.23, L.12.38, L.12.2, L.21.7, L.21.8, L.23.47, L.23.50, L.26.5, L.26.7, L.26.20, E1.17, respectively.

Table 5-1: T-DNA copy number of transformed T1 plants of transgenic wheat expressing OC-I and WT. Copy number estimation by qPCR of the transgene are shown. Blue labels indicate single insertion lines that are selected for subsequent generation work.

Line	Estimated Copy Number By Real-Time PCR	Line	Estimated Copy Number By Real-Time PCR
E1.8	0	L.11.44	2
E1.17	0	L.11.23	2
E1.25	0	L.12.38	2
L.1.4	1	L.12.2	2
L.1.19	1	L.21.7	1
L.7.5	2	L.21.8	1
L.7.40	2	L.23.47	2
L.9.11	2	L.23.50	2
L.9.2	2	L.26.5	2
L.10.1	1	L.26.7	1
L.10.35	1	L.26.20	1

5.2.3 Selection of T3 seeds and confirmation of the presence of the OC-I transgene in T2 plants

Eight transgenic (1.4, 1.19, 10.1, 10.35, 21.7, 21.8, 26.7, and 20.20) with wild-type (E1.8 and E1.17) plants were chosen to select T3 generation, as shown in Figure 5.2. Twenty seeds per line were sown onto media containing kanamycin. About 75% of the resultant seedlings were resistant to kanamycin, which may indicate the presence of OC-I transgene. The remaining 25% did not grow (Table 5.2). Seedlings from 21.8 and 26.7 lines showed 100% resistance to the antibiotic. Seeds with a segregation ratio of 3 (resistant) to 1 (sensitive) were selected to obtain T3 seeds that were potentially homozygous. An additional step to determine whether the experimental segregation ratio results observed from the selection on kanamycin matches the expected results. In order to evaluate this, a chi-square test (goodness-of-fit) was performed. The chi-square results fitted a 3 resistant: 1 sensitive segregation ratio in which the segregation of transgene was identified in a Mendelian ratio (Table 5.2). The segregation ratio of 3:1 that was observed in 7 lines (1.4, 1.19, 10.1, 10.35, 21.8, 26.5 and 26.20) were segregated in a Mendelian ratio of 3:1 (resistant: sensitive; Table 5.2). However, all T2 seedlings of lines 21.8 and 26.7 were 100% resistant to Kan which could be attributed to the transgene having multiple integrations in the genome.

Table 5-2: Overview of the segregation analysis of T2 seeds on kanamycin media and Chi-square test for the segregation of 3 Resistant: 1 Sensitive in T2 generation of transgenic wheat expressing OC-I and WT. 20 seeds per line were grown on media containing 10mg/ml kanamycin.

Line	No of seeds	% Germination	Resistant	Sensitive	% Resistant	X ² (3:1)	p-value
1.4	20	100	15	5	75	0	1
1.19	20	100	13	7	65	1.067	0.3017
10.1	20	100	15	5	75	0	1
10.35	20	100	15	5	75	0	1
21.7	20	100	14	6	70	0.267	0.60558
21.8	20	100	20	0	100	6.667	0.00982
26.5	20	100	19	1	95	4.267	0.03887
26.7	20	100	20	0	100	6.667	0.00982
26.20	20	100	16	4	80	0.267	0.60558
WT	20	100	0	20	0	60	0.00001

χ^2 critical value ($P \leq 0.05$), degree of freedom = 1 = 3.83

5.2.3.1 DNA confirmation the presence of the OC-I transgene in T2 plants:

To confirm Kan results, genomic DNA was extracted from individual leaves of each seedling. PCR analysis was performed to detect the presence of OC-I transgene using primers spanning OC-I (forward primer specific to actin promoter and reverse primer specific to Nos terminal; Table 2.2; Materials and Methods). A 600-bp band was present in the samples from the transgenic plants but not in the plants that were susceptible to kanamycin or untransformed control samples. This analysis confirmed the presence of the OC-I coding sequence and allowed identification of the transformed plants (Figures 5.6-5.10). The results from the selection on kanamycin and the PCR analysis indicate that the plants were correctly transformed with OC-I. In addition, T-DNA copy number was determined in individual transformed plants of each line by TaqMan assay (see section 2.19; Materials and Methods). Most of T2 transgenic lines were showed to be single insertion (Appendix I).

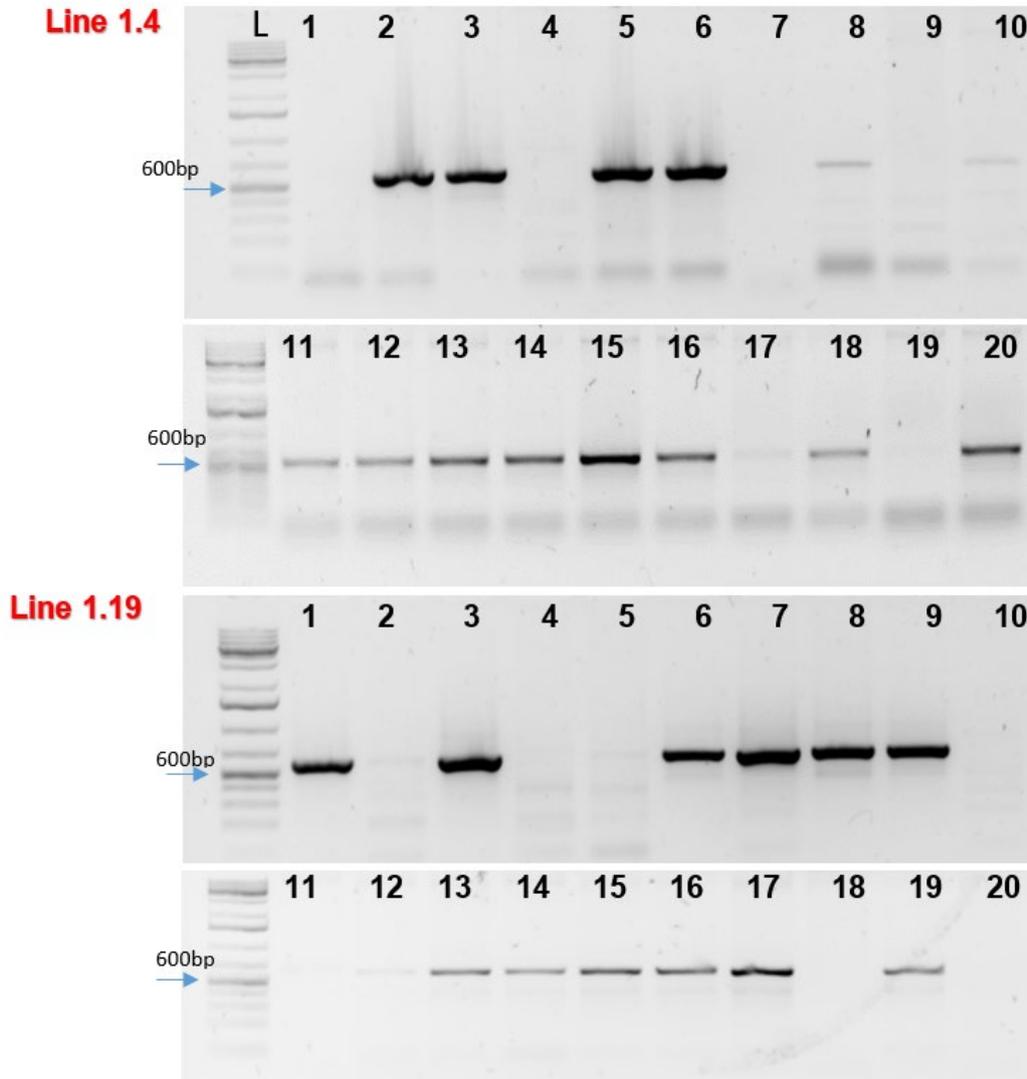


Figure 5.6: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. DNA were extracted at the three-leaf stage. Plants were grown under a under glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C, 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Lanes: L shows DNA marker, lanes 1-20 show DNA extracted from leaves of individual plants of lines 1.4 and 1.19.

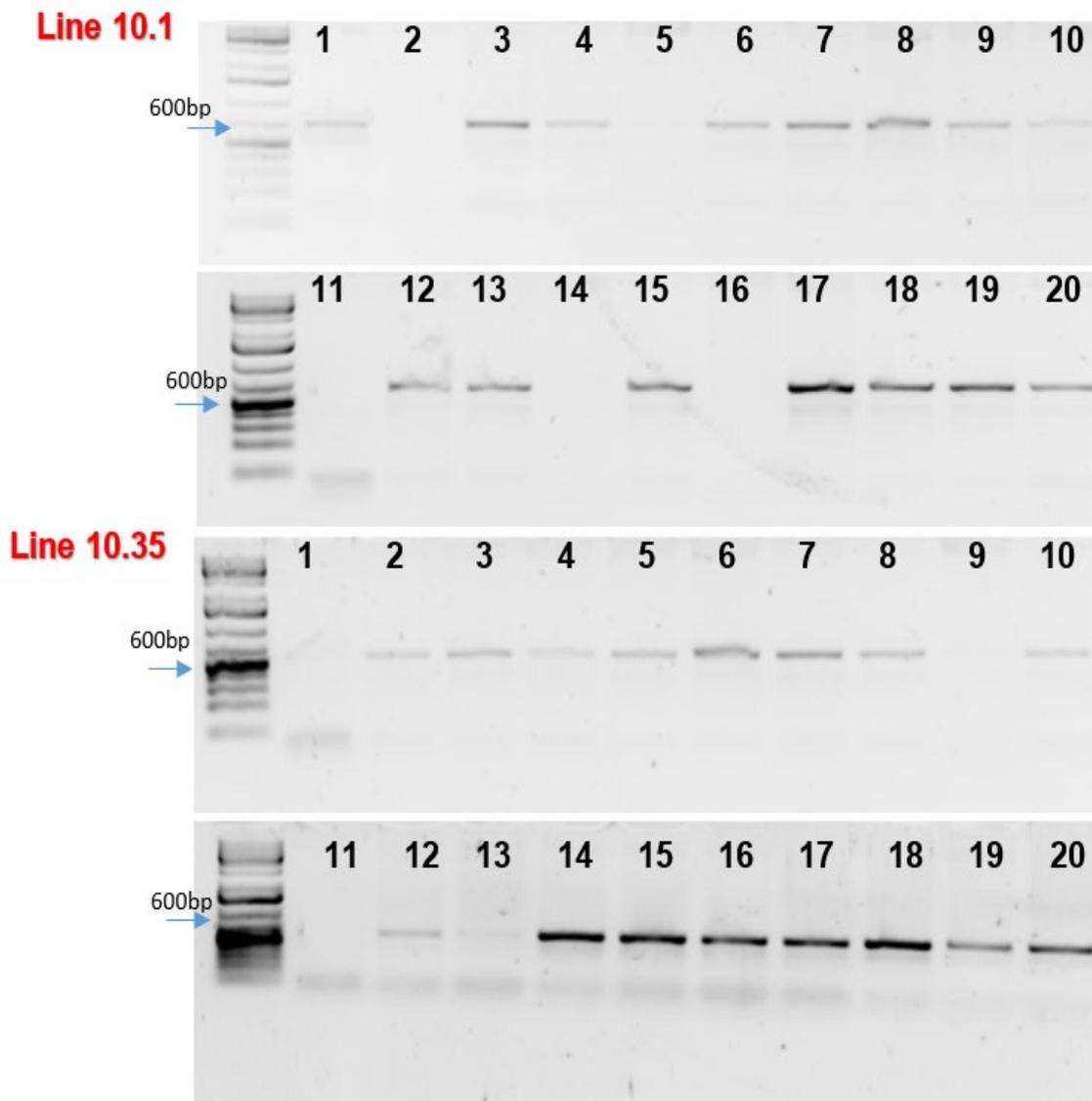


Figure 5.7: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. DNA were extracted at the three leaf stage. Plants were grown under a glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Lanes: L show DNA marker, lanes 1-20 show DNA extracted from leaves of individual plants of lines 10.1 and 10.35.

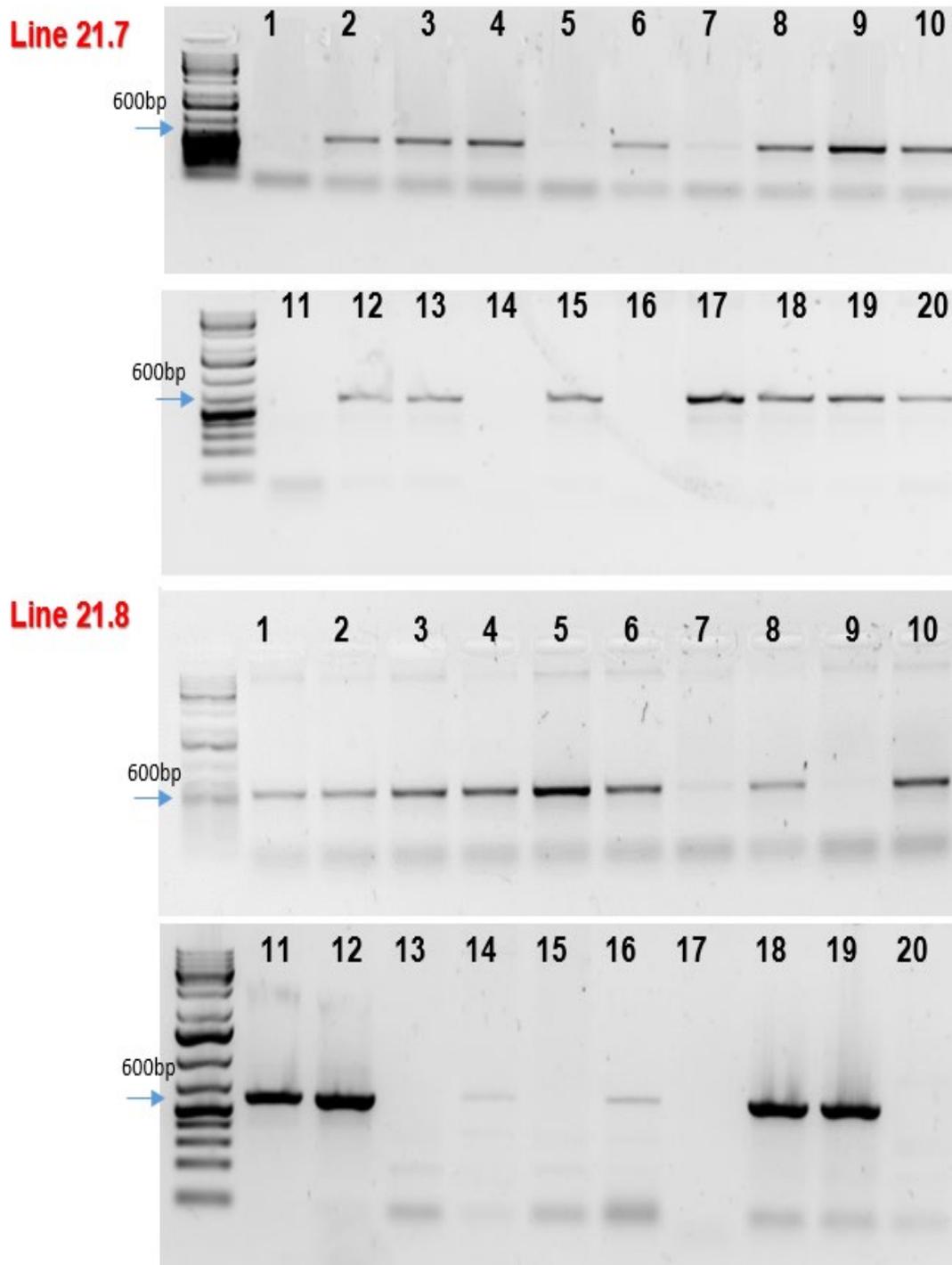


Figure 5.8: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. DNA were extracted at the three leaf stage. Plants were grown under a glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Lanes: L shows DNA marker, lanes 1-20 show DNA extracted from leaves of individual plants of lines 21.7 and 21.8.

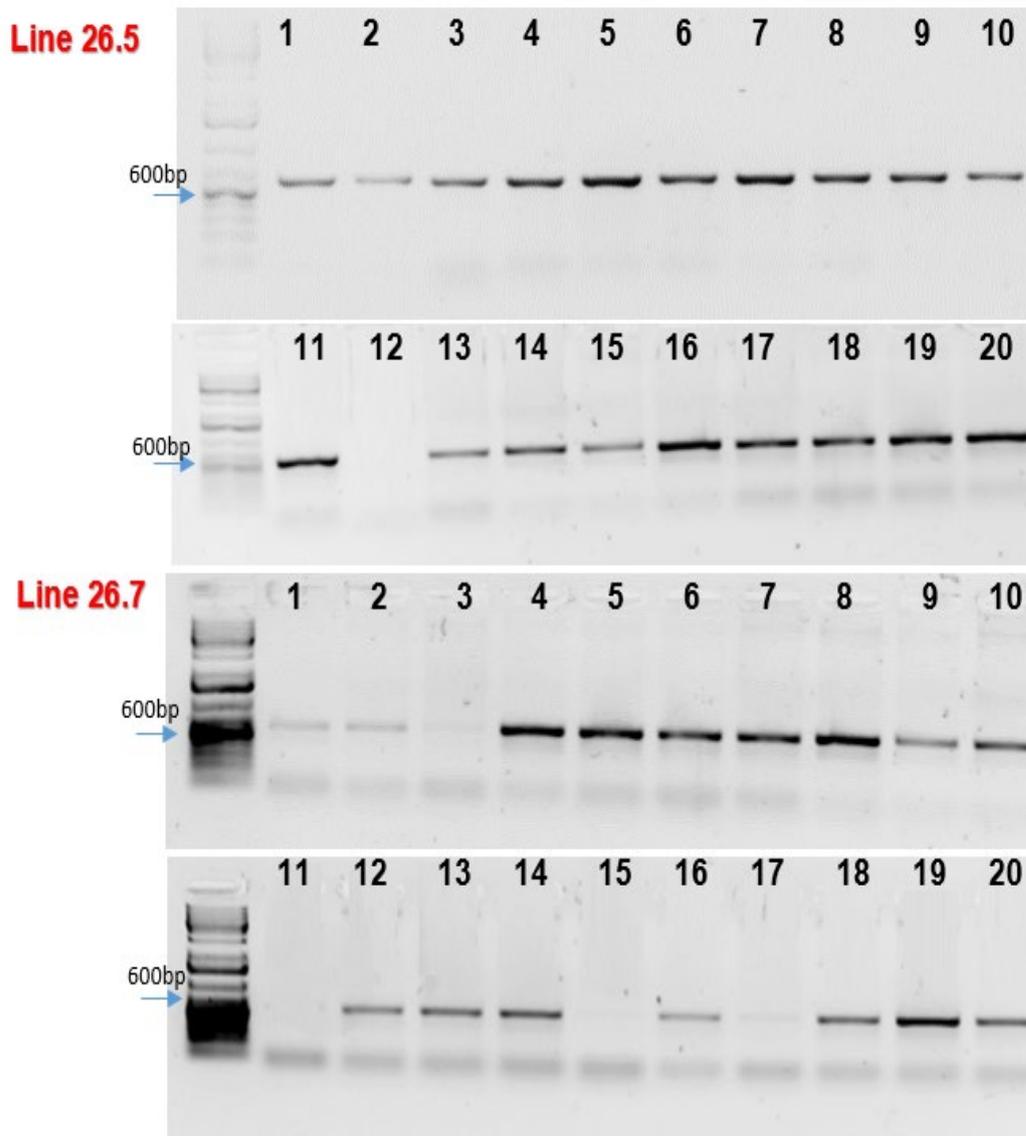


Figure 5.9: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. DNA were extracted at the three leaf stage. Plants were grown under a under glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Lanes: L shows DNA marker, lanes 1-20 show DNA extracted from leaves of individual plants of lines 26.5 and 26.7.

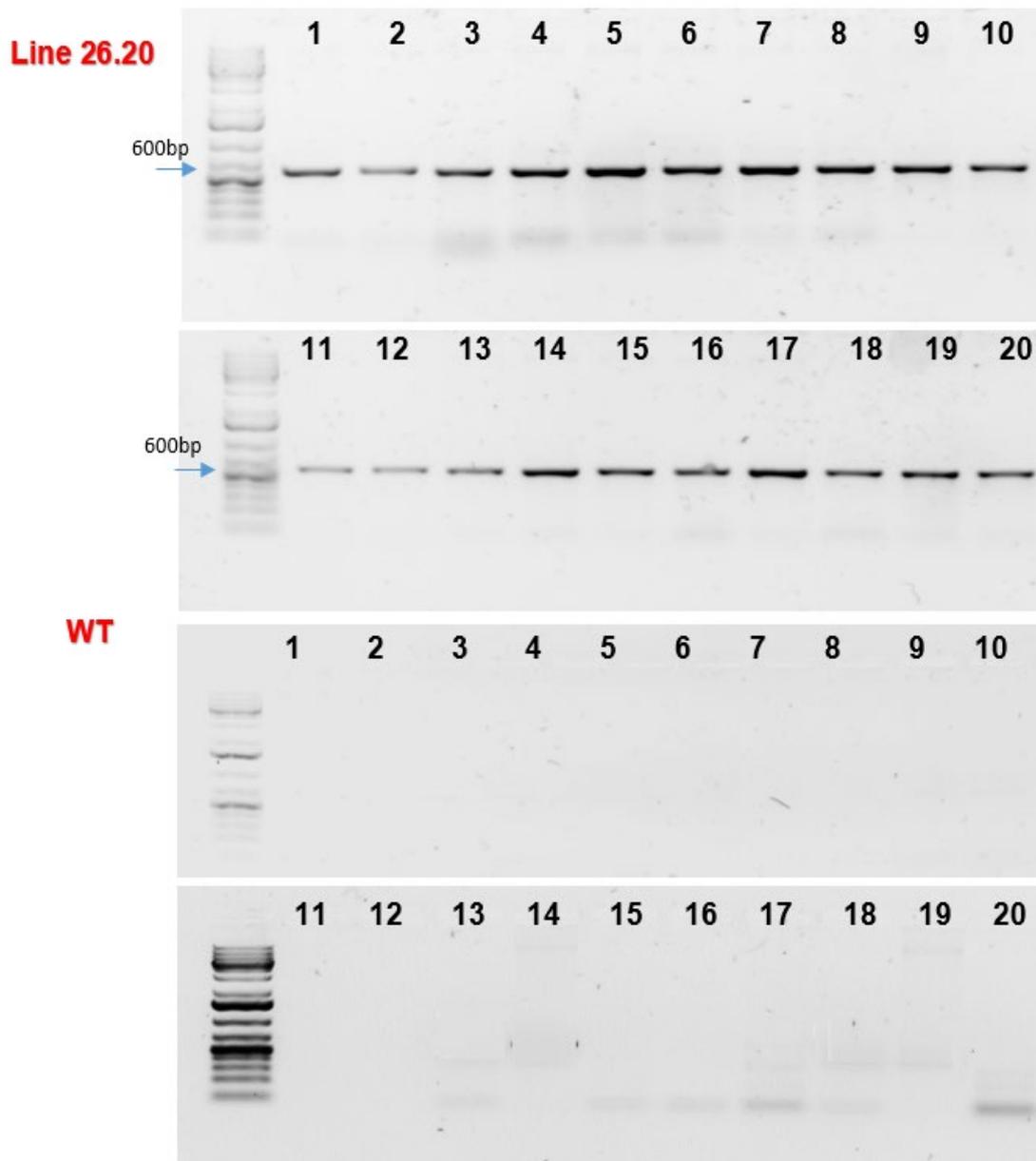


Figure 5.10: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. DNA were extracted at the three leaf stage. Plants were grown under a glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Lane L shows DNA marker, lanes 1-20 show DNA extracted from leaves of individual plants of lines 26.2 and WT (E1.8).

5.2.3.2 Analysis of relative expression level of the OC-I transgene in transgenic wheat lines

The expression of OC-I was analysed in individual transformed plants from each eight transgenic lines (1.4, 1.19, 10.1, 10.35, 21.7, 21.8, 26.7, and 20.20) using qPCR (Figures 5.11 and 5.12). The data showed that the transcript level of OC-I was higher in all plants for the various transgenic lines (Figure 5.12). However, the transcription level of OC-I was varied between selected transgenic lines, these differences were not found in plants of the same line (Figures 5.11 and 5.12). In addition, transcription level of OC-I was compared between lines and a significantly higher transcription ($p < 0.001$) was found in plants of transgenic lines 1.19, 21.7 and 26.20 (Figure 5.13). The seeds of plants 1.19.7, 21.7.2, 26.20.4 and E1.8.5 therefore were used to obtain T4 seeds depended on their PCRs, copy number and transcription level of OC-I.

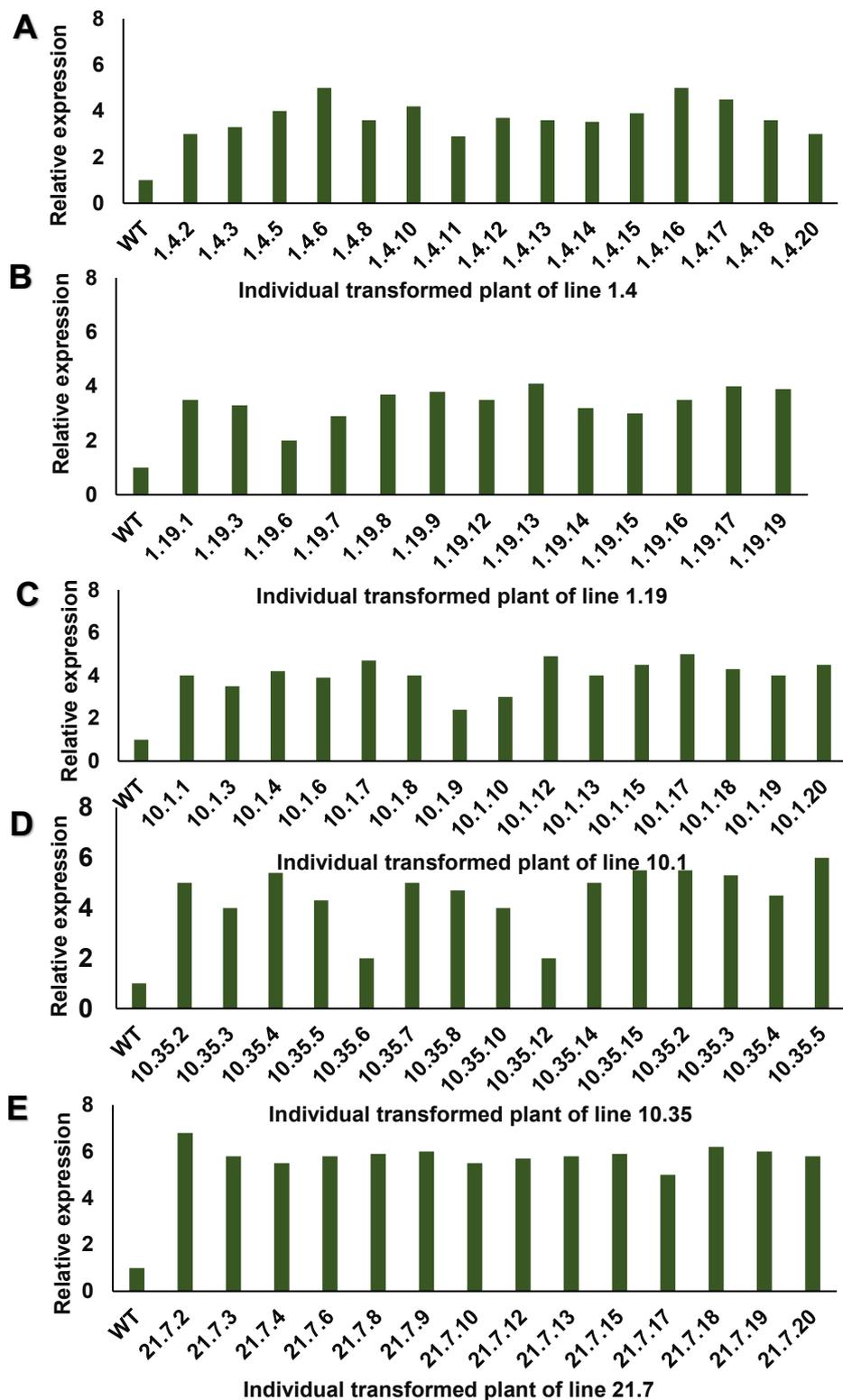


Figure 5.11: Analysis of relative expression level of the OC-I transgene in individual transformed plant from each transgenic wheat lines using specific primers for OC-I (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. Plants were grown under a glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. The expression of OC-I in WT is assigned a value of 1 and the expression of the individual plant in each line are as follows: A: line 1.4; B: line 1.19; C: line 10.1; D: line 10.35; E: line 21.7. The data were normalized to the wheat ACTIN gene.

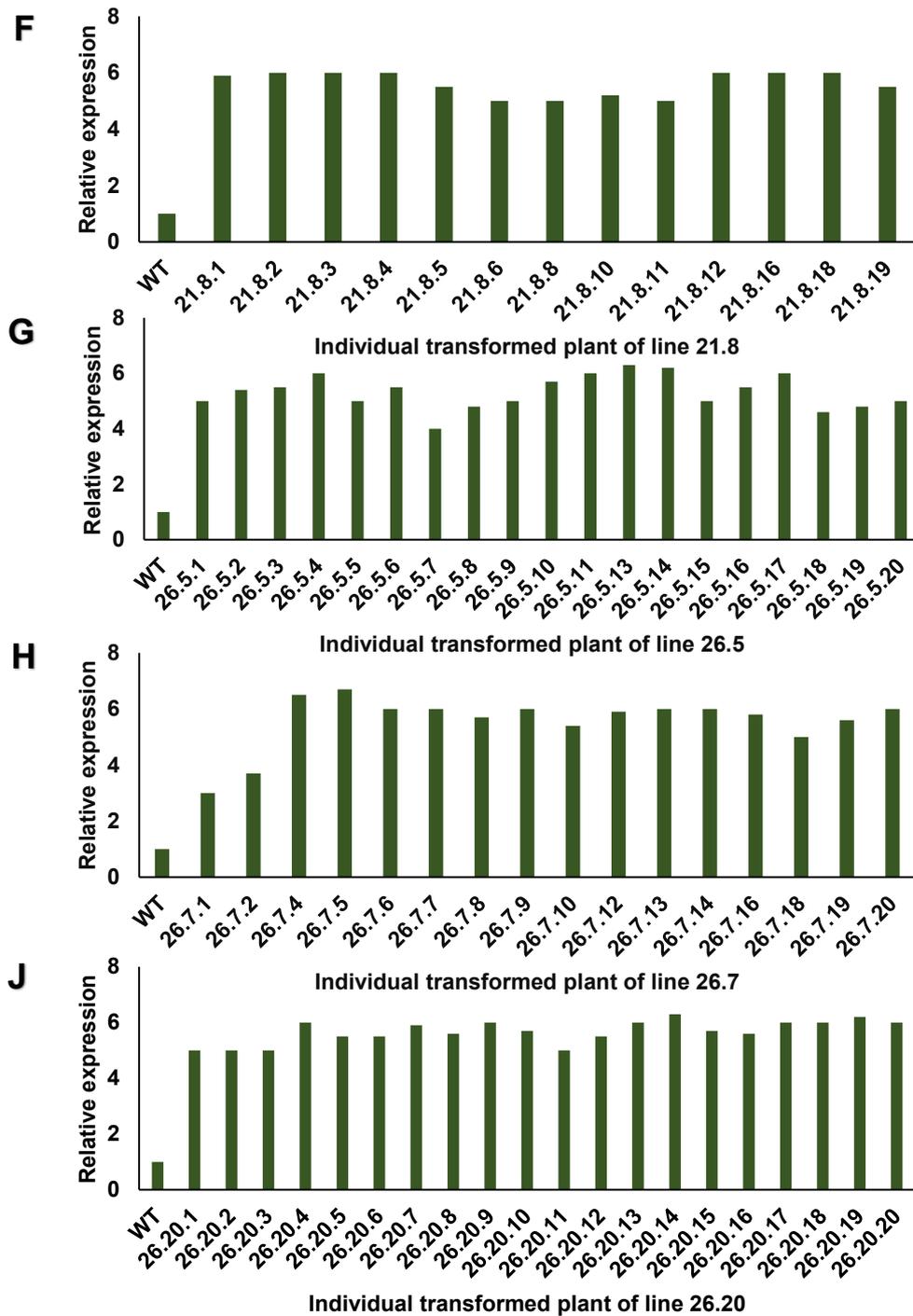


Figure 5.12: Analysis of relative expression level of the OC-I transgene in individual transformed plant from each transgenic wheat lines using specific primers for OC-I (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. Plants were grown under a glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. The expression of OC-I in WT is assigned a value of 1 and the expression of the individual transformed plant of each line are as follows: F: line 21.8; G: line 26.5; H: line 26.7; J: line 26.20. The data were normalized to the wheat ACTIN gene.

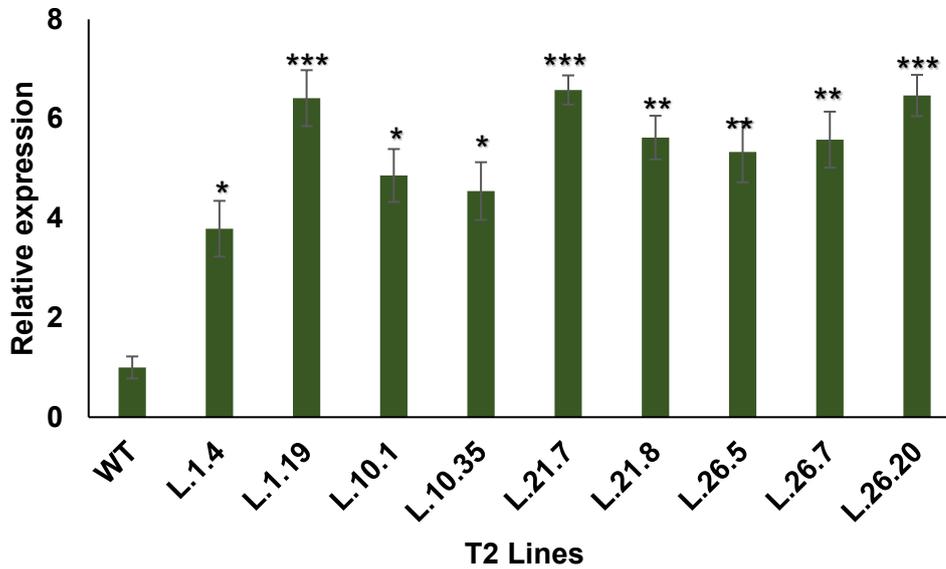


Figure 5.13: A comparison of the expression level of the OC-I transgene in transgenic wheat lines as pooled using specific primers for OC-I (Table 2.5; Materials and Methods) from 3-week-old T2 wheat plants. Plants were grown under a glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. The expression of OC-I in WT was assigned a value of 1 and each line presents the average of total OC-I expression in plants of same line. The data were normalized to the wheat ACTIN gene. Mean \pm SD for the plants in each line are indicated by bars. The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, **p < 0.01 and ***p-value < 0.001, ANOVA).

5.2.4 Confirmation of the presence of OC-I gene in T3 plants and selection of T4 generation seeds

T3 seeds were harvested, and lines of 1.19.7, 21.7.2, and 26.20.4 was selected with the WT (E1.8.5) for T4 generation. A total of 15 seeds were sown for the transgenic line and WT onto media containing kanamycin. The T3 generation was 100% resistant to kanamycin, suggesting that the plants are homozygous (Table 5.3). Seedlings of the same developmental stage were then transplanted into soil. PCR analysis was performed to confirm the presence of the OC-I coding sequence in the transformed plants (Figure 5.14). In addition, all T3 transgenic plants were showed to be single insertion (Appendix II). Moreover, the expression level of OC-I was analysed in individual plants for each line using qPCR (Figure 5.15). This analysis showed that OC-I transcripts were present in all transgenic plants (Figure 5.15A-5.15C). In addition, the level of OC-I transcripts was similar in all lines (Figure 5.15 D). Transformed plants were harvested and seeds of lines 1.19.7.7, 21.7.2.3, and 26.20.4.2 which were named WOC1, WOC2 and WOC3, respectively. They were selected for further characterisation of the properties of the wheat seeds that overexpress OC-I (Figure 5.16A). PCR analysis was performed to confirm the presence of the OC-I in these seeds (Figure 5.16B). The level of OC-I expression was investigated in the WOC1, WOC2 and WOC3 seeds (Figure 5.17). The abundance of OC-I was high in the transgenic seeds but absent from the WT. These seeds were thereafter used for analysis of total protein and the composition of storage proteins (Chapter 6).

Table 5-3: Overview of segregation analysis of T3 seeds on kanamycin media.

Line	Total No of Seeds	Seeds Germinated	Germination %100	Sensitive	Resistant	%resistant
1.19.7	15	13	87	0	13	100
21.7.2	15	15	100	0	15	100
26.20.4	15	14	93	0	15	100
WT	15	15	100	15	0	0

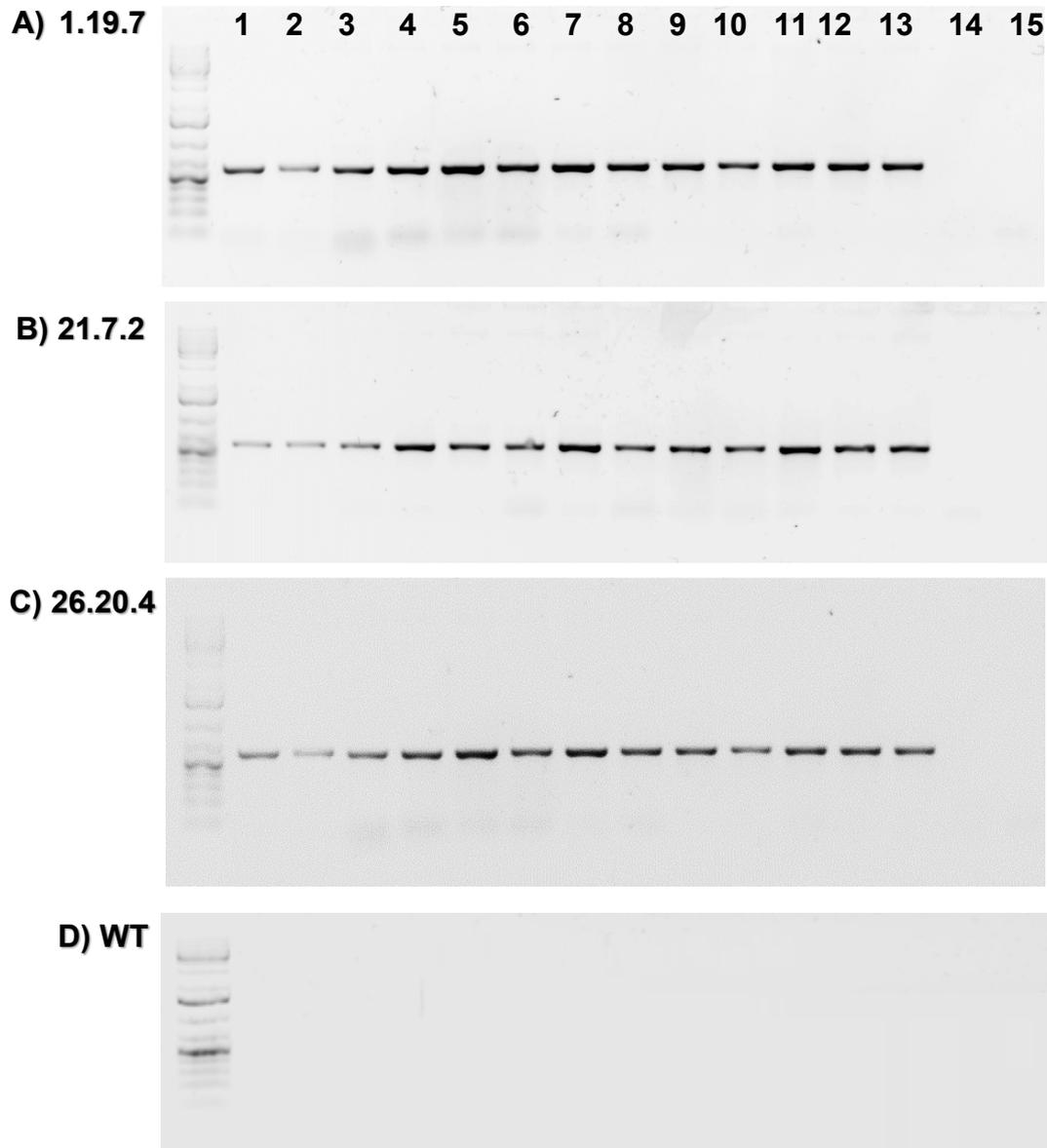


Figure 5.14: PCR amplification of the OC-I transgene (600 bp) using primers Act-GOI and NosT-rev (Table 2.5; Materials and Methods) from 3 old- week- T3 wheat plants. DNA were extracted at the three-leaf stage. Plants were grown under a under glasshouse conditions with a 16h day/8 night photoperiod at a day/night temperature of 20/15°C 60% humidity and a light intensity of 400 $\mu\text{Em}^{-2}\text{s}^{-1}$. Lane L shows the DNA marker; lanes 1-13 show DNA extracted from leaves from individual plants of lines 1.19.7(A), 21.7.2 (B) and 26.20.4 (C) and WT (D). Lanes 14 shows the WT and lane 15 shows the negative control.

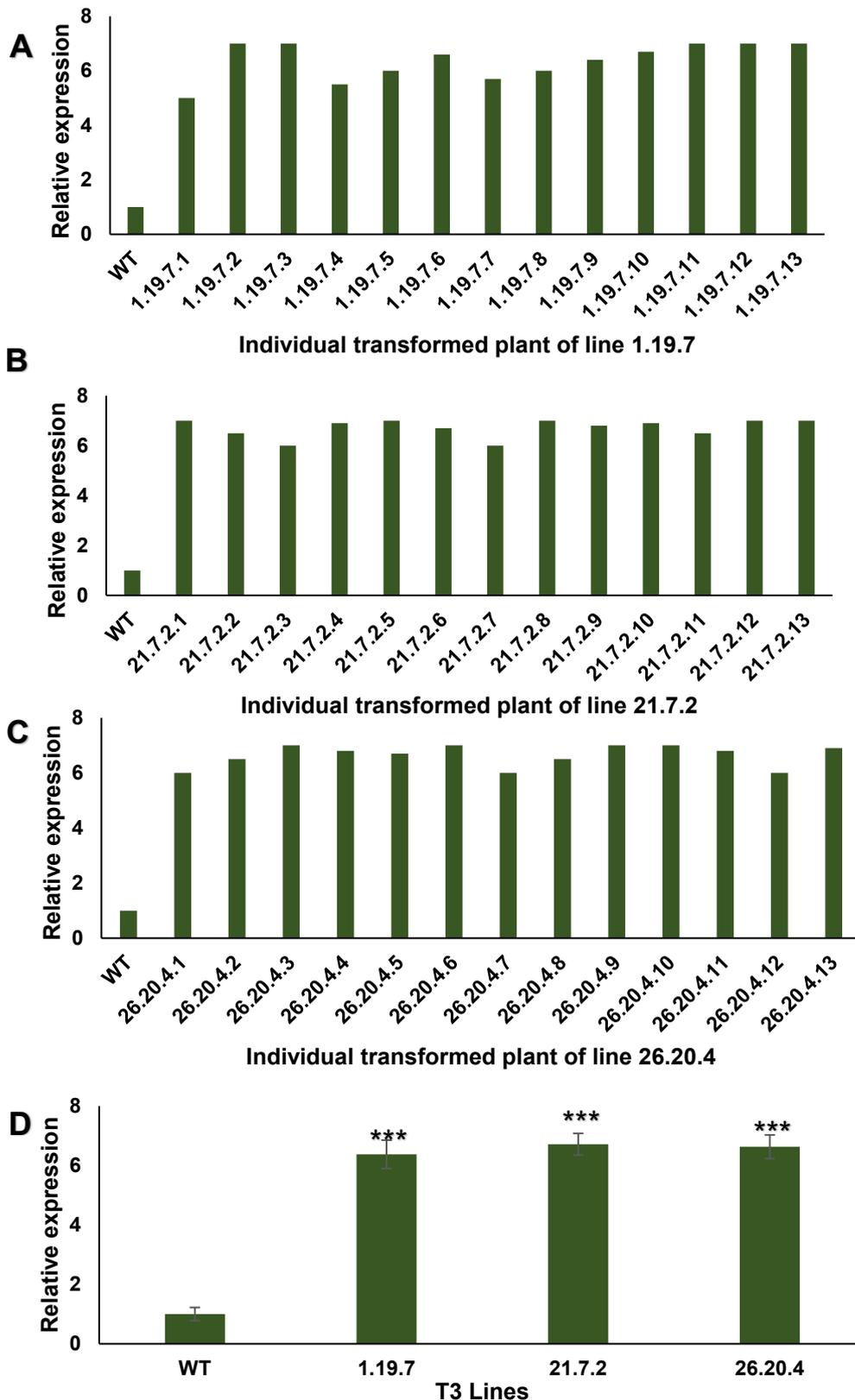


Figure 5.15: Analysis of relative expression level of the OC-I transgene using specific primers for OC-I (Table 2.5; Materials and Methods) from 3-week-old T3 wheat plants. The expression of OC-I in WT was assigned a value of 1 and the expression of the individual plant of each line are as follows: A: line 1.19.7; B: line 21.7.2; C: line 26.20.4; D: relative expression as pooled for each line. The data were normalized to the wheat ACTIN gene. Mean \pm SD for the plants in each line are indicated by bars. The asterisks indicate significant differences to WT plants (***)p-value < 0.001, ANOVA).

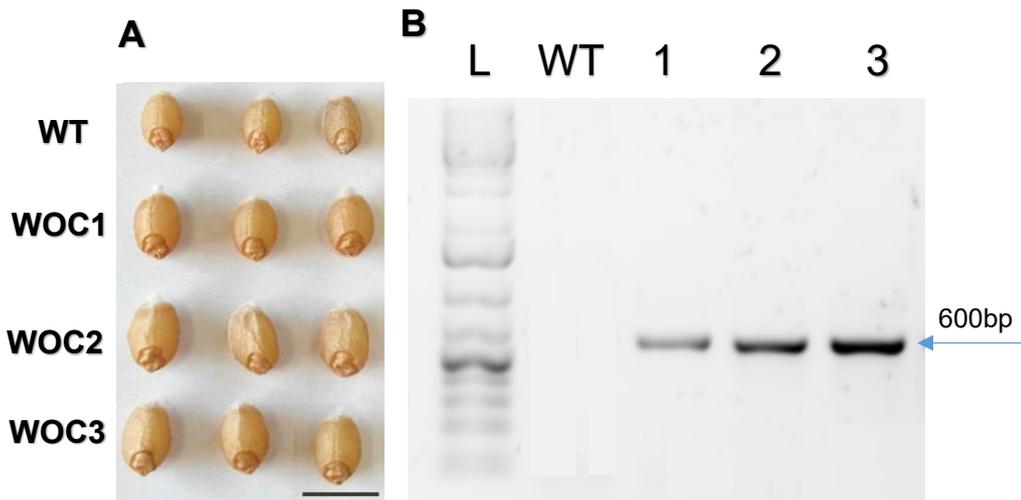


Figure 5.16: A) A visual comparison of the size of the WOC1, WOC2 and WOC3 seeds compared to the WT. B) PCR amplification of the OC-I transgene (600 bp) from T4 wheat seeds. Lane L shows the DNA marker; lanes 1: WT, 2: WOC1, 3: WOC2 and 3: WOC3.

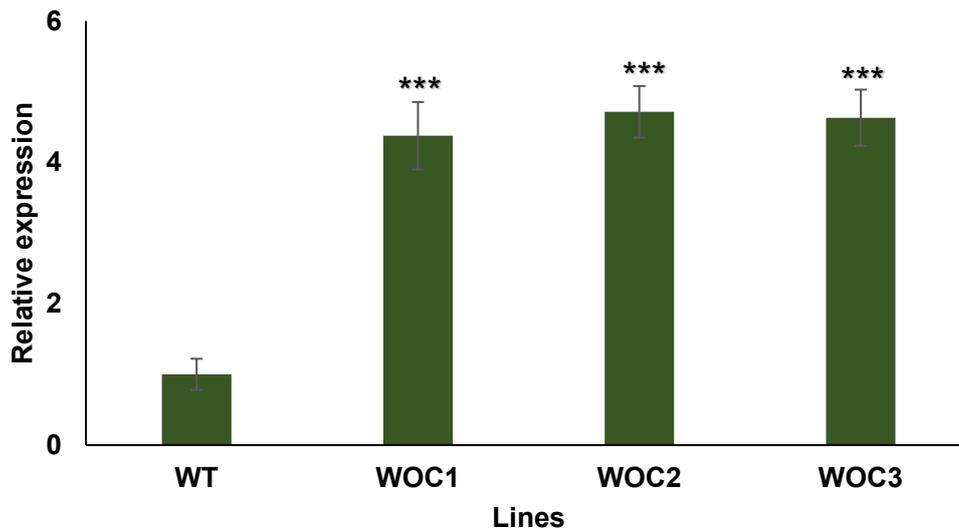


Figure 5.17: Relative expression of the OC-I transgene in transgenic wheat seeds (WOC1, WOC2 and WOC3) and WT. The expression of OC-I in WT was assigned a value of 1. The data were normalized to the wheat ACTIN gene. Mean \pm SD for the seeds are indicated by bars. The asterisks indicate significant differences to WT plants (***)p-value < 0.001, ANOVA).

5.3 Discussion

The data of the present investigation demonstrated the selection of the T4 generation of a number of transgenic wheat lines and showed that the OC-I gene was present and stably integrated into the wheat genome through generations. The protocol followed in this study was to select transgenic wheat lines on media containing kanamycin, followed by confirmation of the presence of the OC-I transgene using PCR. Previous studies indicate that it is possible for epigenetic silencing of the kanamycin transgene to occur in later generations and, furthermore, that in kanamycin sensitive transgenic lines the kanamycin gene should be confirmed by PCR (Yenofsky et al., 1990; Bastaki and Cullis, 2015). The results from kanamycin selection and the PCR analysis indicate that the plants were correctly transformed with OC-I. In addition, it is important to determine copy number which may influence the gene expression and stability of the gene. The traditional method to determine copy number is Southern blotting, however, it requires large amounts of plant material, it is laborious and time consuming (De Preter et al., 2002; Gadaleta et al., 2011). In present study, the detection of OC-I transgene copy number was determined using TaqMan assay of the nptII gene relative to a single copy wheat gene amplicon (Milner et al., 2018). The results obtained in present study confirmed that single insertion lines were selected in each generation. The level of OC-I transcripts was determined for plants in each generation using qPCR. There were differences between transcripts level of OC-I in the transgenic lines as previously observed (Van der Vyver et al., 2003; Quain et al., 2015). A number of different factors can influence the expression of a transgene for example, copy number, methylation, location in the genome and re-arrangement of the transgene (Muskens et al., 2000; Lechtenberg et al., 2003; Tang et al., 2007). PCR, copy number, and transcripts level of OC-I results were allowed selection of

three different transgenic lines (1.19.7, 21.7.2, and 26.20.4) containing the OC-I coding sequence and it clearly shows that T4 seeds were able to be collected. Future experiments will seek to characterise the phenotypes of the transgenic lines compared to the WT under both normal and stress conditions, thereby providing a better understanding of the role of OC-I in crop plants and how it could help to improve performance and quality traits. In the next chapter, the total protein and the composition of storage proteins in transgenic wheat seeds will be examined and compared to those of the WT.

Chapter 6 : Properties of Arabidopsis, soybean and wheat seeds over-expressing *Oryzacystatin* I (OC-I)

6.1 Introduction

Seed storage proteins are important determinants of crop establishment and yield and they are one of the most significant sources of dietary protein. Harvested seeds are important in nutritional intakes for humans, livestock and domesticated animals (Li et al., 2012). The protein levels in seeds varies between plant species. In many monocotyledonous species such as *B. distachyon*, *O. sativa* and *Z. mays*, the seed protein level is typically around 10% of the dry weight. In dicotyledonous seeds, such as *A. thaliana*, *G. max* and *C. sativus*, the seed protein level can be over 30% of the seed dry weight (Jacks et al., 1972; Chileh et al., 2010).

Pulses, or grain legumes, are a mainstay in the diets of much of the world's population. Beans, chickpeas, lentils and peas are important sources of dietary protein, particularly of world's poorest populations (Boye et al., 2010). Soybean proteins contain most if not all of the essential amino acids required in the human diet (Patil et al., 2017). They are therefore nutritionally well-balanced. They also contain compounds, which have significant health benefits such as a reduced risk of cardiovascular disease and hyperlipidaemia (Nishinari et al., 2014). Soybeans contain water and oil and are hence also useful for the mass-production of food because of their emulsifying properties (Nishinari et al., 2014).

Cereal grains contribute in excess of 200 million tonnes of protein annually to the diets of people and livestock across the world (Shewry and Halford, 2002). This is approximately three times more than protein-rich foods, such as legumes, which comprise about 20-40% protein. Cereal grain proteins fulfil a plethora of functions.

About 80% are storage proteins that are found in the endosperm, together with lipids and starch (Shewry et al., 1995; Shewry and Halford, 2002). Storage proteins are synthesised throughout the development of the grain and during grain maturation (Martinez, Manuel et al., 2019). They are degraded during germination, a process which involves the action of various seed proteases. Moreover, the cereal seed proteins play an important role in the processing of grain in the food industry. This is especially true of wheat, which is used to make bread, pastry, breakfast cereals and other products which are common constituents of the human diet.

6.1.1 Dicot and monocot seed storage proteins

There are four categories of seed storage protein: albumins and globulins (both of which are dicot storage proteins), glutelins and prolamins (both of which are monocot storage proteins) (Radhika and Rao, 2015). The 2S albumins are a major class of dicot seed storage protein, which have been studied extensively in the *Cruciferae*, particularly *B. napus* and *A. thaliana* (Li et al., 2012). With the notable exceptions of oats and rice, the main endosperm storage proteins of cereal grains have many subgroupings, including the high molecular weight (HMW) prolamins, as well as the sulphur-rich (S-rich) and sulphur-poor (S-poor) prolamins (Shewry and Halford, 2002). Perhaps the most widely distributed group of storage proteins are the globulins, which are part of the cupin superfamily and are present in both dicots and monocots. β -conglycinin (7S) and glycinin (11S) are the two main storage proteins of soybean (40% and 60% respectively), make up approximately 70% of the soybean seed protein overall (Schmidt et al., 2011; Wei et al., 2020). β -conglycinin is found as various combinations of homologous polypeptide subunits (α' , α and β) and has a molecular weight of 81, 74 and 50 kDa (Thanh and Shibasaki, 2002; Taski-Ajdukovic et al., 2010).

6.1.2 The degradation of storage proteins

Seeds store lipids, starch and proteins that are mobilised during germination, and used to drive metabolism and seedling growth to the point where photosynthesis is established. Seed storage proteins are synthesised in the endoplasmic reticulum (ER) and transported to protein storage vacuoles (PSVs) by Golgi-independent pathways (Jolliffe et al., 2005; Vitale and Hinz, 2005; Galland et al., 2014). The degradation of seed storage portions begins in the imbibition phase of seed germination. Various proteases contribute to the process of protein degradation during germination. Cysteine proteases (CysProt) are particularly important in the degradation and mobilisation of storage proteins (Grudkowska and Zagdanska, 2004; Tan-Wilson and Wilson, 2012; Szewińska et al., 2016). Of these, the most widely studied are the C1A papain-like CysProt and the C13 legumain or vacuolar processing enzymes (VPEs) (Grudkowska and Zagdanska, 2004; Szewińska et al., 2016; Botha et al., 2017).

Phycystatins (PhyCys) are important in the degradation of seed storage proteins and also at different stages of plant growth and development (Solomon et al., 1999; Díaz-Mendoza et al., 2014). However, the precise functions of PhyCys in the regulation of seed protein accumulation and composition germination are largely uncharacterised. Moreover, the role of PhyCys in the control of proteolysis during seed germination remains to be clearly defined.

The analysis of leaf protein contents during vegetative growth and development in transgenic *Arabidopsis* plants overexpressing OC-I in the cytosol or the chloroplasts reported in Chapters 3 and 4 demonstrates that inhibition of cysteine proteases leads to a higher level of protein accumulation in the leaves compared to the wild-type (WT). Therefore, it is important to explore the effects of OC-I overexpression in seeds, particularly with regard to the accumulation and subsequent mobilisation of storage proteins in the seeds with overexpression of OC-I. These may also provide greater insights into the mechanisms by which cysteine proteases regulate the degradation of storage proteins. In the present study, the transgenic *Arabidopsis*, soybean and wheat seeds that overexpress OC-I were characterised in terms of the content and composition of seed protein relative to the WT.

The research question in this chapter is based on the idea that restricting cysteine protease activities may be a general approach to modifying seed protein contents as well as the seed size and weight. The objective of the studies reported in this chapter was to investigate the role of OC-I on the size and properties of *Arabidopsis*, soybean and wheat seeds, in order to address the following questions: (i) Does OC-I as an inhibitor of papain-like cysteine proteinases play a role (or roles) on seed size and weight? (ii) Does OC-I influence the extent of storage protein accumulation and composition in seeds (iii) Do the transgenic seeds show altered dormancy or germination relative to the wild type?

6.2 Results

The following experiments were performed to understand the roles of OC-I in seeds, with a particular focus on storage proteins. The properties of *Arabidopsis*, soybean and wheat seeds overexpressing OC-I in comparison to their respective wild types were characterised in the following studies.

6.2.1 Transgenic *Arabidopsis* seeds overexpression OC-I

Germination was studied in two type of OC-I expressing lines (CYS lines, where the transgene was not targeted and the PC lines, where the transgene was targeted to the plastids) relative to the WT. There were some differences in the germination of the transgenic seeds compared to the WT. For example, the appearance of the radicle was slower in transgenic lines than in the WT line (see Chapters 3 and 4).

The weight of 100 CYS seeds and 100 PC seeds was measured and compared to that of 100 WT seeds. A significant increase in seed weight was observed in both transgenic lines compared to the WT (Fig. 6.1). In particular, there was an increase of approximately 0.14 g and 0.9 g in the PC and CYS seeds, respectively compared to the WT (Figure 6. 1A).

The total protein content of dry PC and CYS seeds was increased significantly compared to the WT (Figure 6. 1B). The PC lines had three times the protein content of the WT, while the CYS lines accumulated twice as much protein as the WT (Figure 6. 1B). The protein composition of the transgenic seeds was compared the WT. SDS-PAGE staining with Coomassie-blue was performed on protein extracts from dry seeds of the WT and the transgenic lines (CYS1, CYS3 and CYS4: Figure 6. 1C; PC2, PC7 and PC9: Figure. 6.1C). The main storage proteins that could be detected in WT *Arabidopsis* seeds are globulin (12S) and albumin (2S). The globulin α and β subunits

have a molecular mass of approximately 34 and 20 kDa, respectively, while the L and S albumin subunits have a molecular mass of 15 and 3 kDa, respectively. The results presented in Fig. 6.1 show that the PC lines accumulate more of the 12S and 2S subunits of the seed storage proteins than the CYS lines and the WT (Fig. 6.1C). Moreover, the level of albumin protein appeared to be lower in the CYS lines than the PC and WT lines (Figure 6. 1C).

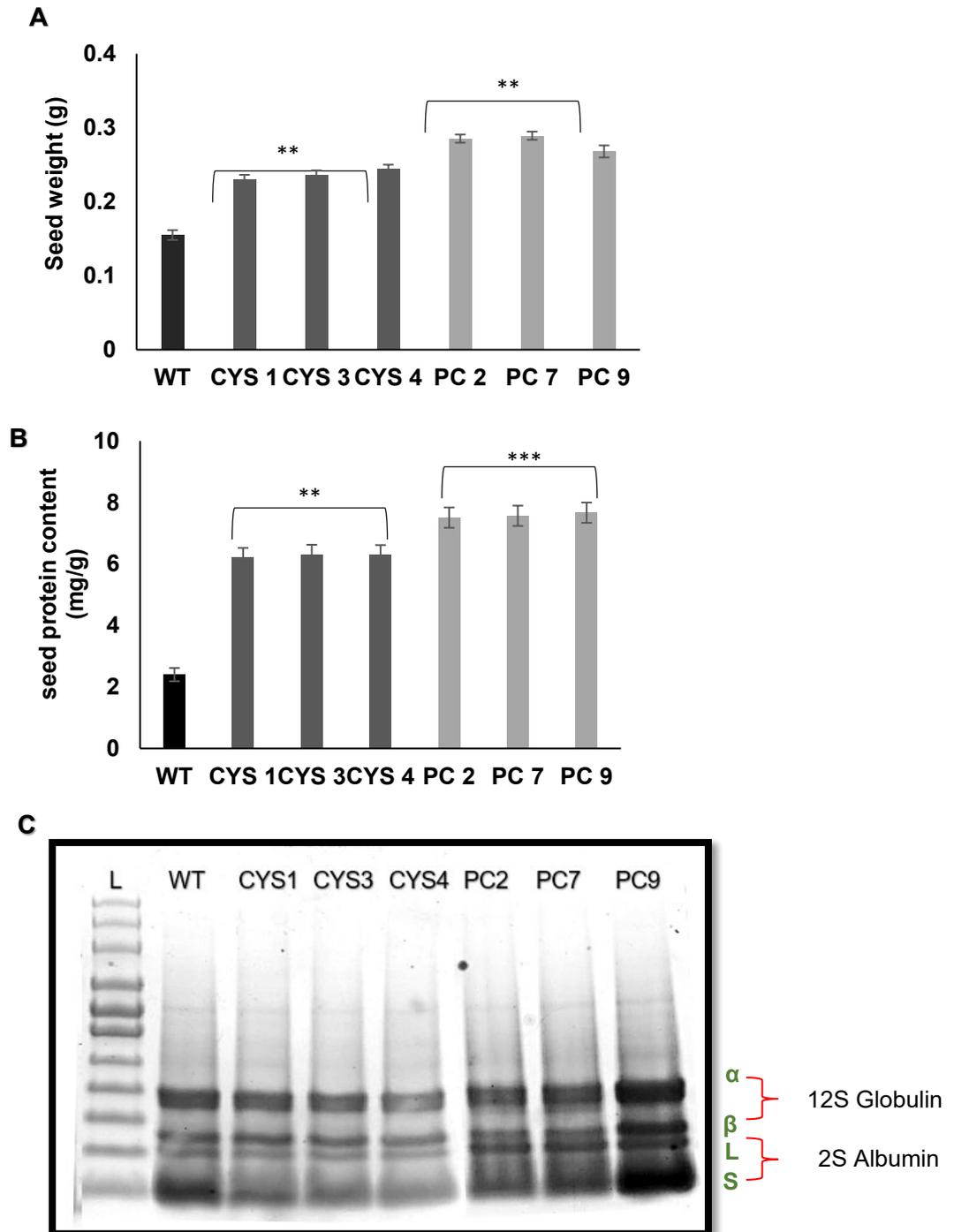


Figure 6.1: A comparison of the seed weight, total protein content and the protein composition of the seeds of wild type (WT) *Arabidopsis* and OC-I expressing lines (CYS1, CYS3, CYS4, PC2, PC7 PC9). (A) Dry seed weight (n=100 seeds) and (B) total protein content of WT (black bar), CYS lines (grey bar) and PC lines (light grey). The data shown represents the mean \pm standard deviation. The asterisks indicate significant differences to WT plants (**p < 0.01 and ***p-value < 0.001, ANOVA). (C) SDS-PAGE profiles after Coomassie-blue staining of protein extracts from dry seeds of both the WT and the transgenic lines (CYS1, CYS3, CYS4, PC2, PC7 and PC9). An equal amount of each extract corresponding to 20 μ g was applied to each lane of a 12% polyacrylamide gel. Lane L is the size of the molecular mass marker (250 kDa). The arrows mark the migration of the 12S globulin subunits (α and β) and the 2S albumin subunits (L and S).

6.2.2 The effects of Oryzacystatin I (OC-I) expression in soybean seeds

6.2.2.1 Seed morphology

The seeds of three transgenic soybean lines (SOC1, SOC2 and SOC3) and WT plants were grown in order to collect the T4 generation seed. Following seed collection, 50 seeds per line were grown in soil under controlled conditions. The percentage germination per transgenic line was estimated as the number of germinated seeds divided by the total number of planted seeds (50) multiplied by 100. All seeds had high germination rates, at around 100%. No significant differences were found in the germination rate between the lines (Figure 6. 2A).

All three transgenic lines had a higher grain yield than the WT, as determined by the number of seeds per plant. The SOC1 and SOC3 lines had an average of 150 seeds per plant, compared to 100 seeds in the WT plants (Figure 6. 2B). In addition, the seeds of the transgenic lines were visibly larger than those of the WT (Figure 6. 2C). The average weight of 50 seeds was significantly greater for SOC1, SOC2 and SOC3 lines (with an average of 20g) compared to the WT (at around 14g) (Figure 6. 2D).

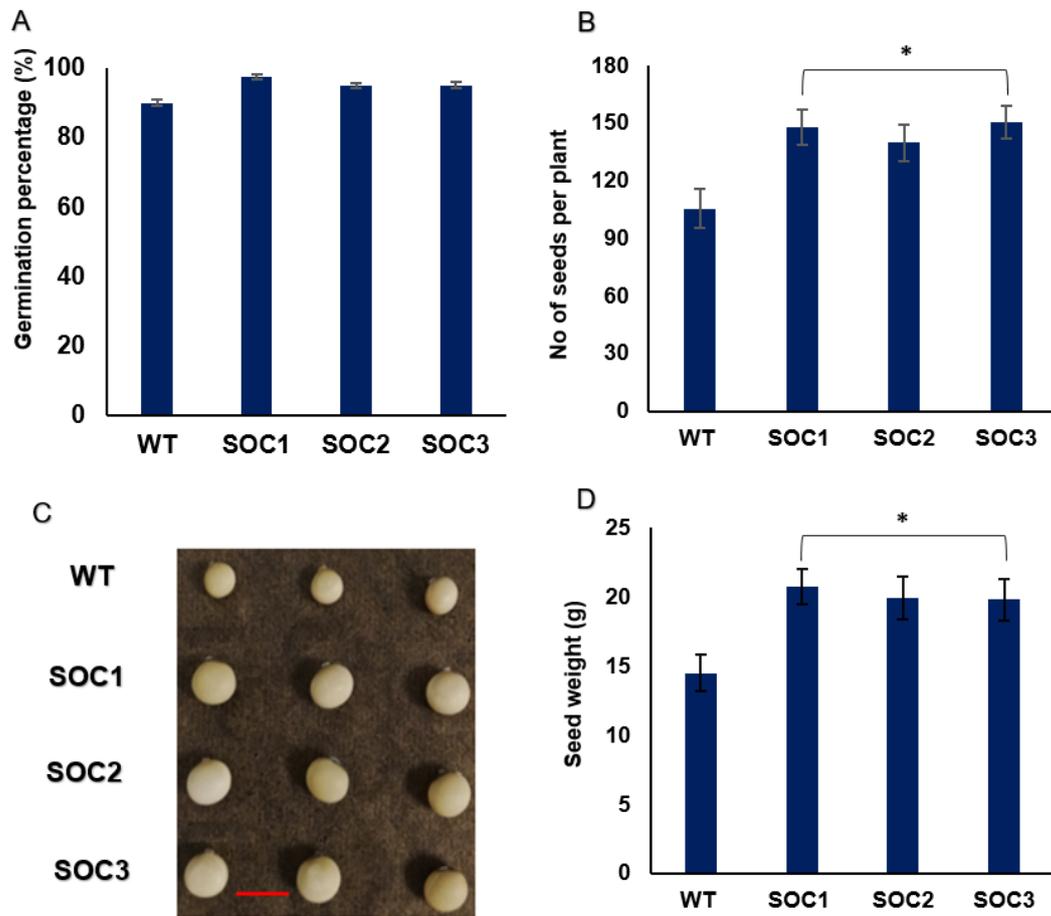


Figure 6.2: A comparison of seed morphology of soybean seeds of the WT and the three independent transformed lines expressing OC-I (SOC-1, SOC-2 and SOC-3). (A) Germination (n=100 seeds); (B) number of seeds per plant; (C) seed size; (D) seed weight (n=50 seeds). The data shown represents the mean \pm standard deviation. The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA).

6.2.2.2 Soybean seed protein content and composition

The seeds of the transgenic and WT lines were processed to make soy flour, as used in markets (see section 2.10: Method). The protein content of all the transgenic and soy flour lines was then measured using three independent samples and compared to that of WT. The SOC1, SOC2 and SOC3 flour had a significantly higher protein content than the WT (Figure 6. 3A). The results shown above demonstrate that the seeds from the SOC1, SOC2 and SOC3 lines had significantly more protein than the wild type. It was therefore important to determine whether the composition of the seed proteins was changed as well as the protent contents. The composition of the soybean seed proteins was then determined in the soybean seeds and the soy flour by SDS-Page gel electrophoresis (Figure 6.3B). This analysis revealed that there were no differences in between the visible intensity of the bands in the seeds from the SOC1, SOC2 and SOC3 lines relative to the WT. However, some differences between the lines were observed in the banding pattern of the samples of soy flour (Figure 6.3B). For example, the bands equivalent to the α' , α and β subunits of β -conglycinin (7S) have molecular weights of 81, 74 and 50 kDa, respectively. The band equivalent to β -conglycinin (7S) was more enriched in the WT soy flour samples than in the transgenic samples (Figure 6. 3B). The acidic and basic subunits of glycinin (11S) have molecular weights of approximately 35 and 14 kDa, respectively. The band equivalent to the glycinin acidic subunit appeared to be enriched in the SOC1, SOC2 and SOC3 soy flour compared to the WT flour, whereas the bands equivalent to the basic subunits were less intense (Figure 6. 3B).

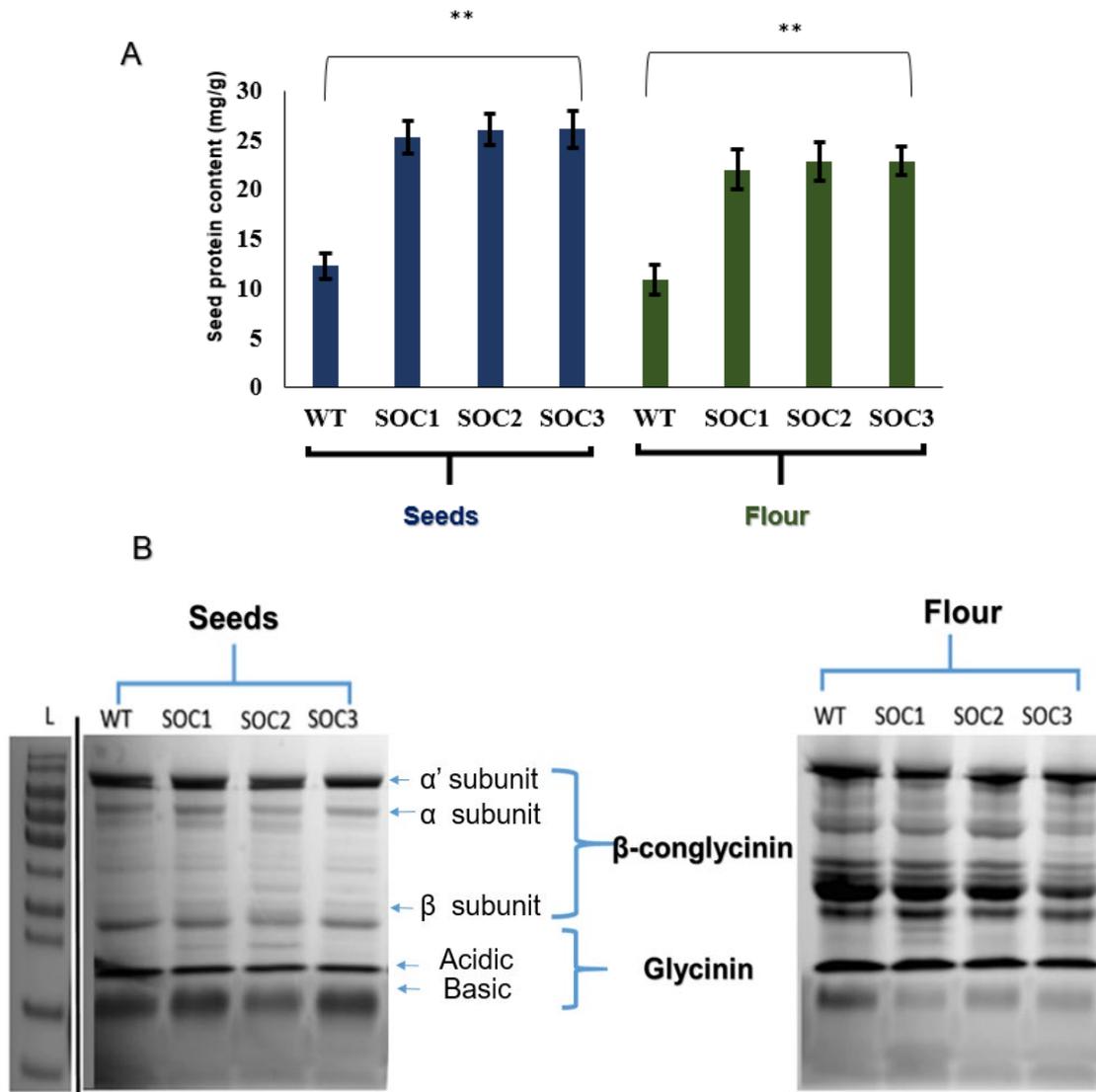


Figure 6.3: Protein content and composition of transgenic soybean seeds. (A) A comparison of protein content of the soybean seeds and soy flour with the WT and three independent transformed lines expressing OC-I: SOC-1, SOC-2 and SOC-3. The data shown represents the mean \pm standard deviation. The asterisks indicate significant differences to WT plants (**p-value < 0.01, ANOVA). (B) SDS-PAGE profile of the seed and flour proteins of the three transgenic lines (SOC1, SOC2, and SOC3) and the WT. An equal amount of each extract (20 μ g protein) was applied to each lane of 12% polyacrylamide gels. Gels were stained with Coomassie-blue. Lane L display the molecular mass marker (120 kDa). The arrows mark the bands equivalent to β -conglycinin and glycinin.

6.2.3 Wheat lines overexpressing OC-I

The production of the T4 generation of a large number of independent transgenic lines of spring wheat overexpressing OC-I was described in Chapter 5. The properties of the seeds of three independent transgenic lines (WOC1, WOC2 and WOC3) were then compared to those of the WT.

6.2.3.1 The seed production and germination

Firstly, 50 seeds per line were sown in soil to measure germination. No statistical differences in seed germination were observed between the transgenic lines and the WT seeds (Figure 6. 4A). The WOC1, WOC2 and WOC3 lines produced significantly more seeds per plant than the WT. The transgenic lines produced on average 350 seeds perplant compared to 250 seeds per plant in the WT (Figure 6 4B). Furthermore, the seeds of the transgenic lines were visibly larger than the WT and they had a greater dry weight than the WT (Figures 6. 4C and 6.4D).

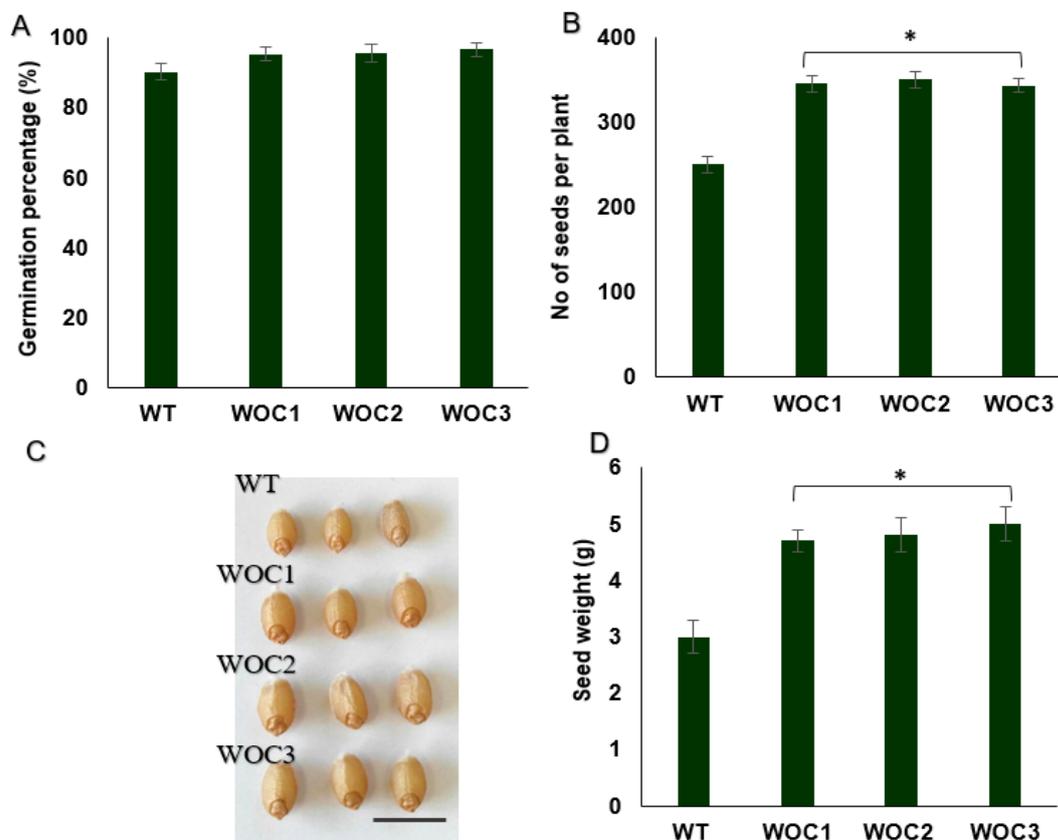


Figure 6.4: A comparison of the seed properties of the WT and three independent transformed lines expressing OC-I: WOC-1, WOC-2 and WOC-3. (A) Seed germination (n=50 seeds); (B) number of seeds per plant; (C) seed size; (D) seed weight (n=50 seeds). The data shown represents the mean \pm standard deviation. The asterisks indicate significant differences to WT plants (*p-value \leq 0.05, ANOVA). Scale bar 3 mm.

6.2.3.2 Seed protein content and composition on SDS-PAGE gels

Transgenic wheat seeds overexpressing OC-I and WT seeds were milled to remove the germ and bran. This produced a white flour, which was used to determine protein content and composition relative to the WT, as described in 2.9 and 2.11. The three WOC lines had a significantly higher protein content than the WT (Figure 6. 5A). The protein content of the transgenic WOC lines was double that of the WT (Figure 6. 5A).

The protein composition of seeds of the independent transgenic lines and WT was examined using SDS-PAGE. Wheat seed proteins form two major groups: non-prolamins (non-gluten), consisting of albumins and globulins (ALGL) and prolamins (gluten), which include gliadins and glutenins (HMW-GS and LMW-GS). The protein bands revealed by Coomassie blue staining showed typical of wheat flour (Schalk et al., 2017). These were as follows, starting from the top of the gel: HMW-GS (67–88 kDa), ω 5- gliadins (49–55 kDa), ω 1,2- gliadins (39–44 kDa), LMW-GS, α - gliadins and γ - gliadins (28–39 kDa) and ALGL (10–25 kDa). A comparison of the bands in the flour from the transgenic lines suggests that they have a greater accumulation of albumins and globulins than the WT flour (Figure 6. 5B).

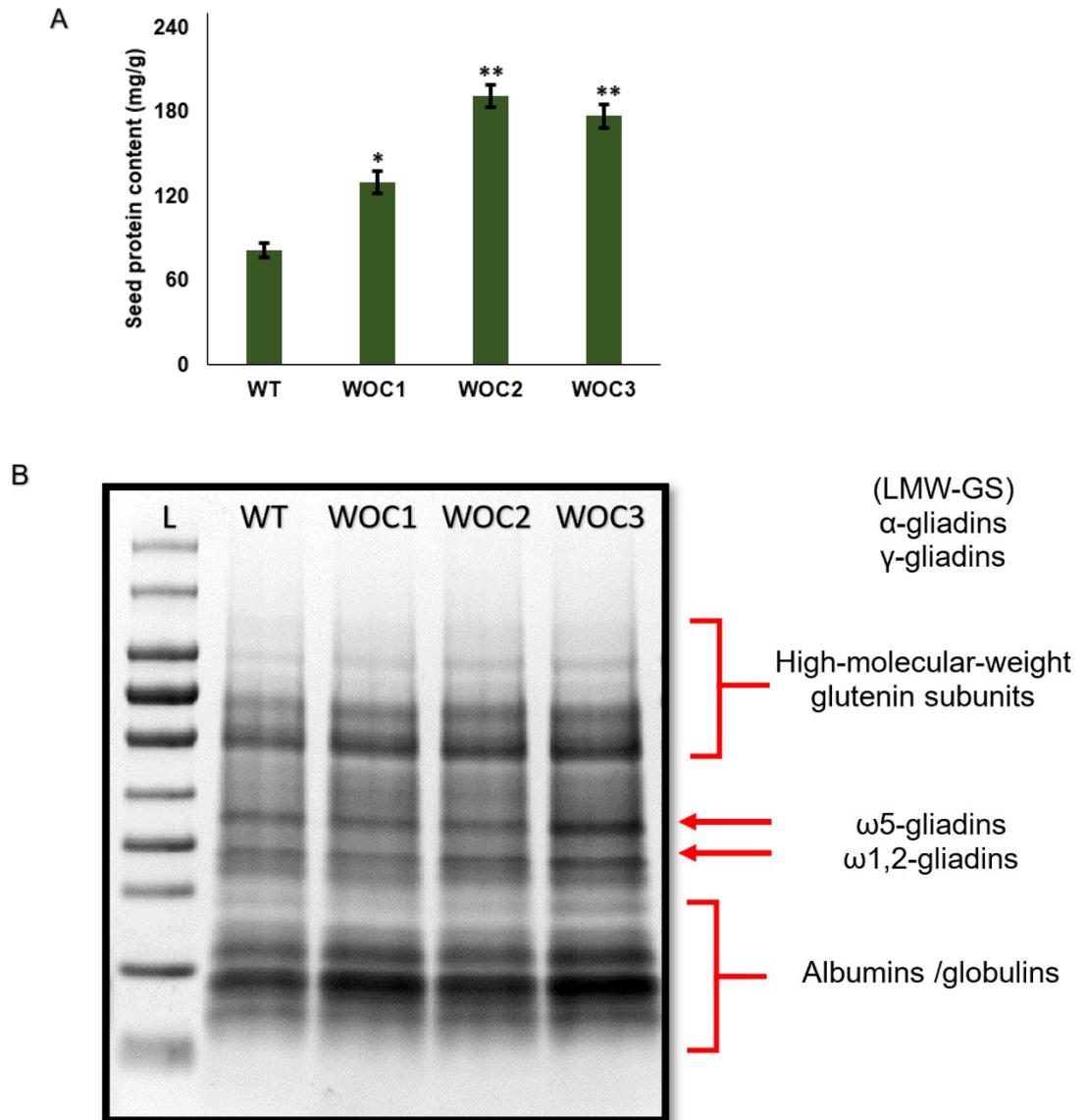


Figure 6.5: Protein content and composition of transgenic wheat seeds. (A) A comparison of the protein content of WT wheat flour and three independent transformed lines expressing OC-I: WOC-1, WOC-2 and WOC-3. The data shown represents the mean \pm standard deviation. The asterisks indicate significant differences to WT plants (**p-value < 0.01, ANOVA). (B) A comparison of the protein composition of the flours prepared from three transgenic lines (WOC1, WOC2, and WOC3) and the WT. An equal amount of protein (20 μ g) was applied to each lane of 12% polyacrylamide gels. After SDS-PAGE electrophoresis gels were stained with Coomassie-blue. Lane L is the size of the molecular mass marker (200 kDa). The arrows mark the migration of seed proteins which are high molecular-weight glutenin subunits (HMW-GS) (ω 5-gliadins and ω 1,2-gliadins), and low-molecular-weight glutenin subunits (LMW-GS) (α -gliadins and γ -gliadins and albumins/globulins (ALGL)).

6.2.4 Proteomic analysis of wheat seeds

6.2.4.1 Quantification Overview

The SDS gel electrophoresis analysis reported above suggested that there were differences in the protein composition between the WOC1, WOC2 and WOC3 lines and the WT. A proteomic analysis of the different seed types was therefore performed to investigate these differences further. Label-free quantitative proteomic analysis was performed in the Advanced Mass Spectrometry Facility at the School of Biosciences at the University of Birmingham (UK). Figures 6.6, 6.7 and 6.8 show the differences in protein composition between the WOC1, WOC2 and WOC3 and WT lines.

A total of 127 proteins were identified in the seeds of all lines (Table 6.1). These were divided into two categories as follows. Proteins with a quantitative ratio greater than 1.5 were considered to have increased in abundance while those with a quantitative ratio of less than 1/1.5 were considered to have decreased in abundance. Proteins with quantitative ratios above 1.5 or below 1/1.5 were considered to show significant changes in abundance. The differentially expressed proteins are shown in Table 6.2, which shows that the WOC1, WOC2 and WOC3 lines have 25, 9 and 22 proteins respectively that are increased in abundance relative to the WT, with 15, 47 and 34 proteins respectively that are decreased in abundance relative to the WT.

Table 6-1: The number of proteins identified in each line. The total number of proteins identified is also shown.

Sample name	Proteins
WT	90
WOC1	89
WOC2	77
WOC3	82
Total	127

Table 6-2: The differentially expressed proteins in each line relative to the wild type.

Group name	Total	Up-regulated (FC>1.5)	Down-regulated (FC<1/1.5)
WT_vs_WOC1	40	25	15
WT_vs_WOC2	56	9	47
WT_vs_WOC3	56	22	34

6.2.4.2 Protein identification

The 127 proteins identified in these studies are involved in a variety of biological processes. Of the identified proteins, particular attention was paid to differences in seed storage proteins between the OC-I-expressing lines and the WT. A comparative analysis of these proteins was performed and they were found to cluster into three groups: (i) the total proteins identified in seeds, (ii) storage proteins that wheat allergies and coeliac disease in humans, and (iii) storage proteins that are responsible for maintaining the quality of the wheat. The intensity profile of seed storage proteins is presented as a heat map, together with a hierarchical clustering of the 40 most differentially changed proteins (Figure 6. 6). All differentially expressed proteins are presented in Appendix III.

For simplicity the proteins that are differentially changed in all three transgenic lines are clustered together (WOC) for comparison to the WT in Figures. 6.6 and 6.7. The abundance of some storage proteins that cause wheat allergy and coeliac disease was decreased in the WOC seeds relative to the WT (Table 6. 3). For example, Avenin-like proteins that belong to the Prolamin superfamily, particularly Avenin-like b8, Avenin-like a2 and Avenin-like a5, were much less abundant in the WOC seeds than the WT (Figure 6. 7).

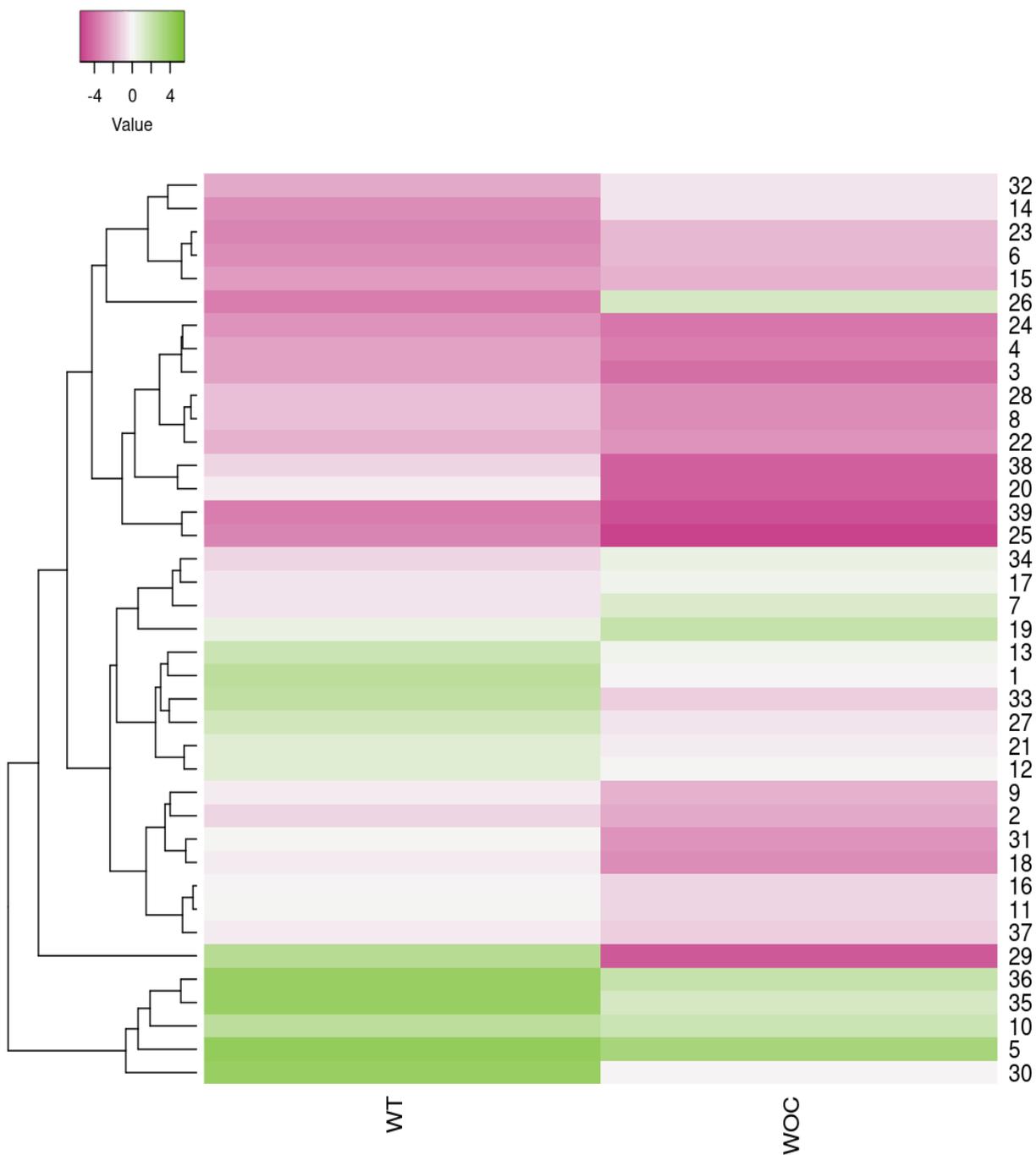


Figure 6.6: Cluster heat map representing the relative abundance of storage proteins in both the WT and WOC lines. The colour scale bar shown in the top left of the heat map presents the average fold change.

Table 6-3: A comparison of differential abundance of storage proteins that can cause wheat allergy and coeliac disease between the WT and OCI-expressing lines.

	Protein accession	Protein description	Protein abundance change (log ₂ FC)
1	D6QZM5	Avenin-like b8	0.69
2	P02863	Alpha/beta-gliadin	1.65
3	P04727	Alpha/beta-gliadin clone PW8142	0.46
4	P04730	Gamma-gliadin (Gliadin B-III)	0.97
5	P06659	Gamma-gliadin B	0.27
6	P0CZ07	Avenin-like a2	0.71
7	P0CZ09	Avenin-like a5 (LMW-gliadin 1111)	0.73
8	P10387	Glutenin, high molecular weight subunit DY10	0.09
9	P18573	Alpha/beta-gliadin MM1	0.66
10	W5A8E0	60S ribosomal protein L	0.62

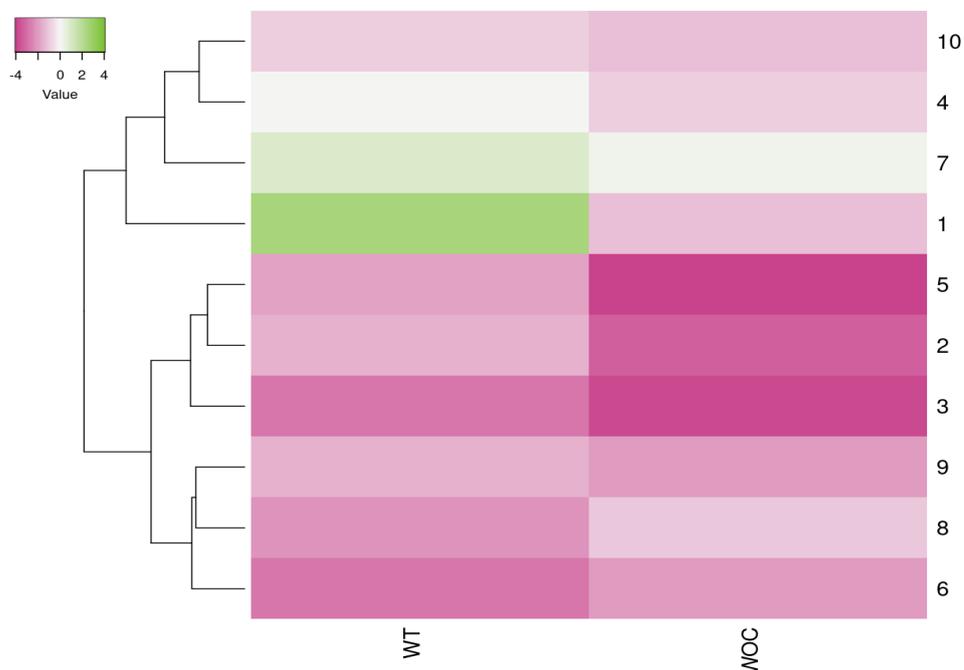


Figure 6.7: Cluster heat map of the abundance of the different storage proteins in the WT and WOC lines that can cause wheat allergy and coeliac disease. The colour scale bar shown in top left of the heat map presents the average fold change.

Table 6-4: A comparison of differential abundance of storage proteins that are important for wheat quality between the WT and WOC lines.

	Protein accession	Protein description	Protein abundance change
1	D2KFG9	Gliadin/avenin-like seed protein	0.68
2	P08453	Gamma-gliadin	0.72
3	P10385	Glutenin, low molecular weight subunit	1.16
4	Q2A784	Avenin-like a1 (LMW-gliadin 2482)	1.19
5	Q43659	15kDa grain softness protein	0.96

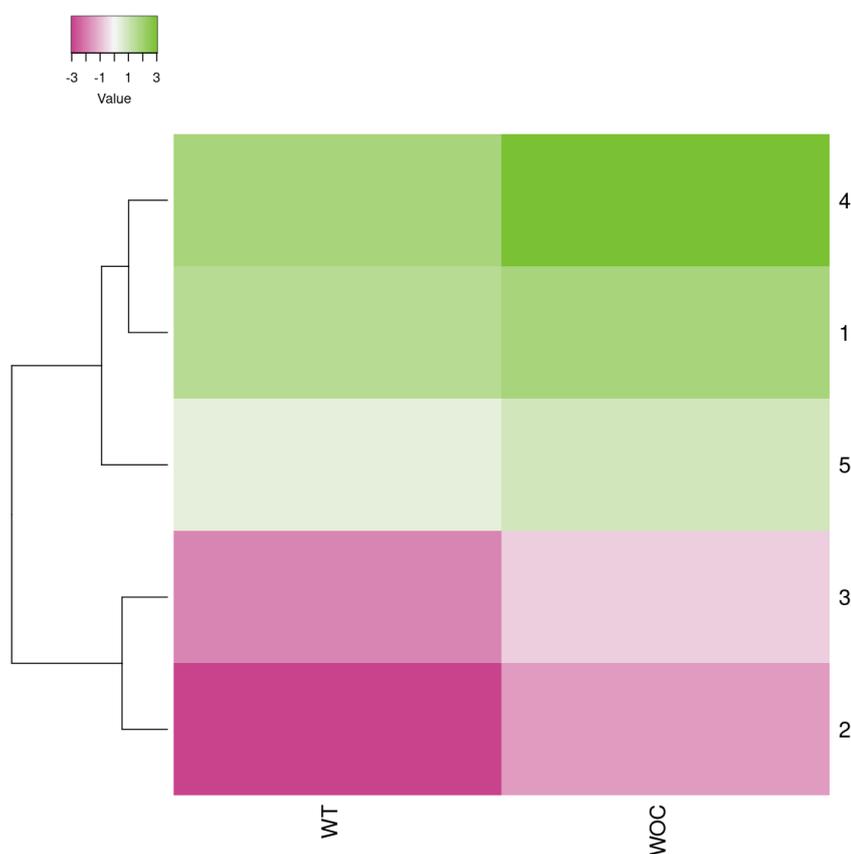


Figure 6.8: Cluster heat map of the relative abundance of various storage proteins in the WOC lines and the WT that are important for wheat quality. The colour scale bar shown in the top left of the heat map presents the average fold change.

6.3 Discussion

Cysteine proteases play important roles in seeds and seed germination. They are required for the processing of seed storage proteins, which they hydrolyse to deliver essential amino acids for the growth and development of the embryo (Gruis, D.F. et al., 2002). Moreover, they are important in the remobilisation of storage proteins during seed germination, morphogenesis and hormone signalling (Grudkowska and Zagdanska, 2004; Cambra et al., 2012; Diaz-Mendoza et al., 2016; Szewińska et al., 2016; Liu, H. et al., 2018; Martinez, Manuel et al., 2019). Cysteine proteases of the C1A and C13 families are responsible for most of the proteolytic activities in cereal and leguminous seeds, respectively leading to the mobilisation of storage proteins (Zhang and Jones, 1995; Hara-Nishimura et al., 1998; Cambra et al., 2012; Diaz-Mendoza et al., 2016). The activities of these cysteine proteases is regulated by endogenous phytocystatins (PhyCys) (Diaz-Mendoza et al., 2016). Since PhyCys regulate papain-like cysteine protease activities during seed germination and seedling development (Hong et al., 2007), the roles of OC-I in seed germination and development were compared in the model species (*Arabidopsis thaliana*), a grain legume crop (soybean; *Glycine max*) and a cereal crop (wheat; *Triticum aestivum*). The data presented here reveal the constitutive expression of OC-I has a strong effect on seed production (in terms of yield), seed size and seed protein content. Moreover, there were some significant differences in the composition of seed storage proteins in the transgenic lines expressing OC-I compared to the WT. Taken together, these findings suggest that OC-I expression improves seed production and quality in soybean and wheat, two major crop species, as well as in *Arabidopsis*.

There were no differences in the germination of the OC-I-expressing *Arabidopsis*, soybean or wheat seeds relative to the wild type. This suggests that despite the

important role of cysteine proteases in seed germination, constitutive OC-I expression has no negative effects on this process. However, the germination rate (the appearance of radicle) was slower in the transgenic *Arabidopsis* seeds with overexpression of OC-I compared to the WT (Chapter 3), confirming the results of a previous study investigating the properties of transgenic *Arabidopsis* seeds overexpressing OC-I (Roa-Roberts, 2014). Moreover, research on the overexpression of AtCYS6, a phytocystatin that regulates seed germination, showed that transgenic *Arabidopsis* plants expressing AtCYS6 had a slowed germination rate and lower CP activities (Hwang et al., 2009). In addition, the overexpression of *Brassica rapa* phytocystatin 1 (BrCYS1) was shown to delay the germination of transgenic *Arabidopsis* seeds (Hong et al., 2007). Similar effects were shown in potato tubers with ectopic expression of cereal cystatins (Munger et al., 2015). However, no adverse effects on seed germination were observed in transgenic soybean lines with OC-I expression (Quain et al., 2014). Taken together, these results of the studies reported here suggest that OC-I has no negative effects on germination in the transgenic soybean and wheat seeds. Furthermore, although seedling development was slower in the transgenic *Arabidopsis* seeds compared to the WT, the total number of germinating seeds was not decreased. Such differences may be due to the interaction of OC-I with the various cysteine proteases present in each species and the affinities of OC-I to the CPs that bind this inhibitor (Quain et al., 2014).

Seed yield and seed size are essential for a plant survival and are, therefore, important agronomic traits in flowering plants. Ectopic overexpression of phytocystatins has previously been shown to improve seed yields (Pillay, P et al., 2012). The data presented here clearly demonstrate that OC-I expression significantly increases seed yield, and seed size and weight in two important crop species. OC-I expression led to

significant increases in seed production per plant in the transgenic lines of all three species (*Arabidopsis*, soybean and wheat) relative to the WT. These increases in seed yield may be a result from increased shoot branching or a delay in the senescence as previously reported (Prins et al., 2008; Quain et al., 2014). Moreover, OC-I expression also improves tolerance to abiotic stresses, for example, drought and low temperature (Prins et al., 2008; Quain et al., 2014). OCI-expression in tobacco significantly increased the pod size, with a significantly greater average seed pod dry weight and a higher number of seeds per plant compared to the WT (Quain et al., 2014). Taken together, these findings suggest that OC-I expression has the potential to significantly improve both grain yield and grain weight in crop species.

Cysteine proteases are abundant in seeds. They are responsible for the mobilisation and degradation of seed storage proteins (Grudkowska and Zagdanska, 2004). Cysteine protease-dependent hydrolysis is responsible for up to 90% of the total prolamin activity in germinated-wheat sourdough (Grudkowska and Zagdanska, 2004). In addition, globulins proteins, which are major protein in *Arabidopsis* seeds, are degraded by cysteine proteases (Jinka et al., 2009). The abundance of both β -conglycinin and glycinin, which are the main components responsible for seed protein quality in soybean, are influenced by OC-I expression. Protease C2 is a cysteine protease enzyme responsible for the degradation of the β subunit of the β -conglycinin storage protein, whereas cysteine protease C1 degrades the α' and α subunit (Seo et al., 2001). The data presented here demonstrate that the expression of OC-I enhanced the seed protein content of all transgenic lines compared to the WT. These results confirm and elaborate the data reported previously for transformed soybean lines, which showed that the SOC lines had significantly more seed protein than the WT (Quain et al., 2014). In addition to total seed protein content, the

composition of seed proteins was investigated using Coomassie blue staining of proteins separated by SDS-PAGE electrophoresis. These studies indicate that the protein composition of the seeds of all the species was changed by OCI-expression in the transgenic lines. The PC lines that express OC-I that is specifically targeted to plastids accumulate more storage proteins than the CYS lines, where OC-I is expressed without a targeting sequence and the WT. An analysis of changes in the intensity of bands on the Coomassie blue stained gels of soybean seed flour proteins suggests that β -conglycinin bands were enriched compared to the WT. Moreover, the Coomassie blue stained gels of wheat seed flour proteins, suggests that the wheat flour contains more albumins and globulins than the flour prepared from the WT seeds. These results suggest that OC-I may still exert specific effects during flour production, preventing the degradation of some storage proteins but not others.

Taken together, these findings provide new insights into the role of OC-I in increasing seed protein content. The reasons why seed protein is increased in the seeds of all three species by OCI-expression are unknown but one may speculate that processes such as protein metabolism and futile cycling of proteins is constrained or inhibited during grain filling leading to a greater level of seed protein accumulation. These findings provide new insights into the functions of OC-I in the accumulation of seed proteins that merits further investigation. Moreover, while the potential of use of these lines by the food industry is constrained by public unease about GM technology, there is clear potential for the use of this information in soybean and wheat breeding programs, in which high yield and high seed protein content, as well as high protein extractability are often preferred traits.

Proteomic analysis of the transgenic wheat seeds overexpressing OC-I revealed that there were some significant differences in seed protein composition compared to the

WT. Cereal grains have lower protein levels (approximately 10-12% dry weight on average) than grain legumes (Shewry and Halford, 2002). However, because of food preferences cereals contribute approximately three times as much protein (in excess of 200mt annually) to the diets of both people and livestock across the world compared to the protein-rich grain legumes (which are about 20-40% protein) (Shewry and Halford, 2002). Moreover, the cereal seed proteins play a crucial role in the processing of grains such as wheat, which is used to make bread, pastry, breakfast cereals and other products which are commonly consumed by humans. A key question therefore concerns the effect of OC-I expression on the abundance of gluten content in the wheat seeds and flour, since cysteine proteases are involved in the degradation of seed storage proteins. Of the 127 individual proteins that were identified in the proteomic analysis of the wheat seeds, five classes of storage protein (15 kD grain softness protein, glutenin, avenin-like a1, gliadin/avenin and gamma-gliadin) were more abundant in the WOC seeds compared to the WT. These proteins not only provide energy for seed germination (Song, X.-J. et al., 2007) but they are also important determinants of wheat flour quality and are hence important to the food industry. In contrast, some storage proteins that cause allergies to wheat and coeliac disease were less abundant in WOC relative to the WT. The studies reported here on the overexpression of OC-I in seeds have therefore yielded some potentially interesting results, with wide implications for agroindustry, the food industry and medicine. The effects of overexpression of OC-I on seed properties are generally similar in all three species studied, suggesting that the overexpression of OC-I may be a generic approach to crop improvement.

Chapter 7 . Genome-wide analysis of cysteine proteases in *Arabidopsis* and wheat

7.1 Introduction

Plant proteases are crucial for proteolysis and play an active role in a wide range of processes throughout the plant life cycle (Van der Hoorn, Renier AL, 2008). For example, they are essential for optimal plant growth, senescence and seed and fruit ripening, and programmed cell death (PCD) (Grudkowska and Zagdanska, 2004; Van der Hoorn, Renier AL, 2008; Liu, H. et al., 2018; Tornkvist et al., 2019). They are localized in different cellular compartments, including the vacuole, plasma membrane and endoplasmic reticulum (see section 1.4.1).

Plant proteases belong to a diverse group of 61 clans and 253 families including aspartic-, cysteine-, metallo-, serine- and threonine-proteases, dependent on their catalytic mechanisms as described in the MEROPS database (Rawlings et al., 2016). Over 800 proteases belonging to 60 families are found in *A. thaliana* (Van der Hoorn, Renier AL, 2008). Papain-like cysteine proteases (PLCPs), which are the most abundant group of cysteine proteases, are characterised by the presence of a nucleophilic cysteine thiol at the active site (Rawlings et al., 2010b). PLCPs or C1A cysteine proteases are one of the largest classes of proteolytic enzymes that are involved in plant development, including flowering (Shahri and Tahir, 2014) embryogenesis (Van der Hoorn, Renier AL, 2008). They also play a key role in immunity and senescence, as well as plant responses to biotic and abiotic stresses (Zamyatnin, 2015). The C1A cysteine proteases belong to the C1A subfamily of the C1 family of the CA clan (Liu, H. et al., 2018). The properties of several PLCPs have

been investigated in different plant species (Zou et al., 2017). However, the genome-wide analysis of PLCPs is limited to a few species.

The current analysis is limited to the PLCPs families that have been identified in the genomes of *Arabidopsis* and wheat. The phylogenetic analysis of plant PLCPs was constructed by utilising the large number of sequences which have been made publicly available. The objectives of the present study are:

1. To identify all PLCPs in both *Arabidopsis* and wheat using public databases and determine whether there is literature evidence of their involvement in plant development.
2. To generate up-to-date phylogenetic trees for PLCPs in *Arabidopsis* and wheat.
3. To attempt to find an ortholog of HvPAP14, a recently identified cysteine protease from barley (*Hordeum vulgare*) chloroplasts, in *Arabidopsis*.

7.2 Results

7.2.1 Phylogenetic analysis of cysteine protease in *Arabidopsis*

About 280 cysteine proteases were identified in the *Arabidopsis* data bank. Entire amino acid sequences were used to determine phylogenetic relationships in most cases. From these alignments, circular phylogenetic trees were constructed using iTol (Figure 7.1). Figure 7.1 shows several sequences that are designated either as “putative” or “probable” candidates. Many such sequences do not have well-defined activities. In some cases, the identified fragments appeared to contain almost full sequences. The sequences were clustered into five families: C1, C2, C13, C14, C48 and C54. The majority of the sequences were identified as C1 proteases.

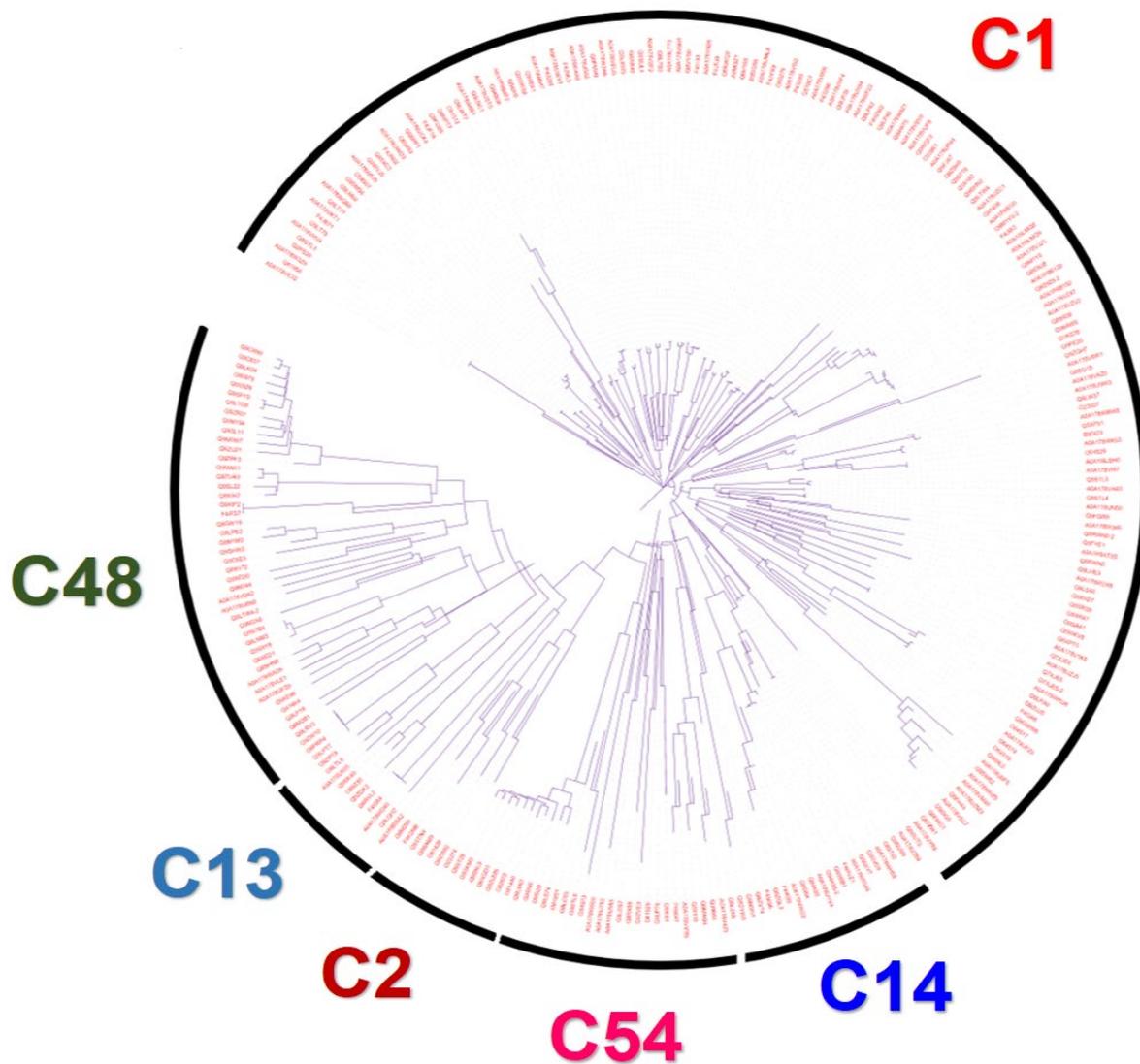


Figure 7.1: Phylogenetic tree providing an overview of the cysteine protease family of *Arabidopsis* in which 280 cysteine protease sequences have clustered into five families: C1, C2, C13, C14, C48 and C54. The majority of the sequences were identified as C1 proteases.

7.2.2 Phylogenetic analysis of well-characterized cysteine proteases in

Arabidopsis

To enable a detailed examination of the evolution of cysteine proteases in *Arabidopsis*, specific sequences were selected based on a UniProt annotation score of 3-5 out of 5, as described in the protocol (Boutet et al., 2016). The UniProt annotation score provides information such as protein properties and function, enzyme regulation and catalytic activity of the more well-characterized proteins (Boutet et al., 2016). In addition, it also highlights sequence similarities that suggest which family a protein belongs (Boutet et al., 2016). Sixty-five cysteine proteases belonging to five families: C1, C2, C13, C14, C48 and C54 were found to be well-characterized proteins (Figure 7.2). A small number of these cysteine proteases have been identified and characterised previously (Figure 7.2).

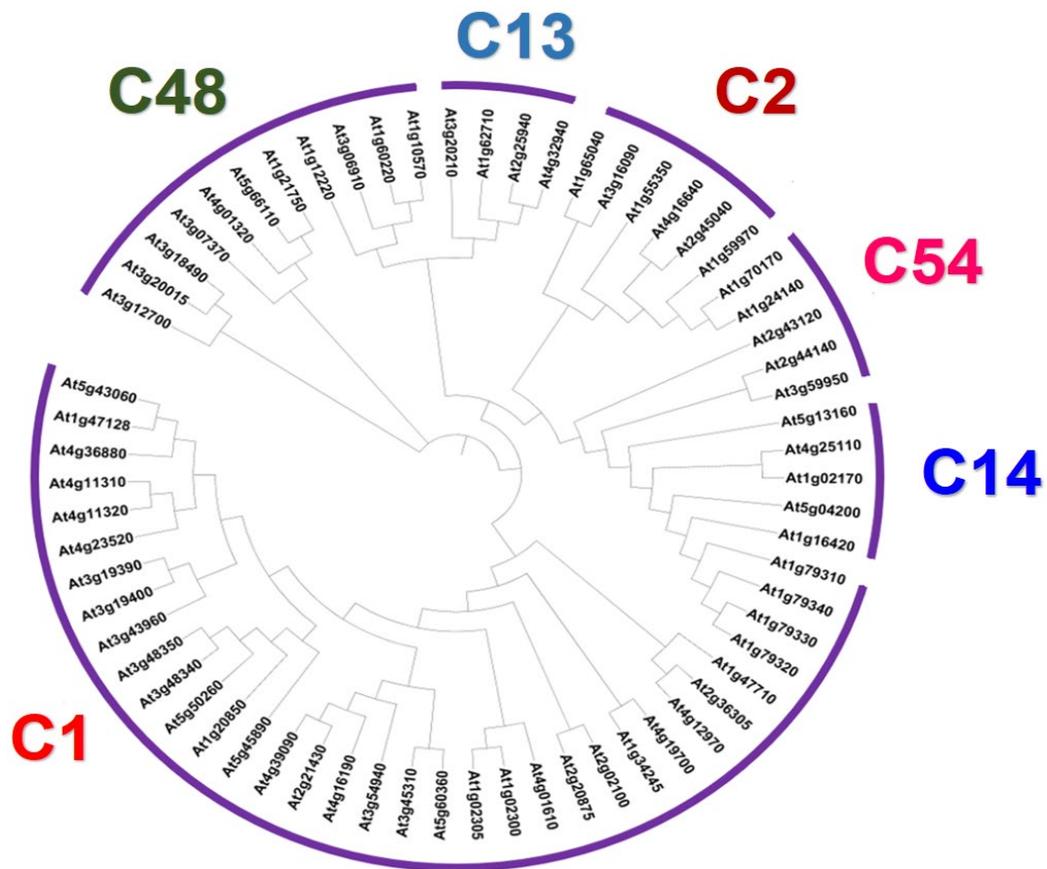


Figure 7.2: Phylogenetic overview of the well-characterized cysteine proteases in Arabidopsis, in which 65 cysteine protease sequences have clustered into five families: C1, C2, C13, C14, C48 and C54.

7.2.3 Phylogeny of selected representative cysteine proteases from different subfamilies of C1A encoded by Arabidopsis

A small number of protein sequences was used here to characterise the features of C1A proteases. However, future studies would benefit from the use of more sequences. In total, 13 cysteine proteases were selected from different subfamilies of the C1A proteases that were identified in the *Arabidopsis* genome (Table 7.1). Their sub-cellular localizations are presented in Table 7.1. These sequences were aligned and a phylogenetic tree for C1A proteases was generated to determine the relationships between each subfamily and its functions (Figure 7.3). The sequences of 13 Cysteine proteases can be divided into four diverse subgroups according to their enzymatic functions (Figure 7.3). The first group is comprised of proteases involved in senescence and/or programmed cell death (Figure 7.3: Group 1). The second group contains KDEL-tailed cysteine endopeptidases, which are involved in developmental cell death (Figure 7.3: Group 2). The third group, which is very diverse, consists of proteases that are involved in germination and/or secondary growth (Figure 7.3: Group 3). Some of these are present in the vascular tissues and are also involved in developmental cell death. The cysteine protease known as SAG12, which is a well-characterised marker of leaf senescence, is the only member of the fourth group that is known to have specific functions in this process (Figure 7.3: Group 4).

Table 7-1: Selected representative cysteine proteases from different subfamilies of C1A involved in different stages of growth in *Arabidopsis*, showing their localization in plant cells. These proteases are located in vacuoles, plasma membrane and endoplasmic reticulum

No.	Gene Name	Protein Name	Location	Reference
1	RD21A	Cysteine proteinase RD21A	Golgi apparatus / Vacuole	(Hayashi et al., 2001; Gu et al., 2012)
2	GCP1	Germination-specific cysteine protease 1	Lysosome / Vacuole	(Tsuji et al., 2013; Ernest, 2015)
3	XCP2	Cysteine protease XCP2	Plasma membrane / Vacuole	(Zhao et al., 2000)
4	XCP1	Cysteine protease XCP1	Plasma membrane / Vacuole	(Funk et al., 2002)
5	RD19B	Probable cysteine protease RD19B	Vacuole	(Bernoux et al., 2008)
6	RD19C	Probable cysteine protease RD19C	Vacuole	(Bernoux et al., 2008)
7	SAG12	Senescence-specific cysteine protease SAG12	Vacuole	(Lohman et al., 1994)
8	CEP1	KDEL-tailed cysteine endopeptidase CEP1	Vacuole / Endoplasmic reticulum	(Höwing et al., 2014; Zhang, D. et al., 2014)
9	CEP2	KDEL-tailed cysteine endopeptidase CEP2	Endoplasmic reticulum	(Hierl et al., 2014)
10	CEP3	KDEL-tailed cysteine endopeptidase CEP3	Endoplasmic reticulum	(Hierl et al., 2014)
11	XBCP3	Papain-like cysteine peptidase	Lysosome / Vacuole	(Beers et al., 2004)
12	RD19A	Cysteine protease RD19A	Vacuole / Nucleus	(Koizumi et al., 1993)
13	ALEU	Thiol protease aleurain	Vacuole	(Ahmed et al., 2000)

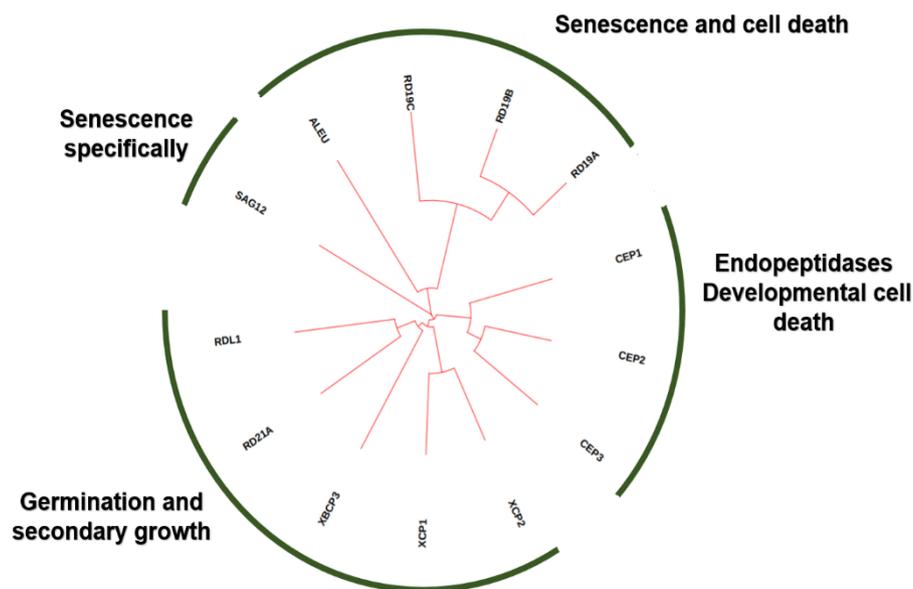


Figure 7.3: Phylogenetic tree for 13 cysteine proteases selected from different subfamilies of C1A in *Arabidopsis*. The figure shows that 13 cysteine proteases sequences have clustered into four groups based on their enzymatic functions.

7.2.4 Homology between protein sequences in Arabidopsis and the barley

PAP14 cysteine protease

A recent study by Frank et al. (2019) reported that the barley cysteine protease called PAP14 is a chloroplast protein and that it is involved in the degradation of chloroplast proteins. To examine the potential homology of HvPAP14 with Arabidopsis cysteine proteases, searches of selected amino acid sequences of *Arabidopsis* cysteine proteases were performed on the basis of homology to the barley PAP14. This analysis highlighted differences as well as similarities between the sequences (Table 7.2). The greatest degree of identity and similarity was found between HvPAP14 and the CEP group. Group 2 contains cysteine proteases that are involved in developmental cell death. All sequences were included in the phylogenetic tree that was constructed using iTol. This analysis showed that the sequences of the CEP group, which are clustered together, seem to be more related to HvPAP14 than to other sequences (Figure 7.4).

Table 7-2: Degree of similarity between selected representative cysteine proteases from different subfamilies of C1A in Arabidopsis and HvPAP14, a recently identified cysteine protease in a *Hordeum vulgare* chloroplast, showing a high similarity in the CEP group.

Cysteine protease	Similarity to APA14	Cysteine protease	Similarity to APA14
RD21A	49.8%	CEP1	74.0%
RDL1	61.5%	CEP2	68.3%
XCP2	63.4%	CEP3	66.5%
XCP1	61.5%	XBCP3	46.9%
RD19B	41.8%	RD19A	42.5%
RD19C	47.0%	ALEU	50.8%
SAG12	57.5%		

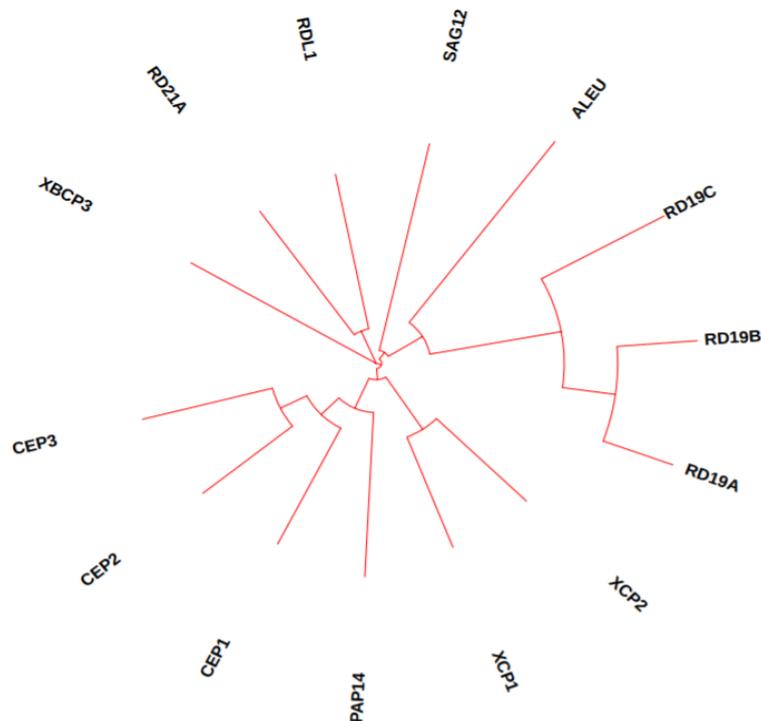
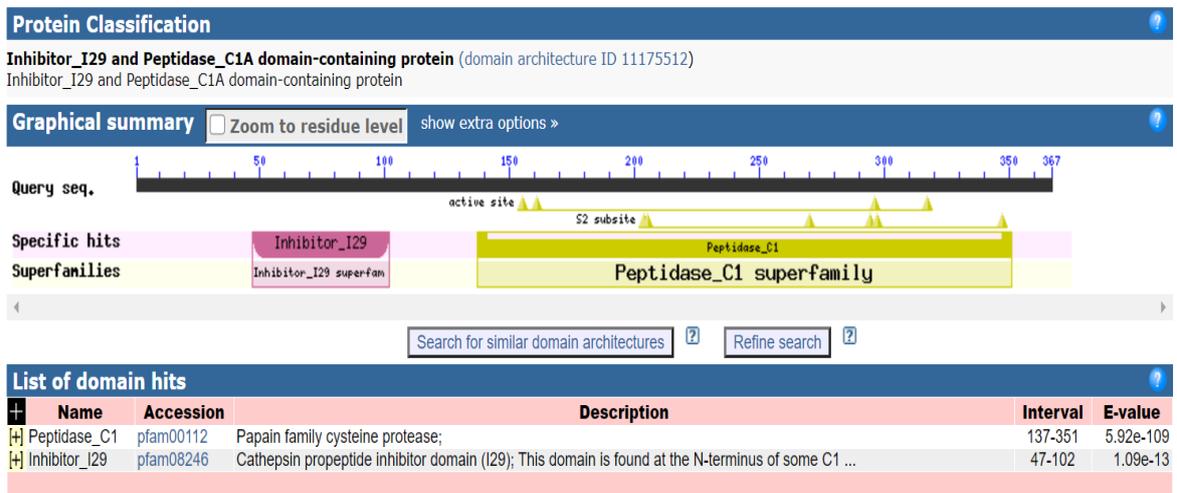


Figure 7.4: Phylogenetic tree showing the relationships between 13 Arabidopsis cysteine proteases and HvPAP14, particularly the clustering of HvPAP14 to the CEP group.

To further investigate the relationships between HvPAP14 and the CEP group, conserved domains were compared between HvPAP14 and CEP1 using the Conserved Domain Database. This analysis showed that HvPAP14 and CEP contain the same domains, namely Peptidase_C1, a Papain family cysteine protease, and Inhibitor_I29, a cathepsin propeptide inhibitor domain (I29). The longer domain is the peptidase_C1 superfamily, as shown in Figure 7.5.

A PAP14



B CEP1

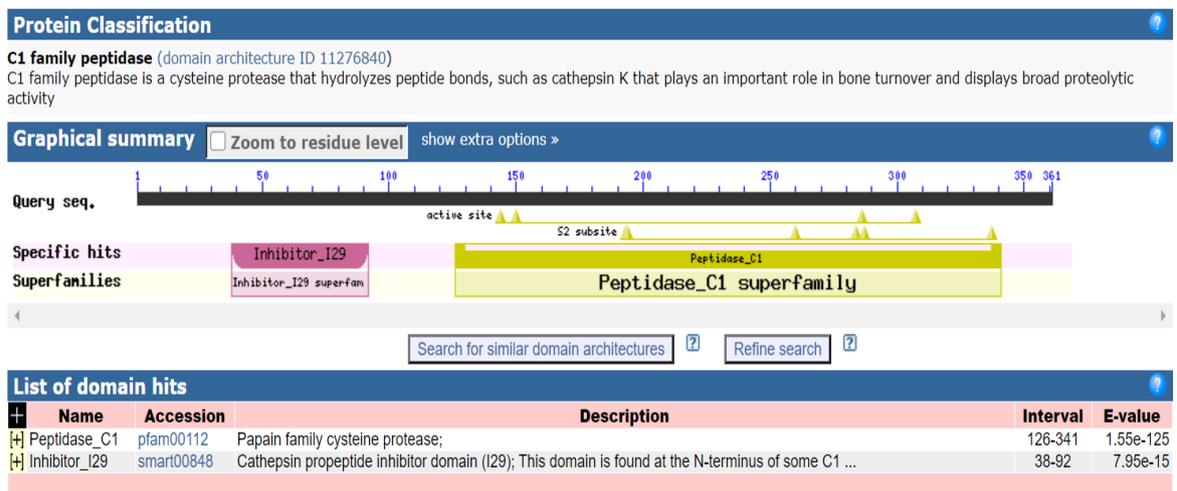


Figure 7.5: A comparison of conserved protein domains in HvPAP14 (A) and CEP1 (B), showing the peptidase_C1 superfamily domain (yellow) and the Inhibitor_I29 domain (pink).

To predict the 3D structure of both HvPAP14 and CEP1 proteins, Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=help>) was used. The prediction of 3D structure is primarily based on the similarity of protein sequences and their domains (Kelley et al., 2015). The structure of these proteins was predicted with 100% confidence and 89% query coverage to the template cysteine protease PDB, c5egwA. Both proteins showed a 3D structure in rainbow colours starting from the N terminal to the C terminal that contains 6 α -helix domains and 9 β -sheets that form the substrate-binding cleft, with the catalytic triad Cys-His-Asn located in between these structures (Figure 7.6). As Figure 7.6 clearly shows, the 3D structure of the CEP1 protein is similar to that of the HvPAP14 protein.

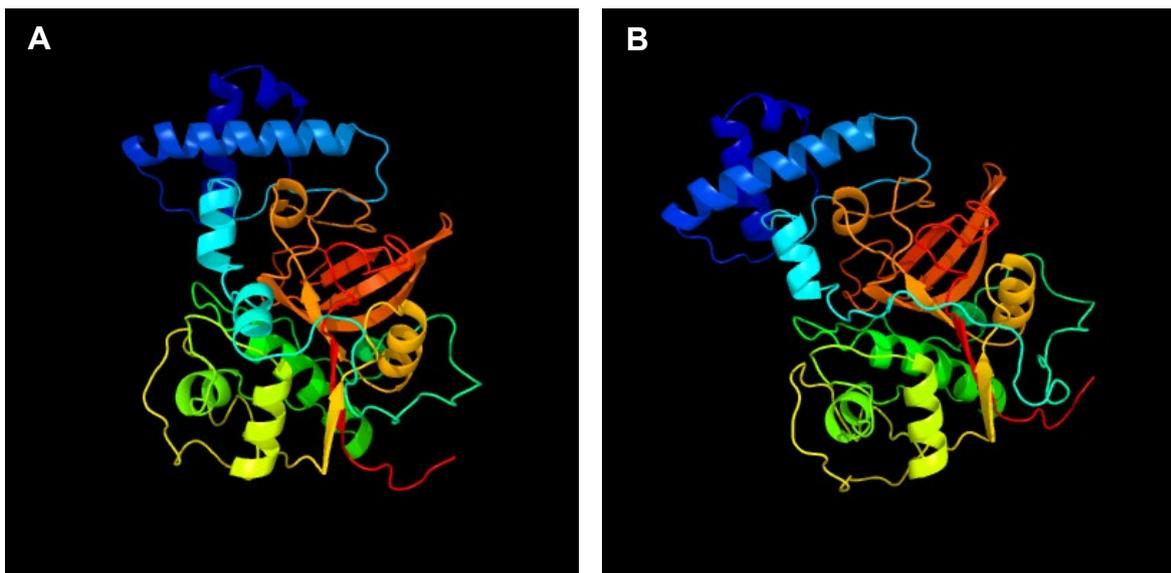


Figure 7.6: Prediction of 3D structure of **(A)** HvPAP14 and **(B)** CEP1 proteins using phyre2. The 3D structure of HvPAP14 and CEP1 proteins are modelled on a sequence of 325 and 323 residues, respectively. The proteins are presented in rainbow colours from the N terminal to the C terminal.

7.2.5 Phylogenetic analysis of cysteine protease in wheat

The cysteine proteases from wheat have not yet been well characterized (Botha et al., 2017). A total of 431 CA cysteine protease sequences were identified in *T. aestivum* obtained from the UniProt website and the Ensembl genomes database, in order to gain new insights into the classification of the CA subfamilies in wheat. The wheat protein sequences were aligned and used to produce the circular phylogenetic tree shown in Figure 7.7. The majority of cysteine proteases in the *T. aestivum* genome, consisting of around 236 members, belong to the C1 family. This family is largely represented by the C1A subfamily of papain-like cysteine proteases (PLCPs) according to the MEROPS database. Two further families, C48 and C2, were identified in the wheat genome. These analysis provides an overview of the cysteine proteases in wheat, which indicates that most members of these families have either not been well characterized or they have not been characterized at all. These data were then used to look for homology between the wheat cysteine proteases and those in other species such as Arabidopsis, with the aim of identifying as many of them as possible. The sequences and classification data will be available to use in Christine Foyer's lab.

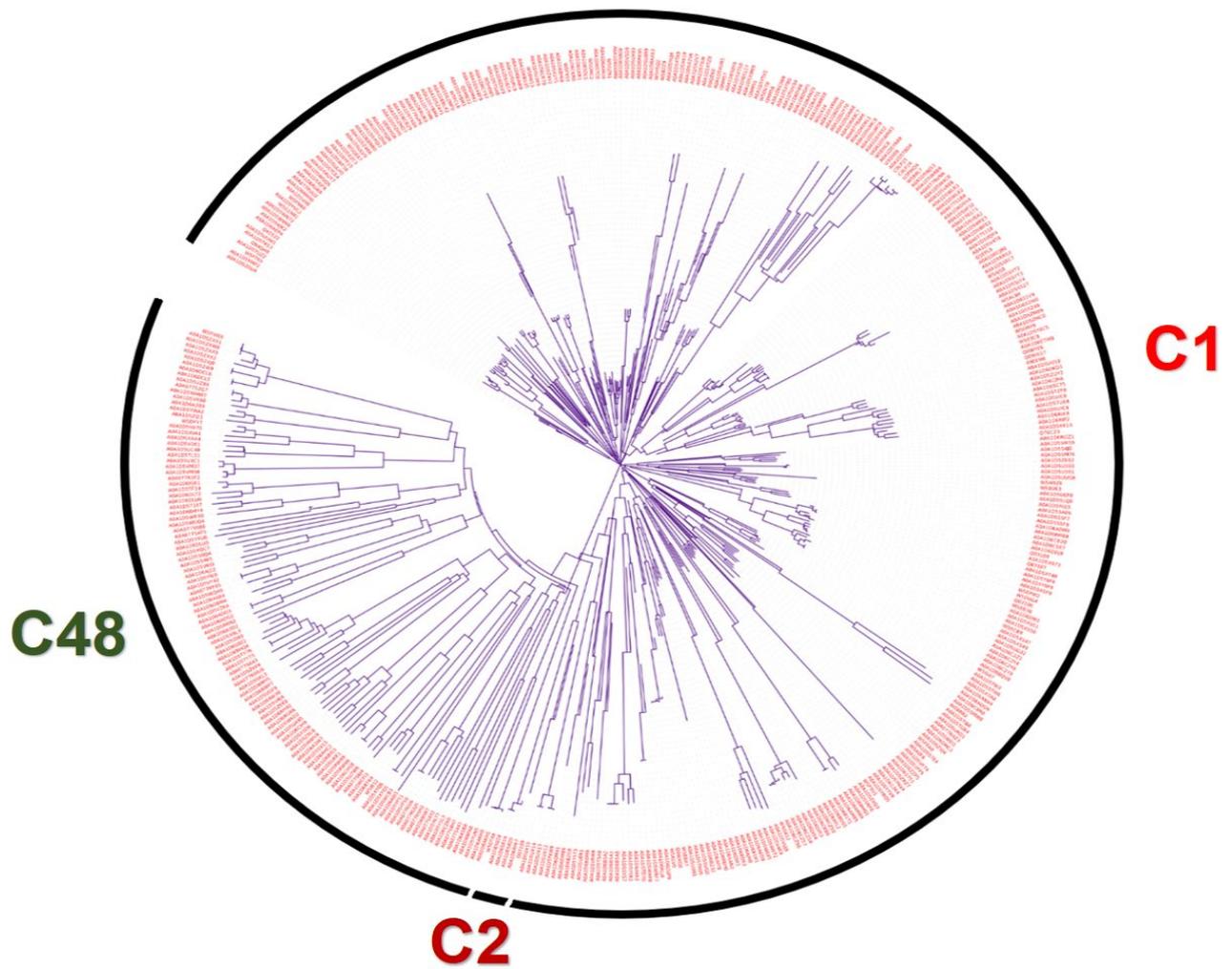


Figure 7.7: Phylogenetic tree providing an overview of the cysteine protease family in wheat genome. The figure shows that 431 cysteine protease sequences have clustered into three families: C1, C2 and C48. The majority of the sequences were identified as C1 proteases.

7.3 Discussion

This chapter describes the first steps towards a comprehensive analysis of plant PLCPs. Two plant species that are available from databanks were used for the clustering analysis and classification into different subfamilies. Phylogenetic analysis identified the cysteine protease family members encoded in the *Arabidopsis* and wheat genomes. Moreover, this analysis used selected representative cysteine proteases from different *Arabidopsis* C1A subfamilies to establish the relationships between the sequences and their functions. Finally, the homological relationships between AtCEP1 and HvPAP14 were investigated. The data presented in this chapter provide a foundation for the development of a framework, through which the characteristics and functions of these gene families can be investigated in the future.

There is a very difference in the sizes of the *Arabidopsis* and wheat (125 Mb and 1700 Mb respectively) genomes (Kaul et al., 2000; Thind et al., 2018). Phylogenetic trees were constructed for both species so that the cysteine protease families in *Arabidopsis* and wheat could be investigated. This analysis provides new insights into the relationships between the different cysteine protease sequences and illustrates the relationships between the CA cysteine proteases in *Arabidopsis* and, albeit less clearly, those in wheat (280 and 431 sequences respectively). The analysis of wheat protein sequences produced a phylogenetic tree that did not allow the clustering of sequences into distinct branches for the different proposed families. Divergence into different groups may have occurred early in evolution (Beers et al., 2004).

PLCPs have a number of specific features. For example, they have a signal peptide that is essential for their transport to the apoplast. Furthermore, the C1-protease domain has an inhibitory pro-domain which is detached upon post-translational activation (Groves et al., 1996). Moreover, the prodomain contains the ERFNIN motif,

which provides structure in a number of PLCP subfamilies (Karrer et al., 1993). In addition, the active site of the protease domain contains an N-terminal Gln and the catalytic triad Cys-His-Asn (Zou et al., 2018).

The activities of PLCPs are likely to be repressed by cystatins, a set of proteins that inhibit cysteine proteases from the papain C1A family through tight binding and reversible interactions (Otto and Schirmeister, 1997; Grzonka et al., 2001). Cystatins are exosite-binding inhibitors i.e. they bind next to the active site, preventing the substrate from accessing the enzyme except through direct interactions with the catalytic centre (Bode and Huber, 2000). Cystatins avoid proteolysis and remain intact and unchanged while attached to the enzyme because they bind in a substrate-like fashion but facing away from the active site of the enzyme (Stubbs et al., 1990).

The data presented in this chapter show that *Arabidopsis* CEP1 is the closest ortholog of the barley HvPAP14 protein. AtCEP1 carries a C-terminal KDEL motif, which leads to localisation in the endoplasmic reticulum (Zou et al., 2018). A putative cleavage site that results in the loss of the KDEL motif has been identified, suggesting that AtCEP1 may be active in the cytosol, i.e. outside the ER (Höwing et al., 2014; Höwing et al., 2017). The barley PLCP, HvPAP14, is located in the ER, as well as vesicular bodies and the chloroplasts, where it is closely associated with the thylakoid membranes (Frank et al., 2019). This enzyme may be responsible for the cleavage of the large subunit of Rubisco, particularly during leaf senescence (Frank et al., 2019).

This study was carried out in part to identify the putative targets for OC-I, which is a cysteine protease inhibitor belonging to the cystatin family. The results presented here suggest that, like HVPAP14, CEP1 may be located in the chloroplasts. Moreover, OC-I might inhibit CEP1 activity. Taken together, the orthology relationships between

CEP1 and HvPAP14 are interesting and reveal new information about potential cysteine proteases that could be located in *Arabidopsis* chloroplasts, and hence be possible targets for OC-I. Several techniques could be used to identify possible protein targets for OC-I, including anti-tag antibodies. These possibilities will be discussed in the next chapter, within the context of opportunities for future research (Section 8.3).

Chapter 8 : General Discussion

Plants contain large numbers of proteases that perform a wide range of functions, including the regulation of metabolism, enzyme activation, transcription factor cleavage and the removal of membrane receptors. The *in vivo* activity of proteases is, in many cases, controlled by endogenous protease inhibitors. Cysteine protease activity can be regulated by the cystatin superfamily proteins. This superfamily is comprised of four families, three of which are cystatins found only in animals. The fourth family are the phytocystatins, which are plant-specific cysteine protease inhibitors (Barrett, 1987). Classification within this family is determined by the presence of disulphide bonds, signal peptides, molecular mass, and sequence homologies (Turk and Bode, 1991). Early studies on phytocystatins focussed primarily on their use as pest control agents (Schlütera et al., 2009). The expression of protease inhibitors improved plant responses to herbivores such as beetles and nematodes (Liang et al., 1991; Davies et al., 2015). Although phytocystatins have been implicated in the control of many important plant processes such as development and stress tolerance, the precise functions of many of these proteins remains poorly characterized. This is particularly important because genetic engineering can be used to alter the structure of phytocystatins to improve binding ability to specific cysteine protease targets. This strategy has great potential in crop improvement programs that are designed to improve stress tolerance (Kunert et al., 2015).

The studies reported in this thesis was that targeting OC-I to *Arabidopsis* chloroplasts (as well as the cytosol) would influence plant growth and development. The results presented in thesis add new information concerning the functions of phytocystatins in plants, making a useful contribution to the literature on the topic (Schlütera et al.,

2009). They also highlight putative new roles of cysteine proteases in plants, particularly in functions such as chloroplast to nucleus signalling.

8.1 Plant growth and development

Although oryzacystatin I from *Oryza sativa* L. *japonica* (OC-I) is perhaps the best-characterised phytocystatin, with data obtained from the characterisation of transgenic plants including tobacco (Prins et al., 2008) and soybean (Quain et al., 2014), much remains uncertain concerning the precise roles of the *in vivo* OCI-targets in plant growth and development. Moreover, earlier studies have focussed only transgenic plants with OC-I expression that was not targetted to specific organelles. In the studies described in this thesis, the phenotypes of the wild type plants was compared to transgenic *Arabidopsis* lines that express OC-I either in the cytosol (CYS) or in the chloroplasts (PC). Moreover, the effects of the expression of OC-I on seed properties was studied in a range of OC-I expressing species (*Arabidopsis*, wheat and soybean).

The data presented in this thesis addressed the question of whether OC-I expression in chloroplasts has similar effects on shoot growth and development to expression in the cytosol. Although the three independent CYS lines and three independent PC lines had a smaller rosette diameter than wild-type (WT) plants up the very last stages of development, the CYS and PC rosettes accumulated less biomass than the WT throughout the vegetative growth period. However, the transgenic lines accumulated significantly more biomass than the WT at the later (reproductive) stages of development. In contrast to the CYS lines, which had significantly fewer leaves than the WT throughout rosette development, PC rosettes had fewer leaves only during the vegetative growth phase. Slow vegetative growth has also been reported in transgenic tobacco plants expressing OC-I in cytosol (Prins et al., 2008). Taken together, these data show that OC-I exerts an effect on vegetative and reproductive development, and

that expression in the chloroplasts has different effects to expression in the cytosol. Moreover, both the CYS and PC rosettes flowered significantly later than the WT, the delay in flowering being most marked in the PC plants suggesting that reproductive development was delayed in the OC-I expressing lines because vegetative growth was slower.

The phenotyping data show that expression of OC-I in the chloroplasts has a marked effect on shoot growth and development, suggesting that there are targets for OC-I in the chloroplasts that have functions related to plant growth and development. The OC-I dependent changes in shoot phenotypes were accompanied by changes in the contents of leaf proteins and pigments. In particular, the CYS and PC rosettes accumulated more leaf protein than the WT and differential effects were observed depending on the intracellular localization of OC-I. The CYS lines had less chlorophyll and carotenoid pigments than the WT, particularly in the later stages of development. In contrast, the PC rosettes accumulated more leaf pigments than the WT at the later stages of leaf development. These results provide further evidence that OC-I is inhibiting different protease targets in the chloroplasts to those in the cytosol. Moreover, the OC-I targets in the chloroplasts regulate the pathways that lead to pigment accumulation. This finding is linked to the observations that the expression of genes that encode photosynthetic proteins is changed in plants expressing OC-I, as well as the responses of photosynthetic gene expression to HL.

Data presented in this thesis showing that photosynthesis is better protected against HL stress in the CYS and PC lines than the WT. This may be explained at least in part by the observation that in contrast to the wild type the Rubisco and D1 proteins accumulate in the OC-I lines under HL stress. Enhanced photosynthetic protein accumulation in these circumstances could lead to improved photosynthetic capacity,

as was observed in the PC lines compared to the WT under HL conditions. The PC lines showed a higher accumulation of Rubisco and D1 proteins than the WT, as well as a greater accumulation of *rbcS* and *psbA* transcripts. These findings are surprising given that no cysteine proteases have been reported in *Arabidopsis* chloroplasts. A phylogenetic analysis was undertaken in these studies to identify the cysteine protease family members encoded in the *Arabidopsis* genomes to determine if there were any proteins with homology to the chloroplast-localised HvPAP14. The *Arabidopsis* CEP1 was found to be the closest orthologue to the barley HvPAP14 protein. However, AtCEP1 carries a C-terminal KDEL motif, which leads to localisation in the endoplasmic reticulum (Zou et al., 2018). While AtCEP1 may also be active in the cytosol (Höwing et al., 2014; Höwing et al., 2017), these findings cannot explain why OC-I expression in the chloroplasts has wide ranging effects on leaf pigment and protein accumulation, as well as chloroplast to nucleus signalling. This means that there is either an as yet unidentified cysteine protease in chloroplasts or other mechanisms that modify chloroplast processes involve cysteine processes that are or become assessable to OC-I.

Chloroplast proteins are degraded through different pathways, as described in Chapter 1 (Section 1.3). These pathways may become more active when plants are exposed to stresses such as high light leading to a more rapid turnover of Rubisco and stromal proteins through for example Rubisco-containing bodies (RCBs) and senescence-associated vacuoles (SAVs) (Figure 8.1). Autophagy is the major system responsible for bulk protein degradation in the vacuole/lysosome (Yoshimoto et al., 2004). If such pathways involve cysteine proteinases, the functions of OC-I in blocking chloroplast protein turnover are likely to be more pronounced under stress conditions leading to a greater accumulation of chloroplast proteins in the OC-I-expressing lines, as was

observed in the studies reported here. Future studies might consider how OC-I expression may affect the autophagy and CVV pathways (Figure 8.1). This could be achieved by crossing the OC-I lines with autophagy mutants, for example, *Arabidopsis atg5* mutants. It is possible to use *in vivo* tracking systems such as GFP-ATG8, in which the expression of ATG8 is linked to a fluorescent marker to enable visualisation of the autophagosomes in the leaves of the WT and OC-I expressing *A. thaliana* lines by confocal microscopy.

Further studies are required to determine how OC-I expression in the chloroplasts leads to increased leaf protein accumulation and associated effects on chloroplast to nucleus signalling. Future investigations are also required to determine the mechanisms that underpin OC-I-dependent regulation of flowering, and particularly how OC-I expression in plastids can exert these effects.

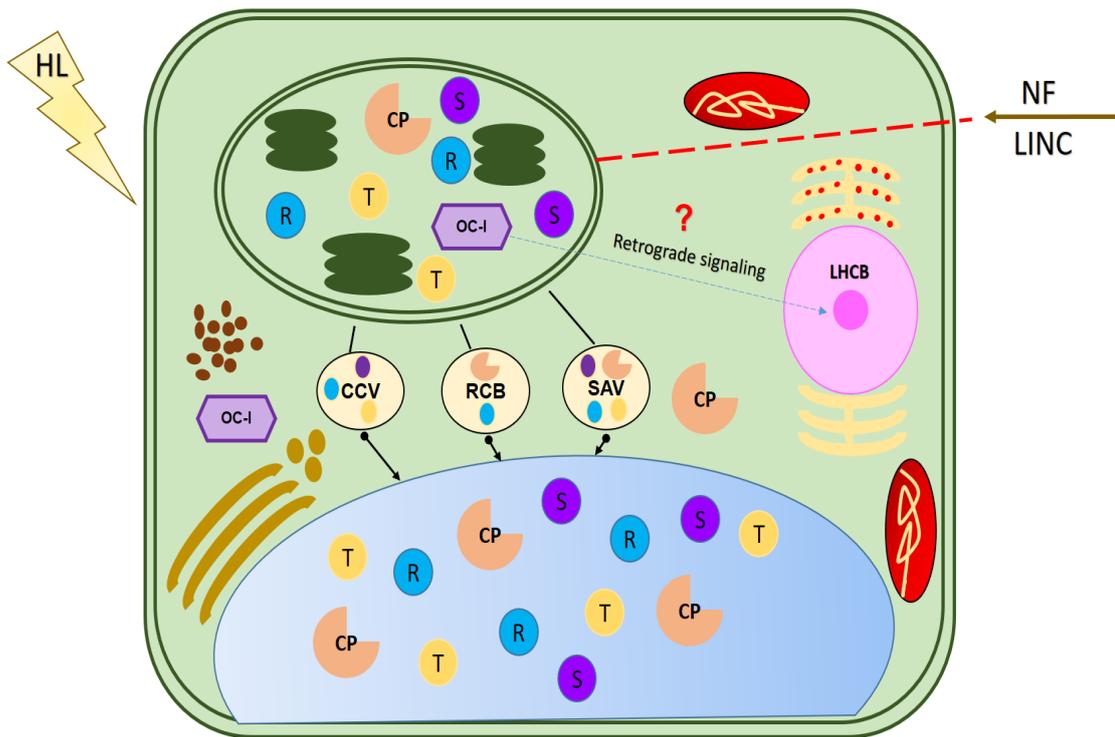


Figure 8.1: Diagram showing the degradation pathways of chloroplast proteins with OC-I located to cytosol and chloroplast, where OC-I = *Oryzacystatin I* (purple); CP = Cysteine protease (orange); T = thylakoid protein (yellow); R = Rubisco (blue); S = stroma protein (dark purple); CVV = senescence-associated vacuole; RCB = Rubisco-containing body; and SAV = senescence-associated vacuoles. Chloroplast proteins can be degraded through three different pathways. **(A)** Chloroplast vesiculation (CV) increases the production of vesicles from the chloroplasts which contain chloroplast proteins, to be targeted for degradation in the vacuole. **(B)** Rubisco is found outside the chloroplast in Rubisco-containing bodies (RCB) and degraded Rubisco products can be found in the vacuoles. **(C)** Stromal proteins are translocated into senescence-associated vacuoles (SAVs) containing cysteine proteases, resulting in the degradation of stroma protein inside the SAVs. It may possible be that OC-I was activated in the RCB and SAVs pathways. The broken lines indicate that OC-I may affect the retrograde signalling, but mechanism remains unknown.

8.2 Seed quality and properties

Wheat is already one of the most commonly grown cereals worldwide. As the global population continues to grow, a 40% increase in wheat production will be needed to meet demand by 2050 (Hall and Richards, 2013). Despite recent developments in biotechnology, improving the relevant traits in wheat remains a challenge. Cysteine protease inhibitors could play an important role in crop improvement strategies, particularly with regards to stress tolerance (Van der Vyver et al., 2003). Transgenic wheat lines overexpressing OC-I were produced and analysed in the present studies, with a particular focus on seed properties (Chapter 5). However, further studies are required to analyse the growth phenotypes and stress tolerance of the independent T4 generation wheat lines expressing OC-I that were prepared here. Future work related to the production of T4 transgenic wheat seeds will be aimed at (i) improving the quality of wheat grains and seed yields and (ii) characterising the shoot and root phenotypes of the transgenic lines in the absence and presence of abiotic stress (e.g. drought) to determine whether the expression of OC-I delays leaf senescence and increases biomass and enhance tolerance to abiotic stress, as has been previously shown in other plant species such as soybean (Quain et al., 2014) .

Cysteine proteases are the main proteases implicated in the germination processes of dicot species, such as *Arabidopsis*, and monocot species, such as Barley (Zhang and Jones, 1995; Szewińska et al., 2016). In particular, they are responsible for the mobilization and degradation of storage proteins (Grudkowska and Zagdanska, 2004; Martinez et al., 2009). Proteolysis is essential for the activation of storage proteins and their hydrolysis in germinating seeds, as well as cereal tissue development. The growth of cereal seeds is controlled by the balance between proteases and their inhibitors. However, this balance can be disturbed if environmental conditions are

detrimental to the plants, as in the case of high humidity, for example, leading to germination of underdeveloped seeds. The expression of OC-I in wheat seeds may help to address this shortfall by improving yield and quality. Moreover, wheat orthologues of OC-I may be useful potential targets in marker-assisted selection for improved traits in wheat breeding programs.

OC-I expression increased seed yields and seed protein contents in all of the three plant species studied in these experiments: *Arabidopsis*, soybean and wheat. These data are surprising because OC-I expression led to significant changes in these important traits. Taken together, these findings indicate that the overexpression of OC-I may be an effective generic approach to crop improvement. The proteomic analysis presented in Chapter 6 provided evidence that the relative abundance of storage proteins was affected in WOC lines compared to the WT. Most of the proteins identified in WOC lines play an important role in providing energy for seed germination as well as determining wheat flour quality, both of which are important to the food industry. Further work is required to determine the targets for OC-I in wheat seeds and to explore the mechanism involved in cysteine protease-dependent increases in seed yields and seed protein contents.

8.3 Identification of papain-like cysteine proteases (PLCPs) in *Arabidopsis* and wheat

A first step towards a comprehensive analysis of plant PLCPs in *Arabidopsis* and wheat was reported in Chapter 7. Wheat (*Triticum aestivum* L.) is allohexaploid with large and complex genomes compared to the *Arabidopsis* genome (Peng et al., 2015). The wheat genome database information is limited, and so the identification of PLCPs in wheat was a challenge. The research reported in this thesis identified 280 and 431 sequences in *Arabidopsis* and wheat, respectively. Bioinformatics approaches will continue to be important in future studies and enable the improved identification of wheat PLCPs orthologues of *Arabidopsis*.

The data presented in Chapter 7 also provide new information about the CEP1 cysteine protease. Further studies are required to determine if CEP1 is located in *Arabidopsis* chloroplasts, as well as other locations. However, in the absence of a chloroplast targeting sequence, chloroplast localisation seems highly unlikely. Nevertheless, these studies strongly suggest that chloroplast proteins are targets for OC-I. It is therefore important to identify OC-I targets and characterize the roles of these proteins in chloroplast functions such as protein turnover and Rubisco degradation in particular. The stock of anti-OC-I antibodies has run out in the Christine Foyer's lab. Therefore, more antibodies need to be made for future experiments designed to identify the protein targets for OC-I for example using anti-OC-I antibodies to fish for interacting proteins. In such experiments, cytosolic and chloroplastic OC-I sequences could be cloned in frame with a tag sequence such as Human influenza hemagglutinin (HA) or green fluorescent protein (GFP) using an immunoprecipitation (IP) technique to isolate interacting proteins with an anti-HA or anti-GFP antibody that are commercially available. Although anti-tag antibodies have the advantage of being

specific for a tag sequence that is not native to plants, the tag can alter the structure and behaviour of the bait protein. Nevertheless, HA and GFP can provide positive outcomes, particularly the short version split GFP and constructs, which can be used to transform the plant or cell culture. It is possible to extract proteins from transgenic plants that express the OCI-tag in plastids or cytosol. Differential centrifugation and a Percoll gradient can be used to purify the protein extraction from plastids and cytosol. Anti-tag antibodies are usually bound to magnetic beads coated with antibodies. Protein extracts can be either mixed with these beads or passed through the column; after repeating this step a number of times, the proteins which have become bound to the column or beads can be extracted. Candidates can then be recognised by mass spectrometry. Further validation of the OC-I interactome could involve the use of a range of techniques such as the two-hybrid system or split-YFP.

8.4 Conclusion and perspective

This thesis has provided new evidence of the importance of cysteine proteinases and their inhibitors such as OC-I in the control of plant growth and development. The findings reported in this thesis not only provide corroborative support for previous reports on the roles of phytocystatins in improving crop yield, but they also provide new evidence for the presence of OC-I targets in the chloroplasts. The comparison of the effects of the expression of OC-I in either the cytosol or chloroplasts, shows that the intracellular compartmentation of OC-I is important in determining the effects on plant growth and development. The data presented in this thesis concerning the expression of OC-I in the cytosol are consistent with previous reports on the effects of similar expression of OC-I in tobacco (Van der Vyver et al., 2003) and soybean (Quain et al., 2014). Moreover, the data presented here demonstrate that overexpression of the OC-I exerts a strong influence on plant performance and grain quality. These findings are important for future plant breeding strategies that are designed to improve the sustainability of yield and quality traits. These findings are particularly important in relation to producing more resilient crop plants that are better able to cope with climate change-induced abiotic stresses. Moreover, they show that phytocystatins such as OC-I are important targets for improving stress tolerance as well as seed production and quality traits. It is noteworthy that phytocystatins are considered to be safe to use for crop improvement as they have no negative effects on humans, when they are present in food that is part of the diet (Atkinson et al., 2004).

References

- Abe, K., Emori, Y., Kondo, H., Suzuki, K. and Arai, S.** 1987. Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds. *Journal of Biological Chemistry*. **262**(35), pp.16793-16797.
- Abe, K., Kondo, H. and Arai, S.** 1987. Purification and Characterization of a Rice Cysteine Proteinase Inhibitor. *Agricultural and Biological Chemistry*. **51**(10), pp.2763-2768.
- Adam, Z. and Clarke, A.K.** 2002. Cutting edge of chloroplast proteolysis. *Trends in plant science*. **7**(10), pp.451-456.
- Adam, Z., Rudella, A. and van Wijk, K.J.** 2006. Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. *Current opinion in plant biology*. **9**(3), pp.234-240.
- Ahmed, S.U., Rojo, E., Kovaleva, V., Venkataraman, S., Dombrowski, J.E., Matsuoka, K. and Raikhel, N.** 2000. The plant vacuolar sorting receptor AtELP is involved in transport of NH₂-terminal propeptide-containing vacuolar proteins in *Arabidopsis thaliana*. *Journal of Cell Biology*. **149**(7), pp.1335-1344.
- Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., Mohellibi, N., Durand, S., Kimmel, E. and Michotey, C.** 2018. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome biology*. **19**(1), p111.
- Allorent, G., Courtois, F., Chevalier, F. and Lerbs-Mache, S.** 2013. Plastid gene expression during chloroplast differentiation and dedifferentiation into non-photosynthetic plastids during seed formation. *Plant molecular biology*. **82**(1-2), pp.59-70.
- Altenbach, S.B., Tanaka, C.K. and Seabourn, B.W.** 2014. Silencing of omega-5 gliadins in transgenic wheat eliminates a major source of environmental variability and improves dough mixing properties of flour. *BMC plant biology*. **14**(1), p393.
- Atkinson, H.J., Johnston, K.A. and Robbins, M.** 2004. Prima facie evidence that a phytocystatin for transgenic plant resistance to nematodes is not a toxic risk in the human diet. *The Journal of nutrition*. **134**(2), pp.431-434.
- Barrett.** 1987. The cystatins: a new class of peptidase inhibitors. *Trends in Biochemical Sciences*. **12**, pp.193-196.
- Bartels, D. and Sunkar, R.** 2005. Drought and salt tolerance in plants. *Critical reviews in plant sciences*. **24**(1), pp.23-58.
- Bastaki, N.K. and Cullis, C.A.** 2015. The instability of the flax element LIS-1 in transgenic *Arabidopsis thaliana*. *Research and Reports in Biology*. **6**, p89.
- Bayer, R.G., Stael, S., Csaszar, E. and Teige, M.** 2011. Mining the soluble chloroplast proteome by affinity chromatography. *Proteomics*. **11**(7), pp.1287-1299.
- Beers, E.P., Jones, A.M. and Dickerman, A.W.** 2004. The S8 serine, C1A cysteine and A1 aspartic protease families in *Arabidopsis*. *Phytochemistry*. **65**(1), pp.43-58.
- Beers, E.P., Woffenden, B.J. and Zhao, C.** 2000. Plant proteolytic enzymes: possible roles during programmed cell death. *Plant molecular biology*. **44**(3), pp.399-415.

- Belenghi, B., Acconcia, F., Trovato, M., Perazzoli, M., Bocedi, A., Polticelli, F., Ascenzi, P. and Delledonne, M.** 2003. AtCYS1, a cystatin from *Arabidopsis thaliana*, suppresses hypersensitive cell death. *The FEBS Journal*. **270**(12), pp.2593-2604.
- Benchabane, M., Goulet, C., Rivard, D., Faye, L., Gomord, V. and Michaud, D.** 2008. Preventing unintended proteolysis in plant protein biofactories. *Plant biotechnology journal*. **6**(7), pp.633-648.
- Benchabane, M., Schluter, U., Vorster, J., Goulet, M.C. and Michaud, D.** 2010. Plant cystatins. *Biochimie*. **92**(11), pp.1657-1666.
- Bernoux, M., Timmers, T., Jauneau, A., Briere, C., de Wit, P.J., Marco, Y. and Deslandes, L.** 2008. RD19, an *Arabidopsis* cysteine protease required for RRS1-R-mediated resistance, is relocalized to the nucleus by the *Ralstonia solanacearum* PopP2 effector. *The Plant Cell*. **20**(8), pp.2252-2264.
- Beyene, G., Foyer, C.H. and Kunert, K.J.** 2006. Two new cysteine proteinases with specific expression patterns in mature and senescent tobacco (*Nicotiana tabacum* L.) leaves. *Journal of Experimental Botany*. **57**(6), pp.1431-1443.
- Bleecker, A.B. and Kende, H.** 2000. Ethylene: a gaseous signal molecule in plants. *Annual review of cell and developmental biology*. **16**(1), pp.1-18.
- Bode, W. and Huber, R.** 2000. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. **1477**(1-2), pp.241-252.
- Botha, A.M., Kunert, K.J. and Cullis, C.A.** 2017. Cysteine proteases and wheat (*Triticum aestivum* L) under drought: a still greatly unexplored association. *Plant, cell & environment*. **40**(9), pp.1679-1690.
- Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A.J., Poux, S., Bougueleret, L. and Xenarios, I.** 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Plant Bioinformatics. Springer*, pp.23-54.
- Boye, J., Zare, F. and Pletch, A.** 2010. Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food research international*. **43**(2), pp.414-431.
- Briefs, I.** 2017. Global status of commercialized biotech/GM crops in 2017: Biotech crop adoption surges as economic benefits accumulate in 22 years. [Online]. (ISAAA Brief No. 53. ISAAA: Ithaca, NY.).
- Brzin, J. and Kidrič, M.** 1996. Proteinases and their inhibitors in plants: role in normal growth and in response to various stress conditions. *Biotechnology and Genetic Engineering Reviews*. **13**(1), pp.421-468.
- Buet, A., Costa, M.L., Martínez, D.E. and Guamet, J.J.** 2019. Chloroplast protein degradation in senescing leaves: Proteases and lytic compartments. *Frontiers in Plant Science*. **10**,p747.
- Cai, Y.-m. and Gallois, P.** 2015. Programmed cell death regulation by plant proteases with caspase-like activity. *Plant programmed cell death. Springer*, pp.191-202.
- Callis, J.** 1995. Regulation of Protein Degradation. *Plant Cell*. **7**(7), pp.845-857.

- Cambra, I., Martínez, M., Dáder, B., González-Melendi, P., Gandullo, J., Santamaría, M.E. and Díaz, I.** 2012. A cathepsin F-like peptidase involved in barley grain protein mobilization, HvPap-1, is modulated by its own propeptide and by cystatins. *Journal of Experimental Botany*. **63**(12), pp.4615-4629.
- Carrillo, L., Martínez, M., Ramessar, K., Cambra, I., Castañera, P., Ortego, F. and Díaz, I.** 2011. Expression of a barley cystatin gene in maize enhances resistance against phytophagous mites by altering their cysteine-proteases. *Plant cell reports*. **30**(1), pp.101-112.
- Carrión, C.A., Costa, M.L., Martínez, D.E., Mohr, C., Humbeck, K. and Guamet, J.J.** 2013. In vivo inhibition of cysteine proteases provides evidence for the involvement of 'senescence-associated vacuoles' in chloroplast protein degradation during dark-induced senescence of tobacco leaves. *Journal of Experimental Botany*. **64**(16), pp.4967-4980.
- Chaves, M.M., Flexas, J. and Pinheiro, C.** 2009. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany*. **103**(4), pp.551-560.
- Chen, D., Chai, S., McIntyre, C.L. and Xue, G.-P.** 2018. Overexpression of a predominantly root-expressed NAC transcription factor in wheat roots enhances root length, biomass and drought tolerance. *Plant cell reports*. **37**(2), pp.225-237.
- Chen, H. and Lin, Y.** 2013. Promise and issues of genetically modified crops. *Current Opinion in Plant Biology*. **16**(2), pp.255-260.
- Chichkova, N.V., Kim, S.H., Titova, E.S., Kalkum, M., Morozov, V.S., Rubtsov, Y.P., Kalinina, N.O., Taliansky, M.E. and Vartapetian, A.** 2004. A plant caspase-like protease activated during the hypersensitive response. *The Plant Cell*. **16**(1), pp.157-171.
- Chileh, T., Esteban-Garcia, B., Alonso, D.L.p. and Garcia-Maroto, F.** 2010. Characterization of the 11S globulin gene family in the castor plant *Ricinus communis* L. *Journal of agricultural and food chemistry*. **58**(1), pp.272-281.
- Christoff, A.P. and Margis, R.** 2014. The diversity of rice phytocystatins. *Molecular genetics and genomics*. **289**(6), pp.1321-1330.
- Christova, P.K., Christov, N.K. and Imai, R.** 2006. A cold inducible multidomain cystatin from winter wheat inhibits growth of the snow mold fungus, *Microdochium nivale*. *Planta*. **223**(6), p1207.
- Ciardi, J.A., Tieman, D.M., Lund, S.T., Jones, J.B., Stall, R.E. and Klee, H.J.** 2000. Response to *Xanthomonas campestris* pv. *vesicatoria* in tomato involves regulation of ethylene receptor gene expression. *Plant Physiology*. **123**(1), pp.81-92.
- Clough, S. and Bent, A.** 2000. Vapor-phase sterilization of *Arabidopsis* seed. [Online]. [Accessed 25 August 2017]. Available from: <https://plantpath.wisc.edu/vapor-phase-sterilization-of-arabidopsis-seed/>
- Clough, S.J. and Bent, A.F.** 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The plant journal*. **16**(6), pp.735-743.
- Comadira, G.** 2015. Regulation of senescence in *Arabidopsis thaliana* and barley. thesis, University of Leeds.
- Cooper, J.W.** 2016. The characterisation of low temperature tolerance in legumes. thesis, University of Leeds.

- Davies, L.J., Zhang, L. and Elling, A.A.** 2015. The *Arabidopsis thaliana* papain-like cysteine protease RD21 interacts with a root-knot nematode effector protein. *Nematology*. **17**(6), pp.655-666.
- De Preter, K., Speleman, F., Combaret, V., Lunec, J., Laureys, G., Eussen, B.H., Francotte, N., Board, J., Pearson, A.D. and De Paepe, A.** 2002. Quantification of MYCN, DDX1, and NAG gene copy number in neuroblastoma using a real-time quantitative PCR assay. *Modern Pathology*. **15**(2), pp.159-166.
- Desclos, M., Dubouset, L., Etienne, P., Le Caherec, F., Satoh, H., Bonnefoy, J., Ourry, A. and Avice, J.-C.** 2008. A proteomic profiling approach to reveal a novel role of *Brassica napus* drought 22 kD/water-soluble chlorophyll-binding protein in young leaves during nitrogen remobilization induced by stressful conditions. *Plant Physiology*. **147**(4), pp.1830-1844.
- Díaz-Mendoza, M., Velasco-Arroyo, B., González-Melendi, P., Martínez, M. and Díaz, I.** 2014. C1A cysteine protease–cystatin interactions in leaf senescence. *Journal of Experimental Botany*. **65**(14), pp.3825-3833.
- Díaz-Mendoza, M., Velasco-Arroyo, B., Santamaria, M.E., González-Melendi, P., Martínez, M. and Díaz, I.** 2016. Plant senescence and proteolysis: two processes with one destiny. *Genetics and molecular biology*. **39**(3), pp.329-338.
- Diop, N.N., Kidrič, M., Repellin, A., Gareil, M., d'Arcy-Lameta, A., Pham Thi, A.T. and Zuily-Fodil, Y.** 2004. A multicystatin is induced by drought-stress in cowpea (*Vigna unguiculata* (L.) Walp.) leaves. *FEBS letters*. **577**(3), pp.545-550.
- Dou, S., Wei, J., Cao, Y. and Lan, J.** 2011. Molecular characterization and preliminary functional analysis of cystatin OC-I in rice. *Journal of Food, Agriculture and Environment*. **9**(2), pp.235-239.
- Duan, B., Yang, Y., Lu, Y., Korpelainen, H., Berninger, F. and Li, C.** 2007. Interactions between drought stress, ABA and genotypes in *Picea asperata*. *Journal of Experimental Botany*. **58**, pp.3025-3036.
- Ernest, J.R.** 2015. Expression patterns of the intracellular ion exchangers AtNHX5 and AtNHX6 and their role in subcellular trafficking to the vacuole. thesis, La Trobe University.
- Evans, I., Rus, A., Belanger, E., Kimoto, M. and Brusslan, J.** 2010. Dismantling of *Arabidopsis thaliana* mesophyll cell chloroplasts during natural leaf senescence. *Plant Biology*. **12**(1), pp.1-12.
- Fahad, S., Bajwa, A.A., Nazir, U., Anjum, S.A., Farooq, A., Zohaib, A., Sadia, S., Nasim, W., Adkins, S. and Saud, S.** 2017. Crop production under drought and heat stress: plant responses and management options. *Frontiers in plant science*. **8**, p1147.
- Food and Agriculture Organization of the United Nations.** 2017. The future of food and agriculture—Trends and challenges. *Annual Report*.
- Faye, L., Boulaflous, A., Benchabane, M., Gomord, V. and Michaud, D.** 2005. Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine*. **23**(15), pp.1770-1778.
- Figueiredo, A., Monteiro, F. and Sebastiana, M.** 2014. Subtilisin-like proteases in plant–pathogen recognition and immune priming: a perspective. *Frontiers in plant science*. **5**, p739.

- Fischer, R., Byerlee, D. and Edmeades, G.** 2014. Crop yields and global food security. *ACIAR: Canberra, ACT*. pp.8-11.
- Frank, S., Hollmann, J., Mulisch, M., Matros, A., Carrión, C.C., Mock, H.-P., Hensel, G. and Krupinska, K.** 2019. Barley cysteine protease PAP14 plays a role in degradation of chloroplast proteins. *Journal of Experimental Botany*. **70**(21), pp.6057-6069.
- Fu, D., Uauy, C., Blechl, A. and Dubcovsky, J.** 2007. RNA interference for wheat functional gene analysis. *Transgenic Research*. **16**(6), pp.689-701.
- Funk, V., Kositsup, B., Zhao, C. and Beers, E.P.** 2002. The Arabidopsis xylem peptidase XCP1 is a tracheary element vacuolar protein that may be a papain ortholog. *Plant Physiology*. **128**(1), pp.84-94.
- Gadaleta, A., Giancaspro, A., Cardone, M. and Blanco, A.** 2011. Real-time PCR for the detection of precise transgene copy number in durum wheat. *Cellular and Molecular Biology Letters*. **16**(4), pp.652-668.
- Gaddour, K., Vicente-Carbajosa, J., Lara, P., Isabel-Lamoneda, I., Díaz, I. and Carbonero, P.** 2001. A constitutive cystatin-encoding gene from barley (Icy) responds differentially to abiotic stimuli. *Plant molecular biology*. **45**(5), pp.599-608.
- Galland, M., Boutet-Mercey, S., Lounifi, I., Godin, B., Balzergue, S., Grandjean, O., Morin, H., Perreau, F., Debeaujon, I. and Rajjou, L.** 2014. Compartmentation and dynamics of flavone metabolism in dry and germinated rice seeds. *Plant and Cell Physiology*. **55**(9), pp.1646-1659.
- Gil-Humanes, J., Pistón, F., Tollefsen, S., Sollid, L.M. and Barro, F.** 2010. Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. *Proceedings of the National Academy of Sciences*. **107**(39), pp.17023-17028.
- Gilroy, E.M., Hein, I., Van Der Hoorn, R., Boevink, P.C., Venter, E., McLellan, H., Kaffarnik, F., Hrubikova, K., Shaw, J. and Holeva, M.** 2007. Involvement of cathepsin B in the plant disease resistance hypersensitive response. *The Plant Journal*. **52**(1), pp.1-13.
- Groten, K., Dutilleul, C., van Heerden, P.D., Vanacker, H., Bernard, S., Finkemeier, I., Dietz, K.-J. and Foyer, C.H.** 2006. Redox regulation of peroxiredoxin and proteinases by ascorbate and thiols during pea root nodule senescence. *FEBS letters*. **580**(5), pp.1269-1276.
- Groves, M.R., Taylor, M.A., Scott, M., Cummings, N.J., Pickersgill, R.W. and Jenkins, J.A.** 1996. The prosequence of procaricain forms an α -helical domain that prevents access to the substrate-binding cleft. *Structure*. **4**(10), pp.1193-1203.
- Grudkowska, M. and Zagdanska, B.** 2004. Multifunctional role of plant cysteine proteinases. *Acta Biochimica Polonica*. pp.609-624.
- Gruis, D., Schulze, J. and Jung, R.** 2004. Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. *The Plant Cell*. **16**(1), pp.270-290.
- Gruis, D.F., Selinger, D.A., Curran, J.M. and Jung, R.** 2002. Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. *The Plant Cell*. **14**(11), pp.2863-2882.
- Grzonka, Z., Jankowska, E., Kasprzykowski, F., Kasprzykowska, R., Lankiewicz, L., Wicz, W., Wiczerzak, E., Ciarkowski, J., Drabik, P. and Janowski, R.** 2001. Structural studies of cysteine proteases and their inhibitors. *Acta Biochimica Polonica*. **48**(1), pp.1-20.

- Gu, C., Shabab, M., Strasser, R., Wolters, P.J., Shindo, T., Niemer, M., Kaschani, F., Mach, L. and van der Hoorn, R.A.** 2012. Post-translational regulation and trafficking of the granulin-containing protease RD21 of *Arabidopsis thaliana*. *PLoS one*. **7**(3), pe32422.
- Gutierrez-Campos, R., Torres-Acosta, J.A., Saucedo-Arias, L.J. and Gomez-Lim, M.A.** 1999. The use of cysteine proteinase inhibitors to engineer resistance against potyviruses in transgenic tobacco plants. *Nature biotechnology*. **17**(12), pp.1223-1226.
- Habib, H. and Fazili, K.M.** 2007. Plant protease inhibitors: a defense strategy in plants. *Biotechnology and Molecular Biology Reviews*. **2**(3), pp.68-85.
- Hall and Richards, R.A.** 2013. Prognosis for genetic improvement of yield potential and water-limited yield of major grain crops. *Field Crops Research*. **143**, pp.18-33.
- Hall, B.G. and Barlow, M.** 2004. Evolution of the serine β -lactamases: past, present and future. *Drug Resistance Updates*. **7**(2), pp.111-123.
- Hara-Nishimura, I., Shimada, T., Hatano, K., Takeuchi, Y. and Nishimura, M.** 1998. Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. *The Plant Cell*. **10**(5), pp.825-836.
- Hara-Nishimura, I., Takeuchi, Y. and Nishimura, M.** 1993. Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. *The Plant Cell*. **5**(11), pp.1651-1659.
- Hatsugai, N., Kuroyanagi, M., Nishimura, M. and Hara-Nishimura, I.** 2006. A cellular suicide strategy of plants: vacuole-mediated cell death. *Apoptosis*. **11**(6), pp.905-911.
- Hatsugai, N., Kuroyanagi, M., Yamada, K., Meshi, T., Tsuda, S., Kondo, M., Nishimura, M. and Hara-Nishimura, I.** 2004. A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. *Science*. **305**(5685), pp.855-858.
- Hayashi, Y., Yamada, K., Shimada, T., Matsushima, R., Nishizawa, N., Nishimura, M. and Hara-Nishimura, I.** 2001. A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of *Arabidopsis*. *Plant and Cell Physiology*. **42**(9), pp.894-899.
- He, Z., Xia, X., Chen, X. and Zhuang, Q.** 2011. Progress of wheat breeding in China and the future perspective. *Acta Agronomica Sinica*. **37**(2), pp.202-215.
- Hierl, G., Höwing, T., Isono, E., Lottspeich, F. and Gietl, C.** 2014. Ex vivo processing for maturation of *Arabidopsis* KDEL-tailed cysteine endopeptidase 2 (AtCEP2) pro-enzyme and its storage in endoplasmic reticulum derived organelles. *Plant molecular biology*. **84**(6), pp.605-620.
- Hinchliffe, A. and Harwood, W.A.** 2019. Agrobacterium-mediated transformation of Barley immature embryos. *Barley*. *Springer*, pp.115-126.
- Hisabori, T., Motohashi, K., Hosoya-Matsuda, N., Ueoka-Nakanishi, H. and Romano, P.G.** 2007. Towards a functional dissection of thioredoxin networks in plant cells. *Photochemistry and Photobiology*. **83**(1), pp.145-151.
- Hong, J.K., Hwang, J.E., Lim, C.J., Yang, K.A., Jin, Z.-L., Kim, C.Y., Koo, J.C., Chung, W.S., Lee, K.O. and Lee, S.Y.** 2007. Over-expression of Chinese cabbage phytoalexin 1 retards seed germination in *Arabidopsis*. *Plant science*. **172**(3), pp.556-563.

- Höwing, T., Dann, M., Hoefle, C., Hückelhoven, R. and Gietl, C.** 2017. Involvement of Arabidopsis thaliana endoplasmic reticulum KDEL-tailed cysteine endopeptidase 1 (AtCEP1) in powdery mildew-induced and AtCPR5-controlled cell death. *PLoS One*. **12**(8), pe0183870.
- Höwing, T., Dann, M., Müller, B., Helm, M., Scholz, S., Schneitz, K., Hammes, U.Z. and Gietl, C.** 2018. The role of KDEL-tailed cysteine endopeptidases of Arabidopsis (AtCEP2 and AtCEP1) in root development. *PloS one*. **13**(12).
- Höwing, T., Huesmann, C., Hoefle, C., Nagel, M.-K., Isono, E., Huckelhoven, R. and Gietl, C.** 2014. Endoplasmic reticulum KDEL-tailed cysteine endopeptidase 1 of Arabidopsis (AtCEP1) is involved in pathogen defense. *Frontiers in plant science*. **5**, p58.
- Huang, Y., Xiao, B. and Xiong, L.** 2007. Characterization of a stress responsive proteinase inhibitor gene with positive effect in improving drought resistance in rice. *Planta*. **226**(1), pp.73-85.
- Humbeck, K., Quast, S. and Krupinska, K.** 1996. Functional and molecular changes in the photosynthetic apparatus during senescence of flag leaves from field-grown barley plants. *Plant, Cell & Environment*. **19**(3), pp.337-344.
- Hwang, J.E., Hong, J.K., Je, J.H., Lee, K.O., Kim, D.Y., Lee, S.Y. and Lim, C.O.** 2009. Regulation of seed germination and seedling growth by an Arabidopsis phytocystatin isoform, AtCYS6. *Plant cell reports*. **28**(11), pp.1623-1632.
- Hwang, J.E., Hong, J.K., Lim, C.J., Chen, H., Je, J., Yang, K.A., Kim, D.Y., Choi, Y.J., Lee, S.Y. and Lim, C.O.** 2010. Distinct expression patterns of two Arabidopsis phytocystatin genes, AtCYS1 and AtCYS2, during development and abiotic stresses. *Plant cell reports*. **29**(8), pp.905-915.
- Inaba, T.** 2010. Bilateral communication between plastid and the nucleus: plastid protein import and plastid-to-nucleus retrograde signaling. *Bioscience, biotechnology, and biochemistry*. pp.1002011852-1002011852.
- Ingram, J. and Bartels, D.** 1996. The molecular basis of dehydration tolerance in plants. *Annual review of plant biology*. **47**(1), pp.377-403.
- Iqbal, N., Masood, A., Khan, M.I.R., Asgher, M., Fatma, M. and Khan, N.A.** 2013. Cross-talk between sulfur assimilation and ethylene signaling in plants. *Plant signaling & behavior*. **8**(1), pe22478.
- Irving, L. and Robinson, D.** 2006. On modelling Rubisco turnover: dynamics and applicability. *New Phytologist*. **170**(2), pp.204-205.
- Ito, K. and Akiyama, Y.** 2005. Cellular functions, mechanism of action, and regulation of FtsH protease. *Annual Review of Microbiology*. **59**, pp.211-231.
- Itzhaki, H., Naveh, L., Lindahl, M., Cook, M. and Adam, Z.** 1998. Identification and characterization of DegP, a serine protease associated with the luminal side of the thylakoid membrane. *Journal of Biological Chemistry*. **273**(12), pp.7094-7098.
- Jacks, T., Hensarling, T. and Yatsu, L.** 1972. Cucurbit seeds: I. Characterizations and uses of oils and proteins. A review. *Economic Botany*. **26**(2), pp.135-141.
- James, M., Masclaux-Daubresse, C., Marmagne, A., Azzopardi, M., Laine, P., Goux, D., Etienne, P. and Trouverie, J.** 2019. A new role for SAG12 cysteine protease in roots of Arabidopsis thaliana. *Frontiers in plant science*. **9**, p1998.

- Järvi, S., Gollan, P.J. and Aro, E.-M.** 2013. Understanding the roles of the thylakoid lumen in photosynthesis regulation. *Frontiers in plant science*. **4**, p434.
- Jiao, Y., Lau, O.S. and Deng, X.W.** 2007. Light-regulated transcriptional networks in higher plants. *Nature Reviews Genetics*. **8**(3), pp.217-230.
- Jinka, R., Ramakrishna, V., Rao, S.K. and Rao, R.P.** 2009. Purification and characterization of cysteine protease from germinating cotyledons of horse gram. *BMC biochemistry*. **10**(1), p28.
- Jolliffe, N.A., Craddock, C.P. and Frigerio, L.** 2005. Pathways for protein transport to seed storage vacuoles. *Biochemical Society Transactions*. **33**(5): pp.1016–1018.
- Kaiserli, E., Paldi, K., O'Donnell, L., Batalov, O., Pedmale, U.V., Nusinow, D.A., Kay, S.A. and Chory, J.** 2015. Integration of light and photoperiodic signaling in transcriptional nuclear foci. *Developmental cell*. **35**(3), pp.311-321.
- Karpinska, B., Alomrani, S.O. and Foyer, C.H.** 2017. Inhibitor-induced oxidation of the nucleus and cytosol in *Arabidopsis thaliana*: implications for organelle to nucleus retrograde signalling. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **372**(1730).
- Karrer, K.M., Peiffer, S.L. and DiTomas, M.E.** 1993. Two distinct gene subfamilies within the family of cysteine protease genes. *Proceedings of the National Academy of Sciences*. **90**(7), pp.3063-3067.
- Kashyap, P.L., Kumar, S., Jasrotia, P., Singh, D.P. and Singh, G.P.** 2020. Nanotechnology in Wheat Production and Protection. *Environmental Nanotechnology Volume 4*. Springer, pp.165-194.
- Kato, Y., Murakami, S., Yamamoto, Y., Chatani, H., Kondo, Y., Nakano, T., Yokota, A. and Sato, F.** 2004. The DNA-binding protease, CND41, and the degradation of ribulose-1, 5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. *Planta*. **220**(1), pp.97-104.
- Kato, Y. and Sakamoto, W.** 2010. New insights into the types and function of proteases in plastids. *International review of cell and molecular biology*. Elsevier, pp.185-218.
- Kaul, S., Koo, H.L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L.J., Feldblyum, T., Nierman, W., Benito, M.I. and Lin, X.** 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. **408**(6814), pp.796-815.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.** 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*. **10**(6), pp.845-858.
- Khan, N.Z., Lindquist, E. and Aronsson, H.** 2013. New putative chloroplast vesicle transport components and cargo proteins revealed using a bioinformatics approach: an *Arabidopsis* model. *PLoS One*. **8**(4), pe59898.
- Kidrič, M., Kos, J. and Sabotič, J.** 2014. Proteases and their endogenous inhibitors in the plant response to abiotic stress. *Botanica serbica*. **38**(1), pp.139-158.
- Koizumi, M., Yamaguchi-Shinozaki, K., Tsuji, H. and Shinozaki, K.** 1993. Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*. *Gene*. **129**(2), pp.175-182.
- Konno, K., Hirayama, C., Nakamura, M., Tateishi, K., Tamura, Y., Hattori, M. and Kohno, K.** 2004. Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. *The Plant Journal*. **37**(3), pp.370-378.

- Kramer, D.M., Sacksteder, C.A. and Cruz, J.A.** 1999. How acidic is the lumen? *Photosynthesis research*. **60**(2-3), pp.151-163.
- Kromdijk, J., Glowacka, K., Leonelli, L., Gabilly, S.T., Iwai, M., Niyogi, K.K. and Long, S.P.** 2016. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science*. **354**(6314), pp.857-861.
- Krüger, J., Thomas, C.M., Golstein, C., Dixon, M.S., Smoker, M., Tang, S., Mulder, L. and Jones, J.D.G.** 2002. A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis. *Science*. **296**(5568), pp.744-747.
- Kunert, K.J., van Wyk, S.G., Cullis, C.A., Vorster, B.J. and Foyer, C.H.** 2015. Potential use of phytocystatins in crop improvement, with a particular focus on legumes. *Journal of Experimental Botany*. **66**(12), pp.3559-3570.
- Kurepa, J. and Smalle, J.A.** 2008. Structure, function and regulation of plant proteasomes. *Biochimie*. **90**(2), pp.324-335.
- Kurepa, J., Wang, S., Li, Y. and Smalle, J.** 2009. Proteasome regulation, plant growth and stress tolerance. *Plant signaling & behavior*. **4**(10), pp.924-927.
- Lawrence, P.K. and Koundal, K.R.** 2002. Plant protease inhibitors in control of phytophagous insects. *Electronic Journal of Biotechnology*. **5**(1), pp.5-6.
- Lechtenberg, B., Schubert, D., Forsbach, A., Gils, M. and Schmidt, R.** 2003. Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. *The Plant Journal*. **34**(4), pp.507-517.
- Leister, D.** 2012. Retrograde signaling in plants: from simple to complex scenarios. *Frontiers in plant science*. **3**, p135.
- Lesk, C., Rowhani, P. and Ramankutty, N.** 2016. Influence of extreme weather disasters on global crop production. *Nature*. **529**(7584), pp.84-87.
- Li, C., Li, M., Dunwell, J.M. and Zhang, Y.-M.** 2012. Gene duplication and an accelerated evolutionary rate in 11S globulin genes are associated with higher protein synthesis in dicots as compared to monocots. *BMC evolutionary biology*. **12**(1), p15.
- Liang, C., Brookhart, G., Feng, G., Reeck, G. and Kramer, K.** 1991. Inhibition of digestive proteinase of stored grain Coleoptera by oryzacystatin, a cysteine proteinase inhibitor from rice seed. *FEBS letters*. **278**(2), pp.139-142.
- Lichtenthaler, H.K.** 1987. Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Methods in enzymology*. Elsevier, pp.350-382.
- Lindahl, M., Spetea, C., Hundal, T., Oppenheim, A.B., Adam, Z. and Andersson, B.** 2000. The thylakoid FtsH protease plays a role in the light-induced turnover of the photosystem II D1 protein. *The Plant Cell*. **12**(3), pp.419-431.
- Liu, H., Hu, M., Wang, Q., Cheng, L. and Zhang, Z.** 2018. Role of papain-like cysteine proteases in plant development. *Frontiers in plant science*. **9**, p1717.
- Liu, Y., Wang, K., Cheng, Q., Kong, D., Zhang, X., Wang, Z., Wang, Q., Qi, X., Yan, J. and Chu, J.** 2020. Cysteine protease RD21A regulated by E3 ligase SINAT4 is required for drought-

- induced resistance to *Pseudomonas syringae* in *Arabidopsis*. *Journal of Experimental Botany*. **71**(18), pp. 5562–5576.
- Livak, K.J. and Schmittgen, T.D.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-(\Delta\Delta C(T))}$. *Methods*. **25**(4), pp.402-408.
- Lohman, K.N., Gan, S., John, M.C. and Amasino, R.M.** 1994. Molecular analysis of natural leaf senescence in *Arabidopsis thaliana*. *Physiologia Plantarum*. **92**(2), pp.322-328.
- Madureira, H.C., Da Cunha, M. and Jacinto, T.** 2006. Immunolocalization of a defense-related 87kDa cystatin in leaf blade of tomato plants. *Environmental and experimental botany*. **55**(1), pp.201-208.
- Mahajan, S. and Tuteja, N.** 2005. Cold, salinity and drought stresses: an overview. *Archives of biochemistry and biophysics*. **444**(2), pp.139-158.
- Maidment, J.M., Moore, D., Murphy, G.P., Murphy, G. and Clark, I.M.** 1999. Matrix metalloproteinase homologues from *Arabidopsis thaliana* expression and activity. *Journal of Biological Chemistry*. **274**(49), pp.34706-34710.
- Majsec, K., Bhuiyan, N.H., Sun, Q., Kumari, S., Kumar, V., Ware, D. and van Wijk, K.J.** 2017. The plastid and mitochondrial peptidase network in *Arabidopsis thaliana*: a foundation for testing genetic interactions and functions in organellar proteostasis. *The Plant Cell*. **29**(11), pp.2687-2710.
- Makarova, K.S., Aravind, L. and Koonin, E.V.** 2000. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends in biochemical sciences*. **25**(2), pp.50-52.
- Mamaeva, A., Taliansky, M., Filippova, A., Love, A.J., Golub, N. and Fesenko, I.** 2020. The role of chloroplast protein remodeling in stress responses and shaping of the plant peptidome. *New Phytologist*. **227**, pp 1326–1334.
- Martinez, Cambra, I., Carrillo, L., Diaz-Mendoza, M. and Diaz, I.** 2009. Characterization of the entire cystatin gene family in barley and their target cathepsin L-like cysteine-proteases, partners in the hordein mobilization during seed germination. *Plant Physiology*. **151**(3), pp.1531-1545.
- Martínez, D.E., Bartoli, C.G., Grbic, V. and Guamet, J.J.** 2007. Vacuolar cysteine proteases of wheat (*Triticum aestivum* L.) are common to leaf senescence induced by different factors. *Journal of Experimental Botany*. **58**(5), pp.1099-1107.
- Martínez, M., Cambra, I., González-Melendi, P., Santamaría, M.E. and Díaz, I.** 2012. C1A cysteine-proteases and their inhibitors in plants. *Physiologia Plantarum*. **145**(1), pp.85-94.
- Martinez, M., Gómez-Cabellos, S., Giménez, M.J., Barro, F., Diaz, I. and Diaz-Mendoza, M.** 2019. Plant proteases: from key enzymes in germination to allies for fighting human gluten-related disorders. *Frontiers in plant science*. **10**, p721.
- Martinez, M., Lopez-Solanilla, E., Rodriguez-Palenzuela, P., Carbonero, P. and Diaz, I.** 2003. Inhibition of plant-pathogenic fungi by the barley cystatin Hv-CPI (gene Icy) is not associated with its cysteine-proteinase inhibitory properties. *Molecular plant-microbe interactions*. **16**(10), pp.876-883.

- Marttila, S., Jones, B.L. and Mikkonen, A.** 1995. Differential localization of two acid proteinases in germinating barley (*Hordeum vulgare*) seed. *Physiologia Plantarum*. **93**(2), pp.317-327.
- Mattoo, A.K. and Handa, A.K.** 2004. Ethylene signaling in plant cell death. *Plant cell death processes*. Elsevier, pp.125-142.
- Mega, R., Abe, F., Kim, J.-S., Tsuboi, Y., Tanaka, K., Kobayashi, H., Sakata, Y., Hanada, K., Tsujimoto, H. and Kikuchi, J.** 2019. Tuning water-use efficiency and drought tolerance in wheat using abscisic acid receptors. *Nature plants*. **5**(2), pp.153-159.
- Michaeli, S., Avin-Wittenberg, T. and Galili, G.** 2014. Involvement of autophagy in the direct ER to vacuole protein trafficking route in plants. *Frontiers in plant science*. **5**, p134.
- Milner, M.J., Howells, R.M., Craze, M., Bowden, S., Graham, N. and Wallington, E.J.** 2018. A PSTOL-like gene, TaPSTOL, controls a number of agronomically important traits in wheat. *BMC plant biology*. **18**(1), pp.1-14.
- Min, C.W., Kim, Y.J., Gupta, R., Kim, S.W., Han, W.Y., Ko, J.M., Kang, H.W., Yoon, W.B., Choung, M.G. and Kim, Y.C.** 2016. High-throughput proteome analysis reveals changes of primary metabolism and energy production under artificial aging treatment in Glycine max seeds. *Applied Biological Chemistry*. **59**(6), pp.841-853.
- Mishra, Y., Jänkänpää, H.J., Kiss, A.Z., Funk, C., Schröder, W.P. and Jansson, S.** 2012. Arabidopsis plants grown in the field and climate chambers significantly differ in leaf morphology and photosystem components. *BMC Plant Biology*. **12**(1), p6.
- Mochizuki, N., Brusslan, J.A., Larkin, R., Nagatani, A. and Chory, J.** 2001. Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proceedings of the National Academy of sciences*. **98**(4), pp.2053-2058.
- Møller, I.M. and Sweetlove, L.J.** 2010. ROS signalling—specificity is required. *Trends in plant science*. **15**(7), pp.370-374.
- Morrell, R. and Sadanandom, A.** 2019. Dealing with stress: A review of plant SUMO proteases. *Frontiers in plant science*. **10**, p1122.
- Munger, A., Simon, M.-A., Khalf, M., Goulet, M.-C. and Michaud, D.** 2015. Cereal cystatins delay sprouting and nutrient loss in tubers of potato, *Solanum tuberosum*. *BMC plant biology*. **15**(1), p296.
- Muskens, M.M., Vissers, A.A., Mol, J.M. and Kooter, J.** 2000. Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing. *Plant Molecular Biology*. **43**(2-3), pp.243-260.
- Nakaune, S., Yamada, K., Kondo, M., Kato, T., Tabata, S., Nishimura, M. and Hara-Nishimura, I.** 2005. A vacuolar processing enzyme, δ VPE, is involved in seed coat formation at the early stage of seed development. *The Plant Cell*. **17**(3), pp.876-887.
- Nelson, C.J. and Millar, A.H.** 2015. Protein turnover in plant biology. *Nature plants*. **1**(3), pp.1-7.
- Neuteboom, L.W., Matsumoto, K.O. and Christopher, D.A.** 2009. An extended AE-rich N-terminal trunk in secreted pineapple cystatin enhances inhibition of fruit bromelain and is posttranslationally removed during ripening. *Plant physiology*. **151**(2), pp.515-527.

- Nishinari, K., Fang, Y., Guo, S. and Phillips, G.** 2014. Soy proteins: A review on composition, aggregation and emulsification. *Food hydrocolloids*. **39**, pp.301-318.
- Nissen, M.S., Kumar, G.M., Youn, B., Knowles, D.B., Lam, K.S., Ballinger, W.J., Knowles, N.R. and Kang, C.** 2009. Characterization of *Solanum tuberosum* multicystatin and its structural comparison with other cystatins. *The Plant Cell*. **21**(3), pp.861-875.
- Otegui, M.S.** 2018. Vacuolar degradation of chloroplast components: autophagy and beyond. *Journal of Experimental Botany*. **69**(4), pp.741-750.
- Otto, H.-H. and Schirmeister, T.** 1997. Cysteine proteases and their inhibitors. *Chemical reviews*. **97**(1), pp.133-172.
- Outchkourov, N.S., De Kogel, W.J., Schuurman-de Bruin, A., Abrahamson, M. and Jongma, M.A.** 2004. Specific cysteine protease inhibitors act as deterrents of western flower thrips, *Frankliniella occidentalis* (Pergande), in transgenic potato. *Plant Biotechnology Journal*. **2**(5), pp.439-448.
- Palma, J.M., Sandalio, L.M., Corpas, F.J., Romero-Puertas, M.C., McCarthy, I. and Luis, A.** 2002. Plant proteases, protein degradation, and oxidative stress: role of peroxisomes. *Plant Physiology and Biochemistry*. **40**(6-8), pp.521-530.
- Pan, J., Fu, Q. and Xu, Z.-F.** 2010. Agrobacterium tumefaciens-mediated transformation of biofuel plant *Jatropha curcas* using kanamycin selection. *African Journal of Biotechnology*. **9**(39), pp.6477-6481.
- Parisi, C., Tillie, P. and Rodríguez-Cerezo, E.** 2016. The global pipeline of GM crops out to 2020. *Nature Biotechnology*. **34**(1), pp.31-36.
- Parry, M.A., Keys, A.J., Madgwick, P.J., Carmo-Silva, A.E. and Andralojc, P.J.** 2008. Rubisco regulation: a role for inhibitors. *Journal of Experimental Botany*. **59**(7), pp.1569-1580.
- Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., Shannon, G.J., Carter, T.C. and Nguyen, H.T.** 2017. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theoretical and Applied Genetics*. **130**(10), pp.1975-1991.
- Paz, M., Martinez, J., Kalvig, A., Fonger, T. and Wang, K.** 2006. Improved cotyledonary node method using an alternative explant derived from mature seed for efficient Agrobacterium-mediated soybean transformation. *Plant Cell Reports*. **25**(3), pp.206-213.
- Pearce, S., Tabbita, F., Cantu, D., Buffalo, V., Avni, R., Vazquez-Gross, H., Zhao, R., Conley, C.J., Distelfeld, A. and Dubcovksy, J.** 2014. Regulation of Zn and Fe transporters by the GPC1 gene during early wheat monocarpic senescence. *BMC Plant Biology*. **14**(1), p368.
- Peng, F.Y., Hu, Z. and Yang, R.-C.** 2015. Genome-wide comparative analysis of flowering-related genes in Arabidopsis, wheat, and barley. *International journal of plant genomics*.
- Pernas, M., Sanchez-Monge, R. and Salcedo, G.** 2000. Biotic and abiotic stress can induce cystatin expression in chestnut. *FEBS Lett*. **467**(2-3), pp.206-210.
- Pillay, P., Kibido, T., Du Plessis, M., Van Der Vyver, C., Beyene, G., Vorster, B., Kunert, K. and Schlüter, U.** 2012. Use of transgenic oryzacystatin-I-expressing plants enhances recombinant protein production. *Applied biochemistry and biotechnology*. **168**(6), pp.1608-1620.

- Pillay, P., Schlüter, U., van Wyk, S., Kunert, K.J. and Vorster, B.J.** 2014. Proteolysis of recombinant proteins in bioengineered plant cells. *Bioengineered*. **5**(1), pp.15-20.
- Poret, M., Chandrasekar, B., van der Hoorn, R.A. and Avice, J.-C.** 2016. Characterization of senescence-associated protease activities involved in the efficient protein remobilization during leaf senescence of winter oilseed rape. *Plant Science*. **246**, pp.139-153.
- Poret, M., Chandrasekar, B., Van der Hoorn, R.A., Coquet, L., Jouenne, T. and Avice, J.-C.** 2017. Proteomic investigations of proteases involved in cotyledon senescence: a model to explore the genotypic variability of proteolysis machinery associated with nitrogen remobilization efficiency during the leaf senescence of oilseed rape. *Proteomes*. **5**(4), p29.
- Prins, A., Van Heerden, P.D., Olmos, E., Kunert, K.J. and Foyer, C.H.** 2008. Cysteine proteinases regulate chloroplast protein content and composition in tobacco leaves: a model for dynamic interactions with ribulose-1, 5-bisphosphate carboxylase/oxygenase (Rubisco) vesicular bodies. *Journal of Experimental Botany*. **59**(7), pp.1935-1950.
- Quain, M.D., Makgopa, M.E., Cooper, J.W., Kunert, K.J. and Foyer, C.H.** 2015. Ectopic phytolectin expression increases nodule numbers and influences the responses of soybean (*Glycine max*) to nitrogen deficiency. *Phytochemistry*. **112**, pp.179-187.
- Quain, M.D., Makgopa, M.E., Márquez-García, B., Comadira, G., Fernandez-Garcia, N., Olmos, E., Schnaubelt, D., Kunert, K.J. and Foyer, C.H.** 2014. Ectopic phytolectin expression leads to enhanced drought stress tolerance in soybean (*Glycine max*) and *Arabidopsis thaliana* through effects on strigolactone pathways and can also result in improved seed traits. *Plant biotechnology journal*. **12**(7), pp.903-913.
- Rabbani, M.A., Maruyama, K., Abe, H., Khan, M.A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K.** 2003. Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant physiology*. **133**(4), pp.1755-1767.
- Radhika, V. and Rao, V.S.H.** 2015. Computational approaches for the classification of seed storage proteins. *Journal of food science and technology*. **52**(7), pp.4246-4255.
- Ramalho-Santos, M., Veríssimo, P., Faro, C. and Pires, E.** 1996. Action on bovine α 1-casein of cardosins A and B, aspartic proteinases from the flowers of the cardoon *Cynara cardunculus* L. *Biochimica et Biophysica Acta -Protein Structure Molecular Enzymology*. **1297**(1), pp.83-89.
- Raven, J.A.** 2013. Rubisco: still the most abundant protein of Earth? *New Phytologist*. **198**(1), pp.1-3.
- Rawlings, N.D., Barrett, A.J. and Bateman, A.** 2010. MEROPS: the peptidase database. *Nucleic acids research*. **38**(suppl_1), pp.D227-D233.
- Rawlings, N.D., Barrett, A.J. and Bateman, A.** 2011. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research*. **40**(D1), pp.D343-D350.
- Rawlings, N.D., Barrett, A.J. and Finn, R.** 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research*. **44**(D1), pp.D343-D350.
- Rawlings, N.D., Morton, F.R. and Barrett, A.J.** 2006. MEROPS: the peptidase database. *Nucleic acids research*. **34**(suppl_1), pp.D270-D272.

- Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., Parry, M. and Slafer, G.** 2012. Achieving yield gains in wheat. *Plant, cell & environment*. **35**(10), pp.1799-1823.
- Richau, K.H., Kaschani, F., Verdoes, M., Pansuriya, T.C., Niessen, S., Stüber, K., Colby, T., Overkleeft, H.S., Bogyo, M. and Van der Hoorn, R.A.** 2012. Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant physiology*. **158**(4), pp.1583-1599.
- Roa-Roberts, D.** 2014. Characterisation of *Arabidopsis thaliana* expressing oryzacystatin-1 (OC-1) in the chloroplast. Master dissertation. University of Leeds.
- Roberts, I.N., Caputo, C., Criado, M.V. and Funk, C.** 2012. Senescence-associated proteases in plants. *Physiologia Plantarum*. **145**(1), pp.130-139.
- Rogers, J.C., Dean, D. and Heck, G.R.** 1985. Aleurain: a barley thiol protease closely related to mammalian cathepsin H. *Proceedings of the National Academy of Sciences*. **82**(19), pp.6512-6516.
- Rojo, E., Zouhar, J., Carter, C., Kovaleva, V. and Raikhel, N.V.** 2003. A unique mechanism for protein processing and degradation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*. **100**(12), pp.7389-7394.
- Romero, L.C., Aroca, M.Á., Laureano-Marín, A.M., Moreno, I., García, I. and Gotor, C.** 2014. Cysteine and cysteine-related signaling pathways in *Arabidopsis thaliana*. *Molecular plant*. **7**(2), pp.264-276.
- Rossel, J.B., Wilson, P.B., Hussain, D., Woo, N.S., Gordon, M.J., Mewett, O.P., Howell, K.A., Whelan, J., Kazan, K. and Pogson, B.J.** 2007. Systemic and intracellular responses to photooxidative stress in *Arabidopsis*. *The Plant Cell*. **19**(12), pp.4091-4110.
- Rustgi, S., Boex-Fontvieille, E., Reinbothe, C., von Wettstein, D. and Reinbothe, S.** 2018. The complex world of plant protease inhibitors: Insights into a Kunitz-type cysteine protease inhibitor of *Arabidopsis thaliana*. *Communicative & Integrative Biology*. **11**(1), pe1368599.
- Sadali, N.M., Sowden, R.G., Ling, Q. and Jarvis, R.P.** 2019. Differentiation of chromoplasts and other plastids in plants. *Plant cell reports*. pp.1-16.
- Sakamoto, W., Miyagishima, S.-y. and Jarvis, P.** 2008. Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. *The Arabidopsis book/American Society of Plant Biologists*. **6**.
- Schalk, K., Lexhaller, B., Koehler, P. and Scherf, K.A.** 2017. Isolation and characterization of gluten protein types from wheat, rye, barley and oats for use as reference materials. *PloS one*. **12**(2).
- Schaller, A.** 2004. A cut above the rest: the regulatory function of plant proteases. *Planta*. **220**(2), pp.183-197.
- Schippers, J.H., Schmidt, R., Wagstaff, C. and Jing, H.-C.** 2015. Living to die and dying to live: the survival strategy behind leaf senescence. *Plant physiology*. **169**(2), pp.914-930.
- Schlütera, U., Vorstera, B., Bishop, Ö.T., Kunerta, K. and Michaudc, D.** 2009. Phytocystatins: A comparative analysis of cysteine protease inhibitors in plants. *South African Journal of Botany*. **75**(2), p419.

- Schmidt, M.A., Barbazuk, W.B., Sandford, M., May, G., Song, Z., Zhou, W., Nikolau, B.J. and Herman, E.M.** 2011. Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. *Plant Physiology*. **156**(1), pp.330-345.
- Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A. and Sakurai, T.** 2002. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal*. **31**(3), pp.279-292.
- Seo, S.-b., Tan-Wilson, A. and Wilson, K.A.** 2001. Protease C2, a cysteine endopeptidase involved in the continuing mobilization of soybean β -conglycinin seed proteins. *Biochimica et Biophysica Acta -Protein Structure Molecular Enzymology*. **1545**(1-2), pp.192-206.
- Shahri, W. and Tahir, I.** 2014. Flower senescence: some molecular aspects. *Planta*. **239**(2), pp.277-297.
- Sheokand, S., Dahiya, P., Vincent, J.L. and Brewin, N.J.** 2005. Modified expression of cysteine protease affects seed germination, vegetative growth and nodule development in transgenic lines of *Medicago truncatula*. *Plant Science*. **169**(5), pp.966-975.
- Shewry, P.R. and Halford, N.G.** 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. *Journal of Experimental Botany*. **53**(370), pp.947-958.
- Shewry, P.R., Tatham, A.S., Barro, F., Barcelo, P. and Lazzeri, P.** 1995. Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. *Biotechnology*. **13**(11), pp.1185-1190.
- Soares, A.S., Driscoll, S.P., Olmos, E., Harbinson, J., Arrabaça, M.C. and Foyer, C.H.** 2008. Adaxial/abaxial specification in the regulation of photosynthesis and stomatal opening with respect to light orientation and growth with CO₂ enrichment in the C₄ species *Paspalum dilatatum*. *New Phytologist*. **177**(1), pp.186-198.
- Sokolenko, A., Altschmied, L. and Herrmann, R.G.** 1997. Sodium dodecyl sulfate-stable proteases in chloroplasts. *Plant physiology*. **115**(2), pp.827-832.
- Solomon, M., Belenghi, B., Delledonne, M., Menachem, E. and Levine, A.** 1999. The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. *Plant Cell*. **11**(3), pp.431-444.
- Song, L., Chen, Z. and Larkin, R.M.** 2018. The genomes uncoupled mutants are more sensitive to norflurazon than wild type. *Plant Physiology*. **178**(3), pp.965-971.
- Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z. and Lin, H.-X.** 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature genetics*. **39**(5), pp.623-630.
- Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. and Turk, V.** 1990. The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *The EMBO journal*. **9**(6), pp.1939-1947.
- Sueldo, D.J. and van der Hoorn, R.A.** 2017. Plant life needs cell death, but does plant cell death need Cys proteases? *The FEBS Journal*. **284**(10), pp.1577-1585.

- Susek, R.E., Ausubel, F.M. and Chory, J.** 1993. Signal transduction mutants of Arabidopsis uncouple nuclear CAB and RBCS gene expression from chloroplast development. *Cell*. **74**(5), pp.787-799.
- Szawińska, J., Simińska, J. and Bielawski, W.** 2016. The roles of cysteine proteases and phytocystatins in development and germination of cereal seeds. *Journal of plant physiology*. **207**, pp.10-21.
- Tan-Wilson, A.L. and Wilson, K.A.** 2012. Mobilization of seed protein reserves. *Physiologia Plantarum*. **145**(1), pp.140-153.
- Tanaka, A. and Makino, A.** 2009. Photosynthetic Research in Plant Science. *Plant and Cell Physiology*. **50**(4), pp.681-683.
- Tang, W., Newton, R.J. and Weidner, D.A.** 2007. Genetic transformation and gene silencing mediated by multiple copies of a transgene in eastern white pine. *Journal of Experimental Botany*. **58**(3), pp.545-554.
- Taski-Ajdukovic, K., Djordjevic, V., Vidic, M. and Vujakovic, M.** 2010. Subunit composition of seed storage proteins in high-protein soybean genotypes. *Pesquisa Agropecuária Brasileira*. **45**(7), pp.721-729.
- Terry, M.J. and Smith, A.G.** 2013. A model for tetrapyrrole synthesis as the primary mechanism for plastid-to-nucleus signaling during chloroplast biogenesis. *Frontiers in plant science*. **4**, p14.
- Tester, M. and Langridge, P.** 2010. Breeding technologies to increase crop production in a changing world. *Science*. **327**(5967), pp.818-822.
- Thanh, V.H. and Shibasaki, K.** 2002. Major proteins of soybean seeds. Subunit structure of beta-conglycinin. *Journal of Agricultural Food Chemistry*. **26**(3), pp.692-695.
- Thind, A.K., Wicker, T., Müller, T., Ackermann, P.M., Steuernagel, B., Wulff, B.B., Spannagl, M., Twardziok, S.O., Felder, M. and Lux, T.** 2018. Chromosome-scale comparative sequence analysis unravels molecular mechanisms of genome dynamics between two wheat cultivars. *Genome biology*. **19**(1), pp.1-16.
- Thoenen, M., Herrmann, B. and Feller, U.** 2007. Senescence in wheat leaves: is a cysteine endopeptidase involved in the degradation of the large subunit of Rubisco? *Acta Physiologiae Plantarum*. **29**(4), pp.339-350.
- Tornkvist, A., Liu, C. and Moschou, P.N.** 2019. Proteolysis and nitrogen: emerging insights. *Journal of experimental botany*. **70**(7), pp.2009-2019.
- Toyooka, K., Okamoto, T. and Minamikawa, T.** 2000. Mass Transport of Proform of a Kdel-Tailed Cysteine Proteinase (Sh-EP) to Protein Storage Vacuoles by Endoplasmic Reticulum-Derived Vesicle Is Involved in Protein Mobilization in Germinating Seeds. *The Journal of Cell Biology*. **148**(3), pp.453-464.
- Tsuji, A., Tsukamoto, K., Iwamoto, K., Ito, Y. and Yuasa, K.** 2013. Enzymatic characterization of germination-specific cysteine protease-1 expressed transiently in cotyledons during the early phase of germination. *The journal of biochemistry*. **153**(1), pp.73-83.
- Turk, V. and Bode, W.** 1991. The cystatins: protein inhibitors of cysteine proteinases. *FEBS letters*. **285**(2), pp.213-219.

- Tuteja, N.** 2007. Abscisic acid and abiotic stress signaling. *Plant signaling & behavior*. **2**(3), pp.135-138.
- Urwin, P.E., Green, J. and Atkinson, H.J.** 2003. Expression of a plant cystatin confers partial resistance to Globodera, full resistance is achieved by pyramiding a cystatin with natural resistance. *Molecular Breeding*. **12**(3), pp.263-269.
- Urwin, P.E., Troth, K.M., Zubko, E.I. and Atkinson, H.J.** 2001. Effective transgenic resistance to Globodera pallida in potato field trials. *Molecular Breeding*. **8**(1), pp.95-101.
- Van der Hoorn, R.A.** 2008. Plant proteases: from phenotypes to molecular mechanisms. *Annual Review of Plant Biology*. **59**, pp.191-223.
- van der Hoorn, R.A. and Jones, J.D.** 2004. The plant proteolytic machinery and its role in defence. *Current opinion in plant biology*. **7**(4), pp.400-407.
- Van der Vyver, C., Schneidereit, J., Driscoll, S., Turner, J., Kunert, K. and Foyer, C.H.** 2003. Oryzacystatin I expression in transformed tobacco produces a conditional growth phenotype and enhances chilling tolerance. *Plant biotechnology journal*. **1**(2), pp.101-112.
- Velasco-Arroyo, B., Diaz-Mendoza, M., Gandullo, J., Gonzalez-Melendi, P., Santamaria, M.E., Dominguez-Figueroa, J.D., Hensel, G., Martinez, M., Kumlehn, J. and Diaz, I.** 2016. HvPap-1 C1A protease actively participates in barley proteolysis mediated by abiotic stresses. *Journal of Experimental Botany*. **67**(14), pp.4297-4310.
- Vierstra, R.D.** 1996. Proteolysis in plants: mechanisms and functions. Post-Transcriptional Control of Gene Expression in Plants. *Springer*, pp.275-302.
- Vierstra, R.D.** 2009. The ubiquitin–26S proteasome system at the nexus of plant biology. *Nature Reviews Molecular Cell Biology*. **10**(6), pp.385-397.
- Vitale, A. and Hinz, G.** 2005. Sorting of proteins to storage vacuoles: how many mechanisms? *Trends in plant science*. **10**(7), pp.316-323.
- Voigt, C., Oster, U., Börnke, F., Jahns, P., Dietz, K.J., Leister, D. and Kleine, T.** 2010. In-depth analysis of the distinctive effects of norflurazon implies that tetrapyrrole biosynthesis, organellar gene expression and ABA cooperate in the GUN-type of plastid signalling. *Physiologia Plantarum*. **138**(4), pp.503-519.
- Wagner, D., Przybyla, D., op den Camp, R., Kim, C., Landgraf, F., Lee, K.P., Würsch, M., Laloi, C., Nater, M. and Hideg, E.** 2004. The genetic basis of singlet oxygen–induced stress responses of Arabidopsis thaliana. *Science*. **306**(5699), pp.1183-1185.
- Wang, F., Liu, J., Chen, M., Zhou, L., Li, Z., Zhao, Q., Pan, G. and Cheng, F.** 2016. Involvement of abscisic acid in PSII photodamage and D1 protein turnover for light-induced premature senescence of rice flag leaves. *PloS one*. **11**(8), pe0161203.
- Wang, P., Mugume, Y. and Bassham, D.C.** 2018. New advances in autophagy in plants: regulation, selectivity and function. In: Seminars in cell & developmental biology: *Elsevier*, pp.113-122.
- Wei, X., Kim, W.-S., Song, B., Oehrle, N.W., Liu, S. and Krishnan, H.B.** 2020. Soybean Mutants Lacking Abundant Seed Storage Proteins Are Impaired in Mobilization of Storage Reserves and Germination. *ACS omega*. **5**(14), pp.8065-8075.

- Wickner, S. and Maurizi, M.R.** 1999. Here's the hook: similar substrate binding sites in the chaperone domains of Clp and Lon. *Proceedings of the National Academy of Sciences*. **96**(15), pp.8318-8320.
- Wisniewski, K. and Zagdanska, B.** 2001. Genotype-dependent proteolytic response of spring wheat to water deficiency. *Journal of Experimental Botany*. **52**(360), pp.1455-1463.
- Woodson, J.D., Perez-Ruiz, J.M., Schmitz, R.J., Ecker, J.R. and Chory, J.** 2013. Sigma factor-mediated plastid retrograde signals control nuclear gene expression. *The Plant Journal*. **73**(1), pp.1-13.
- Wu, G.-Z., Meyer, E.H., Richter, A.S., Schuster, M., Ling, Q., Schöttler, M.A., Walther, D., Zoschke, R., Grimm, B., Jarvis, R.P. and Bock, R.** 2019. Control of retrograde signalling by protein import and cytosolic folding stress. *Nature Plants*. **5**(5), pp.525-538.
- Xiao, Y., Savchenko, T., Baidoo, E.E., Chehab, W.E., Hayden, D.M., Tolstikov, V., Corwin, J.A., Kliebenstein, D.J., Keasling, J.D. and Dehesh, K.** 2012. Retrograde signaling by the plastidial metabolite MEcPP regulates expression of nuclear stress-response genes. *Cell*. **149**(7), pp.1525-1535.
- Xiong, Y., Contento, A.L., Nguyen, P.Q. and Bassham, D.C.** 2007. Degradation of oxidized proteins by autophagy during oxidative stress in Arabidopsis. *Plant physiology*. **143**(1), pp.291-299.
- Yang, A. and Yeh, K.** 2005. Molecular cloning, recombinant gene expression, and antifungal activity of cystatin from taro (*Colocasia esculenta* cv. Kaosiung no. 1). *Planta*. **221**(4), pp.493-501.
- Yenofsky, R.L., Fine, M. and Pellow, J.W.** 1990. A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. *Proceedings of the National Academy of Sciences*. **87**(9), pp.3435-3439.
- Yoshimoto, K., Hanaoka, H., Sato, S., Kato, T., Tabata, S., Noda, T. and Ohsumi, Y.** 2004. Processing of ATG8s, ubiquitin-like proteins, and their deconjugation by ATG4s are essential for plant autophagy. *The Plant Cell*. **16**(11), pp.2967-2983.
- Yu, A.Y.H. and Houry, W.A.** 2007. ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS letters*. **581**(19), pp.3749-3757.
- Zamyatnin, A.** 2015. Plant proteases involved in regulated cell death. *Biochemistry*. **80**(13), pp.1701-1715.
- Zhang and Jones, B.L.** 1995. Characterization of germinated barley endoproteolytic enzymes by two-dimensional gel electrophoresis. *Journal of Cereal Science*. **21**(2), pp.145-153.
- Zhang, D., Liu, D., Lv, X., Wang, Y., Xun, Z., Liu, Z., Li, F. and Lu, H.** 2014. The cysteine protease CEP1, a key executor involved in tapetal programmed cell death, regulates pollen development in Arabidopsis. *The Plant Cell*. **26**(7), pp.2939-2961.
- Zhang, X., Liu, S. and Takano, T.** 2008. Two cysteine proteinase inhibitors from Arabidopsis thaliana, AtCYSa and AtCYSb, increasing the salt, drought, oxidation and cold tolerance. *Plant molecular biology*. **68**(1-2), pp.131-143.
- Zhang, X.M., Wang, Y., Lv, X.M., Li, H., Sun, P., Lu, H. and Li, F.L.** 2009. NtCP56, a new cysteine protease in Nicotiana tabacum L., involved in pollen grain development. *Journal of Experimental Botany*. **60**(6), pp.1569-1577.

- Zhao, C., Johnson, B.J., Kositsup, B. and Beers, E.P.** 2000. Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. *Plant Physiology*. **123**(3), pp.1185-1196.
- Zou, Z., Huang, Q., Xie, G. and Yang, L.** 2018. Genome-wide comparative analysis of papain-like cysteine protease family genes in castor bean and physic nut. *Scientific Reports*. **8**(1), pp.1-13.
- Zou, Z., Liu, J., Yang, L. and Xie, G.** 2017. Survey of the rubber tree genome reveals a high number of cysteine protease-encoding genes homologous to Arabidopsis SAG12. *PloS one*. **12**(2), pe0171725.

Appendix

Appendix I. T-DNA copy number of transformed T2 plants of transgenic wheat expressing OC-I. Copy number estimation by qPCR of the transgene are shown.

Line	Estimated Copy Number By RT- PCR	Line	Estimated Copy Number By RT-PCR	Line	Estimated Copy Number By RT-PCR
1.4.2	1	10.35.5	1	26.5.9	1
1.4.3	1	10.35.6	1	26.5.10	1
1.4.5	1	10.35.7	1	26.5.11	1
1.4.6	1	10.35.8	1	26.5.13	1
1.4.8	3	10.35.10	1	26.5.14	1
1.4.10	1	10.35.12	1	26.5.15	1
1.4.11	1	10.35.14	1	26.5.16	1
1.4.12	1	10.35.15	1	26.5.17	1
1.4.13	1	10.35.2	1	26.5.18	1
1.4.14	1	10.35.3	1	26.5.19	1
1.4.15	1	10.35.4	1	26.5.20	1
1.4.16	2	10.35.5	1	26.7.1	1
1.4.17	1	21.7.2	1	26.7.2	1
1.4.18	1	21.7.3	1	26.7.4	1
1.4.20	1	21.7.4	1	26.7.5	1
1.19.1	1	21.7.6	1	26.7.6	1
1.19.3	1	21.7.8	1	26.7.7	1
1.19.6	1	21.7.9	2	26.7.8	1
1.19.7	1	21.7.10	1	26.7.9	1
1.19.8	1	21.7.12	1	26.7.10	1
1.19.9	1	21.7.13	1	26.7.12	2
1.19.12	1	21.7.15	1	26.7.13	1
1.19.13	1	21.7.17	1	26.7.14	1
1.19.14	1	21.7.18	1	26.7.16	1
1.19.15	1	21.7.19	1	26.7.18	1
1.19.16	1	21.7.20	1	26.7.19	1
1.19.17	1	21.8.1	1	26.7.20	1
1.19.19	1	21.8.2	1	26.20.1	1
10.1.1	1	21.8.3	1	26.20.2	1
10.1.3	2	21.8.4	1	26.20.3	1
10.1.4	1	21.8.5	1	26.20.4	1
10.1.6	1	21.8.6	2	26.20.5	1
10.1.7	1	21.8.8	1	26.20.6	1
10.1.8	1	21.8.10	1	26.20.7	1
10.1.9	1	21.8.11	1	26.20.8	2
10.1.10	2	21.8.12	1	26.20.9	1
10.1.12	1	21.8.16	1	26.20.10	1
10.1.13	1	21.8.18	1	26.20.11	1
10.1.15	1	21.8.19	1	26.20.12	1
10.1.17	1	26.5.1	1	26.20.13	1
10.1.18	1	26.5.2	1	26.20.14	1
10.1.19	1	26.5.3	1	26.20.15	1
10.1.20	2	26.5.4	1	26.20.16	1
10.35.2	1	26.5.5	1	26.20.17	2
10.35.3	1	26.5.6	1	26.20.18	1
10.35.4	1	26.5.7	1	26.20.19	1
		26.5.8	1	26.20.20	1

Appendix II. T-DNA copy number of transformed T3 plants of transgenic wheat expressing OC-I. Copy number estimation by qPCR of the transgene are shown.

Line	Estimated Copy Number By RT- PCR	Line	Estimated Copy Number By RT-PCR	Line	Estimated Copy Number By RT-PCR
21.7.2.1	1	21.8.12.1	1	26.20.4.1	1
21.7.2.2	1	21.8.12.2	1	26.20.4.2	1
21.7.2.3	1	21.8.12.3	1	26.20.4.3	1
21.7.2.4	1	21.8.12.4	1	26.20.4.4	1
21.7.2.5	1	21.8.12.5	1	26.20.4.5	1
21.7.2.6	1	21.8.12.6	1	26.20.4.6	1
21.7.2.7	1	21.8.12.7	1	26.20.4.7	1
21.7.2.8	1	21.8.12.8	1	26.20.4.8	1
21.7.2.9	1	21.8.12.9	1	26.20.4.9	1
21.7.2.10	1	21.8.12.10	1	26.20.4.10	1
21.7.2.11	1	21.8.12.11	1	26.20.4.11	1
21.7.2.12	1	21.8.12.12	1	26.20.4.12	1
21.7.2.13	1	21.8.12.13	1	26.20.4.13	1

Appendix III. Differential abundance of storage proteins expressed in the WT and OCI-expressing line (WOC).

ID	Protein IDs	Protein names	WT	WOC	log2FC
D6QZM5	D6QZM5	Avenin-like b8	2.502163887	-0.192098618	-0.5819
F8RP11	F8RP11	Hsp70-Hsp90 organizing protein (TaHop)	-1.023381233	-2.387369156	-2.475008726
O22263	O22263	Protein disulfide-isomerase like 2-1 (AtPDIL2-1)	-2.555577278	-4.048570633	-0.673859596
P00068	P01083	Alpha-amylase inhibitor 0.28 (CIII) (WMAI-1)	-2.615328789	-3.565031052	-0.297087669
P01084	P01084	Alpha-amylase inhibitor 0.53	4.266724586	3.455438614	-0.47729969
P02276	P02276	Histone H2A.2.1	-3.179215431	-1.95889473	-0.427746773
P02277	P02277	Histone H2A.2.2	-0.445025444	1.136207581	-0.227780342
P02863	P02863	Alpha/beta-gliadin (Prolamin)	-1.593827248	-3.246582031	-2.956627369
P04464	P04464	Calmodulin (CaM)	-0.222107887	-2.049617767	-1.671975136
P04568	P04568	Em protein	2.553992271	1.869394302	-1.035862922
P04730	P04730	Gamma-gliadin (Gliadin B-III)	0.075388908	-0.892576218	-0.524403572
P08453	P08453	Gamma-gliadin	0.894663811	0.176197052	-0.693205834
P08819	P08819	Carboxypeptidase 2 (CPDW-II)	1.755984306	0.385379791	-0.415101051
P09863	P09863	Bowman-Birk type proteinase inhibitor I-2B (Fragment)	-3.133357286	-0.466344833	1.171717882
P0CZ07	P0CZ07	Avenin-like a2	-2.782131433	-2.068267822	-0.414573432
P12299	P12299	"Glucose-1-phosphate adenylyltransferase large subunit, chloroplastic/amyloplastic	-0.027422905	-0.94455719	0.53665638
P12810	P12810	16.9 kDa class I heat shock protein 1 (HSP 16.9)	-0.552893639	0.409864426	1.430567742
P26759	P26759	Oxalate oxidase GF-3.8 (Germin GF-3.8)	-0.272873878	-3.093029022	-2.376269341
P27807	P27807	Histone H2B.1	0.613600731	2.108633041	0.170847893
P29546	P29546	Elongation factor 1-beta (EF-1-beta)	-0.287636757	-4.472923279	-0.544034004
P30569	P30569	EC protein I/II	0.924546242	-0.231773376	-1.526717186
P30570	P30570	EC protein III	-2.004546165	-3.064739227	-1.890892983
P33432	P32032	Alpha-2-purothionin	-3.313580513	-1.85632515	0.677100182
P38076	P38076	Cysteine synthase (CSase A)	-3.003450394	-3.807922363	-0.506597519
P49232	P49232	Profilin-1	-3.49399662	-5.47774744	0.58794117
P68428	P68428	Histone H3.2	-3.589554548	1.467556	2.830841779
P69326	P69326	Ubiquitin	1.691996574	-0.588293076	0.162022591
P80602	P80602	"2-Cys peroxiredoxin BAS1, chloroplastic	-1.647233009	-3.073799133	0.512352943
P82900	P82900	Non-specific lipid-transfer protein 2G (LTP2G)	2.629384041	-4.679842949	-0.019534111
P82901	P82901	Non-specific lipid-transfer protein 2P (LTP2P)	3.975031853	-0.003356934	0.043276787
Q03968	Q03968	"Late embryogenesis abundant protein, group 3 (LEA) (PMA2005)"	2.272679329	-1.274335861	-1.786332131
Q05806	Q05806	Type-5 thionin	0.027422905	-2.908853292	-0.027422905
Q10464	Q10464	Puroindoline-B	-2.191935539	-0.54958725	1.898844719
Q6TCF2	Q6TCF2	Actin	-1.063479424	0.5629673	2.223593712
Q6W8Q2	Q6W8Q2	1-Cys peroxiredoxin PER1	4.013039589	1.491491318	0.166697502
Q8L5C6	Q8L5C6	Xylanase inhibitor protein 1 (XIP-1)	4.106137276	1.997665405	-0.256577492
Q8LRM8	Q8LRM8	Translationally-controlled tumor protein homolog (TCTP)	-0.277401924	-1.266302109	-0.232527733
Q949H0	Q949H0	40S ribosomal protein S7	-0.911036491	-4.534479141	-0.233460427
Q9ZRB0	Q9ZRB0	Tubulin beta-3 chain (Beta-3-tubulin)	-3.633214951	-5.016537189	1.742761612

Appendix IV. Examples of the original gels used for Western-blot analysis of the Rubisco, D1 and phosphorylated D1 proteins in the leaves of the CYS and PC lines compared to WT Arabidopsis plants grown under moderate light (LL) and high light (HL) conditions. This analysis included a gel stained with Coomassie Brilliant Blue and gel of the loading control to ensure appropriate quantitation

