Genetic dissection of the regulation of chloroplast protein import by SP1-type ubiquitin E3 ligases



Mohammad Sabri Bin Mohd. Ali
Linacre College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Plant Sciences

Trinity 2018

Genetic dissection of the regulation of chloroplast protein import by SP1-type ubiquitin E3 ligases

Mohammad Sabri Bin Mohd. Ali
Linacre College
DPhil in Plant Sciences
Trinity 2018

Abstract

Chloroplasts are photosynthetic organelles in plants, and their functions are dependent on the import of thousands of proteins from the cytosol. These proteins engage with the translocon at the outer envelope membrane of chloroplasts (TOC), which acts as a selective conduit for the first stage of their import across the double membrane system. Mutations affecting the TOC components, such as *plastid protein import 1* (*ppi1*) of *Arabidopsis thaliana*, can hinder the efficiency of import activity, and have consequences for organelle development. A screen for second-site suppressors of *ppi1* identified a novel regulator of the TOC machinery called *suppressor of ppi1 locus 1* (*sp1*). Two homologues of SP1, SP1-Like1 (SPL1) and -2 (SPL2), were identified in *Arabidopsis* based on sequence similarity. In this thesis, the function of these two homologues was characterized in relation to chloroplast biogenesis. While *spl1* null mutations had no profound effect on the *ppi1* mutant, knockdown of *SPL2* exhibited a weak suppression phenotype that could be attributed to a recovery in protein import activity and chloroplast development. Immunoprecipitation results revealed a strong physical interaction between SPL2 and the TOC components. Manipulation of *SPL2* gene expression caused phenotypes reminiscent of those caused by *SP1* manipulation during the dynamic process of leaf senescence. *In vitro* ubiquitination analysis showed that SPL1, unlike SP1 and SPL2, does not possess E3 ligase activity, but rather has an inhibitory effect on SP1 activity via an unknown mechanism. Accordingly, *spl1* null mutants displayed an opposite phenotype to *sp1* mutants during leaf senescence (i.e., senescence was accelerated), suggesting that SPL1 may be a negative regulator of SP1. However, SPL1 is only observed within the Brassicaceae family, which raises questions about the wider relevance of its role across different species. Lastly, a screen for new *sp1* enhancer mutants was conducted, and this identified three putative *sp1 enhanced chlorophyll* (*spec*) mutants that display elevated chlorophyll levels, intermediate between those of *sp1 ppi1* and wild-type plants. Mapping of the *spec* mutations by whole-genome sequencing identified several gene candidates in each case, which will require further analysis in order to conclude final identification of the causal locus in each case. Overall, the results revealed new regulatory mechanisms governing the chloroplast protein import machinery.

Acknowledgements

Firstly, I am grateful to the Brunei Government through the Ministry of Education for the scholarship and research funding for my DPhil studies.

I would like to express my sincere gratitude to my supervisors, Professor Paul Jarvis and Professor Jane Langdale, for their excellent guidance. I would like to thank them, especially Professor Jarvis, for their continuous support, patience, motivation, valuable comments, discussion, and the time and effort they contributed throughout the course of completing this DPhil.

I am indebted to many of my colleagues in the Department of Plant Sciences, University of Oxford for the support, assistance, advice and entertainment. Special mention to Na Li for her generosity and enthusiasm whenever I needed some assistance.

I am also grateful to my friends, especially Daniel Wells and Thomas Samuels, who have provided me encouragement and emotional support during stressful times. I am also grateful to my family members who have supported me along the way.

Statement of authorship

The following people made contributions to the work presented in this thesis:

Dr Qihua Ling	Produced the artificial microRNA construct for generating the *SPL2-KD* lines.
Dr Weihua Huang	Performed the *spl1-1* gene characterisation.
Dr Errin Johnson	Prepared the sample fixation for analysis using Transmission Electron Microscopy (Figure 3.5).
Na Li	Performed the immunoprecipitation assay (Figure 3.10).
Daniel Wells	Creating the platform for conducting the SHOREmap analysis and assistance with using the statistical software R.

Abbreviations

ABA	Abscisic acid
ABI5	ABA INSENSITIVE5
aci	altered chloroplast import
ADP	Adenosine diphosphate
AGD	ARF-GAP domain
AIR	Aminoimidazole ribonucleotide
amiRNA	artificial microRNA
Alb	Albino
APS	Ammonium persulphate
ARF-GAP	ADP-ribosylation factor GTPase-activating proteins
at	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
BAM	Binary alignment map
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
CBL	Calcineurin B-like proteins
cDNA	complementary DNA
CDS	Coding sequence
cia	Chloroplast import apparatus
Col-0	Columbia-0
CORVET	Class C core/vacuole/endosome tethering
Cpn	Chaperonin
Cyt	Cytosolic
dCAPS	Derived cleaved amplified polymorphic sequences

DMF	*N,N'*-dimethylforamide
DNA	Deoxyribonucleic acid
DON	Deoxynivalenol
DUF	Domain of unknown function
dNTPs	deoxyribonucleoside triphosphates
DTT	dithiothreitol
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme
E3	Ubiquitin ligase
ECT	Evolutionary conserved C-terminal region
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
EMS	Ethyl methanesulphonate
F_x	Filial generation x
fts	ferrochelatase-two-suppressor
g	g-force
GAP	GTPase-activating protein
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
GST	Glutathione S-transferases
HECT	Homologous to the E6-AP carboxyl terminus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOPS	Homotypic fusion and vacuole protein sorting
HMS	HEPES-$MgSO_4$-sorbitol
Hsp	Heat-shock protein

IP	Immunoprecipitation
IPTG	Isopropyl-β-D-thiogalactopyranoside
I-TASSER	Iterative threading assembly refinement
iTOL	Interactive tree of life
KD	Knockdown
kDa	kilodalton
K-gluc	Potassium gluconate
KOC1	Regulatory kinase at the outer chloroplast membrane 1
LB	Lysogeny broth
MDa	megadalton
MES	2-(*N*-morpholino)ethanesulfonic acid
MgATP	Magnesium ATP
miRNA	microRNA
MS	Murashige and Skoog medium
M_x	Plant generations after mutagenic treatment
NASC	Nottingham Arabidopsis Stock Centre
NCBI	National Center for Biotechnology Information
NERF	NFYA5 enhancing RING finger
NFYA5	Nuclear factor Y subunit A5
NGS	Next-generation sequencing
OB	Oligonucleotide/oligosaccharide-binding
OD_x	Optical density at a wavelength of x nm
Omp	Outer-membrane protein
OR	Overlapping region
OX	Overexpression

PCR	Polymerase chain reaction
PEB	Protein extraction buffer
PEX	Peroxin
PFA	Paraformaldehyde
PIC	Protease inhibitor cocktail
ppi	Plastid protein import
POTRA	Polypeptide transport-associated
pSSU	Precursor small subunit of Rubisco
qRT-PCR	Quantitative RT-PCR
RING	Really interesting new gene
RNA	Ribonucleic acid
RNF	RING finger
rpm	Revolutions per minute
ROS	Reactive oxygen species
RT-PCR	Reverse transcription polymerase chain reaction
SCB	Sodium cacodylate buffer
SCO	Snowy cotyledon
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEM	Standard error of the mean
siRNA	Short interfering RNA
SNARE	Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
SNPs	Single nucleotide polymorphisms
SP1	Suppressor of ppi1 locus1
SPL	SP1-Like

Stic	Suppressor of tic40
T_n	Generation of transgenic plants
TBE	Tris/Borate/EDTA
TBS	Tris-buffered saline
T-DNA	Transfer DNA
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
TGS	Tris-glycine
TIC	Translocon at the inner envelope membrane of the chloroplasts
TL	Total lysate
TOC	Translocon at the outer envelope membrane of the chloroplasts
UDP	Uridine diphosphate
UGT73C5	UDP-glycosyltransferase 73C5
UPS	Ubiquitin-proteasome system
UTR	Untranslated region
v/v	$\frac{Volume\ (mL)}{Volume\ (mL)}$
VCL1	Vacuoleless1
VPS	Vacuolar protein sorting
WMD	Web microRNA designer
w/v	$\frac{Weight\ (g)}{Volume\ (mL)}$
YFP	Yellow fluorescent protein

TABLE OF CONTENTS

Abstract i

Acknowledgements ii

Statement of authorship iii

Abbreviations iv

Chapter 1 (General introduction) 1

Summary 2

1. General introduction 3

1.1 Chloroplasts and other plastids 3

1.2 The protein import machinery 4

1.2.1 TOC 5

1.2.1.1 The TOC complex 5

1.2.1.2 The TOC receptor proteins 5

1.2.1.3 The TOC translocation channel 6

1.2.1.4 Mutants of the TOC proteins 7

1.2.2 TIC 8

1.2.2.1 The TIC complex 8

1.2.2.2 The 1 MDa TIC complex 9

1.3 Regulation of the TOC machinery by the ubiquitin-proteasome system . 11

1.3.1 The ubiquitin-proteasome system (UPS) 11

1.3.2 The SP1 E3 ligase 12

1.3.3 Modulation of plastid transitions by SP1 13

1.3.4 Involvement in abiotic stress tolerance 13

1.3.5 SP1 homologues in *Arabidopsis* 14

1.3.6 SPL1 15
1.3.7 SPL2 16
1.4 Aim and structure of thesis 18
Chapter 2 (Materials and methods) 20
2. Materials and methods 21
2.1 Introduction 21
2.2 Plant growth conditions 21
2.2.1 Preparation of plant growth media 21
2.3 Plant growth and physiological studies 22
2.3.1 Crossing of *Arabidopsis* plants 22
2.3.2 Chlorophyll measurements 23
2.3.3 Senescence 24
2.4 Phylogenetic analysis 24
2.5 Protein structural prediction 25
2.6 Transmission electron microscopy 25
2.6.1 Sample fixation 25
2.6.2 Dehydration 26
2.6.3 Epoxy resin infiltration 27
2.6.4 Embedding 27
2.6.5 Imaging 27
2.7 DNA and RNA 28
2.7.1 Plant DNA extraction 28
2.7.2 PCR genotyping 28
2.7.3 Agarose gel electrophoresis 29
2.7.4 RNA extraction 30

2.7.5 cDNA synthesis 30
2.7.6 Semi-quantitative RT-PCR 30
2.7.7 qRT-PCR 31
2.8 Bacterial work, gene cloning and plant transformation 31
2.8.1 Preparation of lysogeny broth (LB) medium 31
2.8.2 Gateway cloning 32
2.8.3 Bacterial transformation and plasmid isolation 32
2.8.4 Plant transformation 34
2.9 Analysis of proteins from plants 35
2.9.1 Protein extraction 35
2.9.2 SDS-PAGE 36
2.9.3 Coomassie staining 37
2.9.4 Immunoblotting 37
2.10 Protein import into isolated chloroplasts 38
2.10.1 Preparation of radiolabelled preproteins 38
2.10.2 Chloroplast isolation and protein import 38
2.11 Protoplast transient assays and immunoprecipitation 40
2.11.1 Protoplast preparation from *Arabidopsis thaliana* 41
2.11.2 Protoplast transfection 41
2.11.3 Protein pulldown by immunoprecipitation 42
2.12 *In vitro* autoubiquitination assays 43
2.12.1 Protein expression in bacteria 43
2.12.2 Purification of recombinant proteins 44
2.12.3 Ubiquitination assay 45
2.13 Forward genetic screen 46

2.14 Mapping by sequencing.....47
2.1.1 DNA preparation for Next generation Sequencing (NGS)..... 47
2.1.2 Mapping..... 47
2.15 Tables.....49
Chapter 3 (Results I) 51
Abstract..... 52
3. Characterization of the *SP1-LIKE* genes 53
3.1 Introduction..... 53
3.2 Results..... 55
3.2.1 Overexpression of *SPL2* partially complements the *sp1* mutation 55
3.2.2 Identification of *spl1* mutant lines 55
3.2.3 Generation of *SPL2* knockdown lines by artificial microRNA 56
3.2.4 Genetic interaction analysis with *ppi1* 57
3.2.5 *SPL2* regulates the chloroplast protein import machinery..... 59
3.2.6 Genetic analysis of *SPL2-KD* with other chloroplast protein import mutants..... 60
3.2.7 SPL2 E3 ligase interacts directly with the TOC proteins 61
3.3 Discussion 63
3.4 Figures..... 66
Chapter 4 (Results II) 83
Abstract..... 84
4. Functional analyses of the SP1-Like E3 ligases..... 85
4.1 Introduction..... 85
4.2 Results 86
4.2.1 Phylogenetic analysis of the SP1 homologues..... 86

4.2.2 Mutation of two amino acid residues alters the predicted conformation of SPL1 cytosolic domain 88
4.2.3 SPL1 is a non-functional E3 ligase *in vitro* 89
4.2.4 *SPL1* and *SPL2* are both important during senescence 91
4.2.5 SPL1 negatively influences the ubiquitination activity of SP1 *in vitro* 92
4.2.6 Overexpression of *SPL1* abrogates the effect of *SP1* overexpression on the *ppi1* mutant phenotype 93
4.3 Discussion 94
4.4 Figures 99
Chapter 5 (Results III) 114
Abstract 115
5. New screen for *sp1* enhancers 116
5.1 Introduction 116
5.2 Results 117
5.2.1 Forward genetic screen 117
5.2.2 Generation of the *spec* mutant mapping populations and genomic datasets 118
5.2.3 Computational analysis of the sequence data to identify candidate *spec* loci 119
5.2.3.1 Candidate loci for *spec1* 120
5.2.3.2 Candidate loci for *spec2* 120
5.2.3.3 Candidate loci for *spec3* 121
5.3 Discussion 122
5.4 Figures 128

5.5 Tables 136
Chapter 6 (General discussion) 142
6. General discussion 143
6.1 Summary of key findings 145
6.2 Appraisal of the new regulators of the chloroplast protein import machinery 147
6.2.1 SPL1 147
6.2.2 SPL2 149
6.3 Conclusion and perspective 151
Bibliography 152

Chapter 1

GENERAL INTRODUCTION

Summary

Chloroplasts belong to a diverse group of plant organelles called the plastids, and are mainly responsible for primary production on Earth through photosynthesis. They entered the plant lineage as a result of ancient endosymbiosis between a cyanobacterial organism and a eukaryotic progenitor. The majority of the present-day plastid proteome is now encoded in the nucleus, with the proteins synthesized in the cytosol. The chloroplast thus evolved a machinery, called the translocons at the outer/inner envelope membranes of chloroplasts (TOC/TIC), for transporting these proteins across the double membrane system surrounding each organelle. Recently, a gene called *SUPPRESSOR OF PPI1 LOCUS1* (*SP1*) was reported to encode an E3 ligase of the ubiquitin-proteasome system (UPS). Manipulation of this gene can influence plastid development and homeostasis via protein turnover of the import machinery. Interestingly, two homologues of SP1 (SPL1 and SPL2) exist in the model plant *Arabidopsis thaliana*. Early insight concerning the function of the closest relative of SP1 (SPL1) did not suggest functional redundancy with SP1. The aim of this thesis is, therefore, to expand our knowledge on the regulatory mechanisms controlling the chloroplast protein import machinery by characterizing the SP1 homologues in greater detail, as well as by conducting a new mutant screen for enhancers of *sp1*.

CHAPTER 1: GENERAL INTRODUCTION

1.1 Chloroplasts and other plastids

The emergence of plastids in plants is believed to have involved endosymbiosis, between an ancient photosynthetic prokaryote that is similar to extant cyanobacteria and a eukaryotic progenitor (Reyes-Prieto *et al.*, 2007). The modern organelles still retain a fully functional genome, although it has been greatly reduced to fewer than 100 protein-coding genes (Timmis *et al.*, 2004). It is estimated that around 3000 plastid proteins are encoded in the nucleus and are required to be transported into the chloroplast upon synthesis in the cytosol (Jarvis, 2008; Li & Chiu, 2010).

The chloroplast is a defining feature of plants and various algae, belonging to a diverse group of organelles known as the plastids (Whatley, 1978; Keeling, 2010). Chloroplasts are responsible for carrying out the process of photosynthesis, which relies on the presence of the light-absorbing molecules known as the chlorophylls (Nelson & Ben-Shem, 2004). Photosynthesis is a vitally important chemical process as biomass production in plants and, consequently, on Earth is dependent on it (Evans & von Caemmerer, 2011). Other plastid forms include the starch-rich amyloplasts that play important roles in energy storage and gravitropism, and the carotenoid pigment accumulating chromoplasts which act as attractants in flowers and fruits (Lopez-Juez & Pyke, 2005; Sakamoto *et al.*, 2008). There are also small, undifferentiated plastids present in the meristematic and reproductive tissues of plants, known as the proplastids. These proplastids function as an organelle transmission medium acting between different cell types as well as generations, and

will later differentiate into a more specialized plastid form (Jarvis & López-Juez, 2013).

Aside from photosynthesis, chloroplasts are also an important site for the biosynthesis of many precursor molecules such as certain amino acids, fatty acids, and purine and pyrimidine bases, and starch, as well as for other metabolic processes (Neuhaus & Emes, 2000; Lopez-Juez & Pyke, 2005).

1.2 The protein import machinery

Proteins destined to the chloroplast are typically synthesized in a precursor form containing an N-terminal targeting signal called a transit peptide (Bruce, 2000). The precursor proteins engage with cytosolic sorting factors (Flores-Pérez & Jarvis, 2013) before being transported into the organelle in an energy-dependent manner via what is known as the chloroplast protein import machinery (Li & Chiu, 2010; Strittmatter *et al.*, 2010; Jarvis & López-Juez, 2013; Paila *et al.*, 2015). Protein translocation across the double envelope membrane of the chloroplast is mediated by the TOC (Translocon at the Outer envelope membrane of Chloroplasts) and TIC (Translocon at the Inner envelope membrane of Chloroplasts) (Jarvis, 2008; Li & Chiu, 2010), which has been known as the canonical protein import pathway (Shi & Theg, 2013). There have been several reports of non-canonical protein import pathways (Jarvis & López-Juez, 2013) which will not be discussed due to their non-relevance in this thesis. Once the precursor protein reaches the stroma, the transit peptide is cleaved off by the stromal processing peptidase and the mature protein is folded to its native structure with the help of stromal chaperones (Hsp93, Hsp90, Hsp70 and Cpn60 (Flores-Pérez & Jarvis, 2013; Inoue *et al.*, 2013)) or it engages

with other components for downstream targeting processes (for example, those leading to the photosynthetic thylakoid membranes) (Jarvis, 2008; Paila *et al.*, 2015).

1.2.1 TOC

1.2.1.1 The TOC complex

The TOC and TIC complexes each comprise multiple proteins, with the core components identified for the first time in the 1990s through various biochemical analyses of isolated pea chloroplasts (Jarvis, 2008; Ling *et al.*, 2013). The core TOC complex consists of two receptor proteins, Toc34 and Toc159, and a translocation channel protein with a typical β-barrel structure, Toc75 - the numbers denotes their molecular weight in kDa (Schnell *et al.*, 1997; Jarvis, 2008). The stoichiometry of the TOC components in the order Toc34:Toc75:Toc159 was reported to be 4-5:4:1 (Schleiff *et al.*, 2003) or 3:3:1 (Kikuchi *et al.*, 2006). The differences between these stoichiometric estimates may be due to the application of different experimental techniques or other reasons such as the presence of unidentified components (Kikuchi *et al.*, 2006) or the instability of Toc159 (Schleiff *et al.*, 2003).

1.2.1.2 The TOC receptor proteins

The Toc34 and Toc159 proteins are both integral membrane proteins with C-terminal membrane anchors (M-domain) and a homologous GTPase domain (G-domain) facing the cytosol. They are considered to be receptors for their ability to recognize and bind precursor protein directly (Perry & Keegstra, 1994; Sveshnikova *et al.*, 2000; Smith *et al.*, 2004). While the mechanistic significance of the GTPase cycling of the receptors has been much discussed and debated, it is still not certain how it works (Jarvis, 2008).

In addition to the G-domain, Toc159 possesses an unstable acidic domain (A-domain) at the N-terminus with unclear function. Genetic complementation analysis suggests the presence of the A-domain is dispensable for atToc159 function (Lee *et al.*, 2003), although it has also been proposed to control preprotein specificity (Inoue *et al.*, 2010).

Multiple isoforms of the TOC receptors exist in *Arabidopsis thaliana* (and indeed in other plants, although *Arabidopsis* has been most studied). There are three isoforms of Toc159: atToc120, atToc132 and atToc159, and two isoforms of Toc34: atToc33 and atToc34, each with different recognition specificities ("at" denotes the species) (Jarvis *et al.*, 1998; Bauer *et al.*, 2000). The atToc33 and atToc159 receptors preferentially act in the import of highly-abundant photosynthetic preproteins whereas the atToc34, atToc120 and atToc132 receptors act in the import of less abundant housekeeping preproteins (Gutensohn *et al.*, 2000; Kubis *et al.*, 2003, 2004). The existence of such a multiple receptor isoforms organization enables client-specific import pathways to operate. This arrangement can potentially avoid damaging competition effects between very abundant and scarce preproteins, therefore increasing the efficiency of the protein import process, or it might influence which types of proteins are imported with effects on plastid functions and development (Ling *et al.*, 2013).

1.2.1.3 The TOC translocation channel

The Toc75 translocation channel belongs to a group of Omp85-related proteins (Outer-membrane protein of 85 kDa), a superfamily of β-barrel outer membrane proteins present in all Gram-negative bacteria, mitochondria and chloroplasts (Day *et al.*, 2014). Like other members of the Omp85 family, the Toc75

protein is composed of a conserved C-terminal β-barrel domain that forms the membrane integral portion of the protein, and an N-terminal domain with three polypeptide transport-associated (POTRA) domains facing the intermembrane space (Paila *et al.*, 2015; Chen *et al.*, 2016). Multiple isoforms of Toc75 also exist in *A. thaliana*, with atToc75-III being the major isoform and atToc75-IV being a minor isoform lacking a POTRA domain (the Roman numerals indicate the chromosome on which the isoform is encoded) (Baldwin *et al.*, 2005; Huang *et al.*, 2011). A third *Arabidopsis* Toc75 locus on chromosome I is most likely a pseudogene (Baldwin *et al.*, 2005).

1.2.1.4 Mutants of the TOC proteins

Mutations in the TOC component genes of *Arabidopsis* affect the protein import capacity of plastids for several types of preproteins and the biogenesis of different types of plastids (Kubis *et al.*, 2003, 2004; Constan *et al.*, 2004; Baldwin *et al.*, 2005). In some cases, mutations can result in lethality during embryogenesis or early seedling growth (Kubis *et al.*, 2004; Baldwin *et al.*, 2005). The *plastid protein import 1* (*ppi1*) mutant of *Arabidopsis* was the first protein import apparatus mutant to be isolated and studied (Jarvis *et al.*, 1998). This mutant carries an *Agrobacterial* T-DNA knockout mutation for the Toc34 homologue, atToc33. The *ppi1* mutants display a severe chlorotic phenotype with altered chloroplast ultrastructure and reduced import capacity (Jarvis *et al.*, 1998). Mutants of atToc34 (termed *ppi3*), on the other hand, display significant root growth defects presumably due to defective non-photosynthetic root plastid development (Constan *et al.*, 2004). Several reports indicate that the activities of atToc33 and atToc34 are partially redundant; for example, the recovery of the *ppi1* phenotype through overexpression of atToc34

(Jarvis *et al.*, 1998) and the embryo lethality of the double mutation, *ppi1 ppi3* (Constan *et al.*, 2004).

Unlike the Toc34 type receptors, the null mutant of atToc159 (termed *ppi2*) results in an albino phenotype due to severe disruption in chloroplast development (Bauer *et al.*, 2000; Kubis *et al.*, 2004). The single mutants of atToc132 exhibited only a weakly chlorotic mutant phenotype, while no differences observed in the atToc120 mutant. However when both atToc132 and atToc120 were knocked out together, a severe albino phenotype similar to *ppi2* was observed (Ivanova *et al.*, 2004; Kubis *et al.*, 2004). This indicates redundancy and the importance of the role of the two proteins in chloroplast biogenesis.

Null mutations of the major isoform of atToc75, atToc75-III, cause embryo lethality, clearly indicating the importance of the protein translocating channel of the TOC complex for protein import, even during embryogenesis (Baldwin *et al.*, 2005; Hust & Gutensohn, 2006). The use of inducible knockdown lines of atToc75-III, and a hypomorphic *toc75-III* mutant allele (termed *toc75-III-3*), enabled studies on the role of atToc75 in chloroplast biogenesis beyond embryogenesis. These mutant plants showed reduced chlorophyll levels, abnormal development of plastids, and reductions in the import of photosynthetic and non-photosynthetic preproteins (Stanga *et al.*, 2009; Huang *et al.*, 2011), further underlining the general importance of the channel to the import of many different plastid proteins.

1.2.2 TIC

1.2.2.1 The TIC complex

The identity of the TIC complex is highly disputed for reasons which will be discussed only briefly here, owing to its lack of direct relevance to the work

presented in this thesis. Until recently, the most widely-discussed, proposed subunits of the TIC machinery were Tic110, Tic40, Tic20 and Tic21, along with other associated molecular chaperones such as Tic22 in the intermembrane space and stroma Hsp70, Hsp93, Hsp90 and Cpn60 (Flores-Pérez & Jarvis, 2013; Jarvis & López-Juez, 2013; Paila *et al.*, 2015; Richardson *et al.*, 2017). The Tic110 has been proposed to form a channel (Heins *et al.*, 2002), but is more likely a scaffold for coordinating stromal chaperones (Inaba *et al.*, 2003; Tsai *et al.*, 2013). Tic40 is a co-chaperone component of the import motor, proposed to regulate stromal chaperones (Kovacheva *et al.*, 2004; Chou *et al.*, 2006; Bédard *et al.*, 2007). Tic20 is a polytopic membrane protein that is proposed to form (part of) the translocation channel (Kouranov *et al.*, 1998; Kikuchi *et al.*, 2009). Tic21 has a similar role to Tic20 that it functions as part of the inner membrane protein-conducting channel (Teng *et al.*, 2006; Kikuchi *et al.*, 2009). Mutants of these components lead to defects in chloroplast biogenesis ranging in severity from chlorosis to albinism and, in worst cases, to embryo lethality, providing a clear indication of their importance for the import of preproteins into the chloroplasts (Kovacheva *et al.*, 2004; Su & Li, 2008; Kasmati *et al.*, 2013).

1.2.2.2 The 1 MDa TIC complex

However, the proposed composition of the TIC channel complex was challenged after the identification of a stable 1 MDa multisubunit translocon complex that associates with protein import intermediates at the inner envelope membrane (Kikuchi *et al.*, 2013). This 1 MDa complex contains three newly-identified components: the plastome-encoded Tic214 (ycf1) protein, and the nucleus-encoded components Tic100 and Tic56. In addition to these, the Tic20 (and Tic21) proteins

are present, but the 1 MDa complex was lacking both Tic110 and Tic40. This 1 MDa complex was proposed to constitute the channel-forming TIC translocon, instead of the previously-identified components. Mutants of Tic100 and Tic56 in *Arabidopsis* displayed severe albino phenotypes (Kikuchi *et al.*, 2013), and total disruption/deletion of Tic214 in tobacco plants is lethal (Drescher *et al.*, 2000), indicating the essential role of these components in plant and/or plastid development. Nevertheless, controversial questions remain over the significance of this 1 MDa complex, related to the relationship of this 1 MDa complex with the previously described TIC complex, and in particular concerning the absence (de Vries *et al.*, 2015) of the 1 MDa complex in grasses (Nakai, 2015). It was suggested that a Tic110/Tic40 containing complex might act downstream of the 1 MDa TIC channel complex, possibly functioning as a motor complex to drive protein import at the expense of ATP hydrolysis, by analogy with what is known to occur in mitochondria (Ling *et al.*, 2013; Nakai, 2015).

1.3 Regulation of the TOC machinery by the ubiquitin-proteasome system

1.3.1 The ubiquitin-proteasome system (UPS)

Cellular proteins inside a cell are subjected to various posttranslational modification events as a way to regulate their activity and functions. In addition to that, the protein modification event can also lead to the turnover of the modified protein through a proteolytic system known as the UPS. This is a major protein degradation system in eukaryotic organisms including plants (Vierstra, 2009). Targeting proteins for degradation by this system is achieved by the attachment of highly conserved 76-amino acid ubiquitin molecules through a cascade of three important enzymes: E1 (ubiquitin-activating enzyme), E2 (ubiquitin-conjugating enzyme) and E3 (ubiquitin ligase). The specificity of ubiquitination is determined largely by the E3 ligases, which explains their diversity in number compared to E2 and E1 enzymes (~1400 E3s in *A. thaliana*, compared to only ~40 E2s and two E1s) (Vierstra, 2009). Attachment of the ubiquitin molecule is typically made to a lysine residue in the substrate via an isopeptide bond.

There are four categories of E3 ligases identified in plants based on their mode of action or complex composition. They are HECT (Homologous to the E6-AP Carboxyl terminus), U-box, RING (Really Interesting New Gene), and cullin-RING ligases. The diversity of the E3 ligases mean that they can each take on a specific role, recognizing specific targets, and act to regulate a wide variety of substrates involved in diverse aspects of plant development and function (Vierstra, 2009; Ling & Jarvis, 2015a). Typically, the ubiquitinated proteins then become substrates for the 26S proteasome (a protease complex) in order to be degraded, and the corresponding ubiquitin molecules are then recycled.

1.3.2 The SP1 E3 ligase

A RING-type ubiquitin E3 ligase encoded by *SUPPRESSOR OF PPI1 LOCUS1* (*SP1*) was discovered through a genetic screen in *Arabidopsis* aimed at shedding light on the regulatory mechanisms governing protein import via the TOC complex (Ling *et al.*, 2012). It was observed that *sp1* mutations were able to recover the chlorotic phenotype of *ppi1* plants. The *sp1 ppi1* double mutant chloroplasts displayed an improvement in chloroplast development as a result of better protein import activity. The *sp1* mutation was also able to suppress other TOC mutant phenotypes, such as that of *toc75-III-3*, but not those of TIC component mutations, suggesting specificity of SP1 for proteins on the outer envelope membrane. The E3 activity of the SP1 was shown to be dependent on the presence of an intact C3HC4 motif in the RING finger domain, as disruption of the conserved motif abolished *in vitro* ubiquitination activity (Ling *et al.*, 2012). Confocal microscopy analysis of fluorescent protein fusions, and biochemical analyses of isolated chloroplasts, revealed that SP1 is localized to the outer envelope membrane of the chloroplast. The SP1 protein was also reported to be present in other organelles such as mitochondria and peroxisomes (Pan *et al.*, 2016), although such observations have been disputed as being most likely due to mislocalization artefacts caused by overexpression (Ling *et al.*, 2017). Furthermore, a proposed involvement of SP1 in regulating the peroxisome protein import machinery was also disputed, because of the irreproducibility of critical peroxin (PEX) protein abundance data (Pan *et al.*, 2016; Ling *et al.*, 2017).

1.3.3 Modulation of plastid transitions by SP1

The suppression mediated by the *sp1* mutation in the *ppi1* background was shown to be linked to an increase in the abundance of TOC proteins (Ling *et al.*, 2012). Thus, the loss of SP1-mediated ubiquitination in *sp1* resulted in reduced degradation of TOC components (other than Toc33, which is absent in *ppi1*), and this compensated for the absence of Toc33 through the formation of stabilized, functional TOC complexes (Kessler, 2012). The *sp1* mutant plants displayed a delay in plastid transition processes, during de-etiolation and leaf senescence. Conversely, overexpression of *SP1* resulted in an accelerated rate of these transition processes. During de-etiolation in wild type, the atToc132 and atToc120 proteins were rapidly degraded by SP1, and replaced with atToc159 (Ling *et al.*, 2012). This demonstrated a requirement for a suitable TOC complex configuration for the import of photosynthetically related proteins upon the illumination of dark-gown plants. Not only that, this finding supported the hypotheses on client specificity of different TOC receptors that were discussed earlier (see Section 1.2.1.2), and, it showed that there is a requirement for TOC reorganization for plastid proteome changes to occur. This also suggested a broader relevance of SP1, with involvement in other plastid transition process such as the conversion of chloroplasts to chromoplasts during food ripening in crops like tomatoes.

1.3.4 Involvement in abiotic stress tolerance

A follow-up study on the role of SP1 focused on its involvement in plant responses to abiotic stress. *Arabidopsis* plants overexpressing *SP1* display a greater tolerance to high salinity, osmotic and oxidative stress conditions than wild type, as demonstrated by measuring seedling greening, development and survival (Ling &

Jarvis, 2015b). In contrast, under such conditions, the mutants of *sp1* displayed impaired tolerance linked to a greater accumulation of reactive oxygen species (ROS), which can be detrimental to cellular processes; ROS can do damage to many cellular components by reacting with proteins, DNA and lipids (Saibo *et al.*, 2009). In order to avoid the harmful effects of ROS overproduction, plants adapt to environmental stresses by reducing their photosynthetic activity, which can be brought about by changes in the chloroplast proteome (Saibo *et al.*, 2009). Indeed, this is what SP1 was proposed to be involved in: combating the overaccumulation of ROS during unfavourable growth conditions. During stress treatments, it was observed that the TOC abundance was specifically depleted, and that this activity is dependent on SP1. The data were correlated with changes in the import rate of photosynthesis-related proteins (Ling & Jarvis, 2015b). Therefore, it was proposed that SP1 lowers ROS production by attenuating the import of photosynthetic proteins through the turnover of the relevant TOC receptors, and consequently reducing photosynthetic activity to a level that is optimal for the conditions.

1.3.5 SP1 homologues in *Arabidopsis*

In the report by Ling *et al.* (2012), two homologues of *SP1* were identified in *A. thaliana*: *SP1-Like1* (*SPL1*) and *SP1-Like2* (*SPL2*). They share sequence identity (60.3% for SPL1; 20.5% for SPL2) and topological similarity with SP1, but it was unknown whether they participated in the same regulatory pathway. Just like SP1, both homologues were observed to reside in the chloroplast envelope via confocal microscopy analysis of YFP fusion proteins (Ling *et al.*, 2012).

Although the aforementioned original data showing chloroplast localization were clear, some conflicting reports have since appeared in the literature. For

example, the two homologues were said to be present in other organelles such as the peroxisomes and mitochondria (Pan *et al.*, 2016, 2018). In another study, the SPL2 protein was reported to have been observed in the nucleus of plant roots (Gao *et al.*, 2015). As was noted earlier in relation to SP1 (Section 1.3.2), such discrepancies could be related to mislocalization artefacts caused by overexpression, especially as none of the figures in the original datasets suggested functions elsewhere in the cell (Ling *et al.*, 2012, 2017). Even if the proposed multiple organelle localization was not an artefact, the significance of the unusual partial peroxisome localization would be debatable for a reason that will be discussed in the following section. As for the proposed nuclear localization of SPL2, the failure to include control experiments and the dubious choice of plant anatomy for imaging draw the results into question, especially considering that the proposed physiological function in that study involved the stomata (Gao *et al.*, 2015).

1.3.6 SPL1

The SPL1 protein was reported to be a close homologue of SP1, and the initial study on its possible involvement in chloroplast biogenesis was investigated by conducting a genetic complementation experiment (Ling *et al.*, 2012): the overexpression of *SPL1* did not complement the *sp1 ppi1* double mutant, indicating that SPL1 is not redundant to SP1. This suggested that SPL1 might have different targets to SP1, but its function remained to be investigated.

The report by Pan *et al.* (2018) suggested that SPL1 has an antagonistic effect on SP1, and that SP1 and SPL1 are able to promote each other's destabilization via unknown mechanisms. The significance of such an SP1 and SPL1 destabilization mechanism in the peroxisomes would be difficult to comprehend

considering that SPL1 is proposed to be only partially localized to the organelle and only in a small number of the organelles (Pan *et al.*, 2016, 2018).

1.3.7 SPL2

The SPL2 protein is a more distantly related homologue of SP1 and not much work has been done to characterize its role in the context of chloroplast biogenesis. However, in 2015, Gao *et al.* presented a paper on a gene called *NERF* (*NFYA5 Enhancing RING FINGER*) that encodes a RING E3 ligase (which is *SPL2*). The expression of this gene was proposed to contribute to drought resistance in plants. According to the authors, the 3' untranslated region (UTR) of the *NERF* (*SPL2*) gene possesses an overlapping region (*OR*) which produces a short interfering RNA (siRNA). This siRNA can positively affect the transcript level of the adjacent drought-stress inducible transcription factor gene, *NFYA5*, by antagonizing the expression level of the inhibitory microRNA, miR169 (Gao *et al.*, 2015).

This model was supported by the results of several experiments carried out in various knockdown and overexpressor mutant lines for *NERF*, *OR* and *NFYA*, by measuring the transcript levels of the three loci to assess the associations between them. Knockdown of *NERF* resulted in a reduction in *NFYA5* transcripts, whereas overexpression of *NERF* or *OR* resulted in an elevated level of *NFYA5* transcript (Gao *et al.*, 2015). On the other hand, the expression of *NFYA5* had no effect on the level of *NERF* transcripts. Knockdown of *NERF* resulted in plants being drought sensitive, whilst *NERF* (containing the *OR*) overexpressor lines were more drought tolerant (Gao *et al.*, 2015).

Although strong evidence of the relationship between the *OR* containing transcripts and the transcript level of *NFYA* were presented, the authors failed to

include appropriate controls (overexpression of RING mutant version of *NERF/SPL2* with the *OR*) in their drought experiments. This raises a doubt about their claim concerning the functional significance of the *OR* within the *NERF* transcripts for plant drought tolerance. Therefore, it is unclear whether the reported drought-tolerant phenotype was due to the *OR*, as proposed, or instead to the NERF/SPL2 protein function as an E3 ligase.

1.4 Aim and structure of the thesis

The aims of this project were to: (1) elucidate the roles of two putative regulators of chloroplast biogenesis in *Arabidopsis thaliana*, namely the *SPL1* and *SPL2* homologues of *SP1*; and, (2) conduct a new genetic screen for enhancers of *sp1*, in order to build upon what was already known about SP1 and identify new regulators or components of the chloroplast protein import machinery. Various approaches were undertaken in pursuing the aims of the thesis, including various molecular biology techniques, genetics, protein biochemistry, microscopy, and bioinformatics. The key research questions addressed in each of the results chapters are described below.

Chapter 3 builds on the foundations of the published genetic studies to elucidate the roles of *SPL1* and *SPL2* in chloroplast biogenesis. A reverse genetics approach was carried out, and the mutants identified were characterized in detail. In particular, they were crossed to various chloroplast protein import mutants, and the changes in chlorophyll level in the mutants were recorded. To follow up, a perturbation to the chloroplast protein import was investigated, and abundance of the components of import machinery analyzed. Immunoprecipitation assays were then conducted to confirm the interacting partners of the relevant E3 ligase. The results were interpreted in order to elucidate the molecular mechanisms of SPL function.

Chapter 4 presents a more advanced analysis of the functions of the SPL proteins. I explored the evolutionary relationships among the SP1 homologues across the Viridiplantae, measured the E3 ligase activity of the *Arabidopsis* proteins *in vitro* in order to test their functionality, and investigated their physiological relevance by analysing the dynamic process of leaf senescence in various *Arabidopsis* genotypes.

Chapter 5 describes the screen for new *sp1* enhancer mutants, and presents analysis of the putative mutants of interest. Analysis of Next-Generation Sequencing (NGS) data derived from the mutants using a computational approach was employed to fine map the causative genetic lesion in each case. Based on the data, a number of candidate genes were identified for each mutant. For each of these, I sought further information from the literature and from data repositories in order to assess the likelihood of it being responsible for the relevant enhancer phenotype.

Chapter 2

MATERIALS AND METHODS

CHAPTER 2: MATERIALS AND METHODS

2.1 Introduction

This chapter presents the research methodology, growth conditions, materials, chemical reagents and instruments used to carry out the study.

2.2 Plant growth conditions

All *Arabidopsis thaliana* plants used in this study were of the Columbia-0 (Col-0) ecotype. For most of the analysis, the plants were grown in a chamber (Percival Scientific, Inc.) with ~100 μmol m^{-2} s^{-1} white light under a long-day photoperiod (16 hours light, 8 hours dark). The temperature was maintained at 20°C with 70% relative humidity.

A greenhouse or controlled environment rooms were only used to grow the plants for propagation purposes under standard growth conditions at 22-24°C with a long-day photoperiod 120 μmol m^{-2} s^{-1} light (16 hours light, 8 hours dark).

2.2.1 Preparation of plant growth media

Seeds were either germinated on Petri dishes containing 0.65% Murashige and Skoog (MS) medium (4.3 g/L MS basal salt mixture, 0.5% (w/v) sucrose, 0.05% (w/v) 2-(*N*-morpholino)ethanesulfonic acid (MES), pH adjusted to 5.7 with potassium hydroxide (KOH); phyto-agar was added to a concentration of 0.65% (w/v) before sterilization by autoclaving at 110°C for 15 minutes), or on soil (Levington Advance Seed and Modular F2). The following antibiotics were only included in the MS

medium for the selection of plants carrying transgene(s): 10 μg/mL DL-phosphinotricin and 200 μg/mL cefataxime.

For *in vitro* growth, seeds were surface sterilized for 5 minutes with 70% (v/v) ethanol and 0.05% (v/v) Triton X-100, followed by an additional 10 minutes with 100% (v/v) ethanol before sowing. The 100% (v/v) ethanol was evaporated by drying the seeds on a sterile filter paper in the flow hood. Sowed seeds were stratified in the dark at 4°C for a minimum of 2 days.

2.3 Plant growth and physiological studies

2.3.1 Crossing of *Arabidopsis* plants

Plants were grown on the soil as described in Section 2.2. Typically, a homozygous mutant with clear, recessive phenotype was designated as female acceptors. This eases the identification of self-pollinated progeny in the F_1, based on the appearance of the mutant phenotype. Transgenic plants containing resistance markers were used as male donors. In this case, selection by antibiotic resistance could be carried out by screening the F_1 progeny on MS media supplemented with the appropriate antibiotics.

Before crossing, all mature siliques and fully opened flowers of the acceptor plant were removed using fine scissors. Emasculation of the largest unopened flowers was then performed using fine-tipped forceps to remove all sepals, petals and stamens. An opened and mature flower from the donor plant was selected and used to pollinate the already emasculated acceptor flower by covering the stigma with pollen grains as much as possible. This procedure was conducted under a dissecting microscope to ensure the acceptor flower was originally free from

contaminating pollen grains, and that the stigma was covered with the desired pollen grains. The pollinated inflorescence was marked and allowed to mature.

2.3.2 Chlorophyll measurements

The chlorophyll content of 10-day-old seedlings was measured using a spectrophotometer following extraction with *N,N'*-dimethylforamide (DMF). The moisture on the surface of 10 seedlings (per genotype) grown in vitro was removed gently with paper wipes (Kimtech Science) before weighing them in a 1.5 mL centrifuge tube. Then, 1 mL of DMF was added and left to incubate overnight in the dark at 4°C on an orbital shaker. The absorbance of each sample was measured in a quartz cuvette (Hellma Analytics) for the wavelengths 646 nm and 664 nm using an Ultrospec 100 Pro visible spectrophotometer (Amersham Biosciences). The absorbance readings were ideally between 0.01 and 0.8; where the readings exceeded the ideal range, samples were diluted further in DMF. The absolute chlorophyll concentration in nmol/mg fresh weight was calculated using published equations (Porra *et al.*, 1989), taking into account the dilution factor:

$$Chlorophyll\ a = \frac{(13.43\ \times\ A663.8) - (3.47\ \times\ A646.8)}{fresh\ weight}\ \times\ dilution\ factor$$

$$Chlorophyll\ b = \frac{(22.9\ \times\ A646.8) - (5.38 - A663.8)}{fresh\ weight}\ \times\ dilution\ factor$$

$$Chlorophyll\ a + b = \frac{(19.43\ \times\ A646.8) + (8.05\ \times\ A663.8)}{fresh\ weigh}\ \times dilution\ factor$$

For mature plants (> 4 weeks old), the chlorophyll contents of the leaves were measured using a Konica-Minolta SPAD-502 meter according to the manufacturer's

guidelines. The SPAD values were converted into nmol/mg fresh weight using a calibrated equation (Ling *et al.*, 2011).

2.3.3 Senescence

The leaves of equally developed 28-day-old plants were wrapped with aluminium foil to induce senescence. Whilst covered, the plants were left to grow in a chamber under standard growth conditions for 5 days. The chlorophyll contents of the leaves were measured using a Konica-Minolta SPAD-502 meter (Section 2.3.2) and the resultant photosynthetic performance (F_v/F_m) was recorded using a CF Imager (Technologica).

2.4 Phylogenetic analysis

BLAST (Basic local alignment search tool) searches using *Arabidopsis thaliana* SP1, SPL1 and SPL2 amino acid sequences as queries were performed at, and datasets were downloaded from, both Phytozome (Goodstein *et al.*, 2012) and Ensembl Plants (Zerbino *et al.*, 2018). Plant species with truncated sequences in the highly-conserved C3HC4 RING domain of the E3 ligases were discarded and, if available, replaced by full-length sequences from BLAST searches conducted at the National Center for Biotechnology Information (NCBI). The sequences were aligned using MergeAlign (Katoh *et al.*, 2005; Collingridge & Kelly, 2012) and phylogeny was inferred by maximum likelihood using the IQ-TREE (Trifinopoulos *et al.*, 2016) web server. The resultant trees were visualized and annotated using interactive Tree Of Life (iTOL) version 4 (Letunic & Bork, 2016).

2.5 Protein structural prediction

The structures of the cytosolic domain of the E3 ligases were predicted by submitting the corresponding amino acid sequences to I-TASSER (Iterative Threading Assembly Refinement) online server (Zhang, 2008; Roy *et al.*, 2010). Predicted models with the highest confidence scores (*C*-scores) were selected and visualized using PyMOL (https://pymol.org/2/)

2.6 Transmission electron microscopy (TEM)

Samples were processed by Dr Errin Johnson at the Sir William Dunn School of Pathology, University of Oxford, as described previously (Hyman & Jarvis, 2011) with modifications presented in the following section.

2.6.1 Sample fixation

For primary fixation, 4-week-old *Arabidopsis* leaves were removed from the plants, and then 2 mm discs from the distal region of the leaf (next to the midvein) were excised using a biopsy punch and cut in half with a razor blade in a droplet of fixative (2.5% (v/v) glutaraldehyde and 4% (v/v) paraformaldehyde (PFA) in 0.1 M sodium cacodylate buffer (SCB) at pH 6.9). This was done for 5 leaves from different plants for each genotype, and the work was done as quickly as possible. Leaf pieces were transferred to plastic baskets and fixed in a Leica AMW microwave processing unit for a total of 10 minutes at 37°C with five alternating 2-minute cycles of 20 W and 0 W power. Leaves were then transferred to glass vials over-filled with fixative so that there was no air in the vial when capped, and the vials were shaken vigorously until the leaf pieces sunk to the bottom. Samples were fixed further at room temperature for ~5 hours and then at 4°C overnight.

On the next day, samples were washed twice for 5 minutes with 0.1 M SCB pH 6.9, and then transferred to 2 mL centrifuge tubes. They were then washed twice (for 45 minutes, then 30 minutes) in SCB, once (for 30 minutes) in 50 mM glycine in SCB, and finally once (for 20 minutes) in SCB. All washing steps employed rotation.

For secondary fixation, samples were incubated in 1% (v/v) osmium tetroxide and 1.5% (w/v) potassium ferrocyanide in SCB, first for 2 hours at 4°C with vigorous shaking, and then for a further 50 minutes at room temperature with rotation. Samples were washed once for 5 minutes, once for 1.5 hours, and three times for 15 minutes, each time with Milli-Q water and at room temperature with rotation.

For tertiary fixation, samples were incubated in 2% uranyl acetate (w/v) at room temperature in the dark with rotation. Samples were then washed with Milli-Q water three times over 15 minutes in total.

2.6.2 Dehydration

After washing, samples were incubated at room temperature with rotation in ice-cold 10%, 20%, 30% and 40% (v/v) ethanol, consecutively for 10 minutes each, followed by further incubations at room temperature in 50% (v/v) ethanol for 20 minutes, 60% (v/v) ethanol for 10 minutes, and then 70% (v/v) ethanol overnight at 4°C, again with rotation at each step. On the next day, samples were incubated with rotation at 4°C in ice-cold 80%, 90%, 100% (v/v) ethanol and 100% (v/v) anhydrous ethanol, consecutively for 30 minutes each, and then for 30 minutes in ice-cold acetone at room temperature with rotation.

2.6.3 Epoxy resin infiltration

All resin steps were performed at room temperature with rotation and samples were centrifuged for 30 seconds at 2000 revolution per minute (rpm) in a microcentrifuge after each solution change to aid resin infiltration. Samples were incubated with 25% (v/v) resin (TAAB Hard Plus epoxy resin) in pure acetone for 2 hours, and then 50% (v/v) resin in acetone for 3 hours, 75% (v/v) resin in acetone for 1 hour and 100% resin overnight, with each step being conducted in a fume hood with rotation.

On the following day, changes of the 100% (v/v) resin with centrifugation at each step were performed twice, at 11:30 am and 5 pm; and then the samples were left overnight.

On the last day, a resin change was carried out at 10:30 am with centrifugation, and then embedding was conducted in the afternoon.

2.6.4 Embedding

Flat dish (coffin) embedding moulds were prepared by pre-polymerizing a thin layer of resin in each mould for ~2 hours in the 60°C oven, before adding the leaf pieces (which were pushed using a disposable plastic transfer pipette to the end of the coffin so that the razor-cut surface of the disc could be sectioned transversely) and then covering with fresh resin. Blocks were polymerized in the oven for 48 hours at 60°C.

2.6.6 Imaging

Samples were imaged using a Technai 12 TEM. At least 10 different whole chloroplast images per sample were captured - a total of 30 chloroplasts per

genotype, 10 from each of the three different plants. All images were taken at the same magnification level. Total chloroplast cross-sectional area was measured using Fiji, a distribution of the open-source software ImageJ (Schindelin *et al.*, 2012).

2.7 DNA and RNA

2.7.1 Plant DNA extraction

A medium-sized rosette leaf cut from a ~4-week-old plant of each genotype of interest was collected into a 1.5 mL centrifuge tube containing a 2.5-3.5 mm glass bead, and flash-frozen in liquid nitrogen. The leaf tissue was homogenized to a powder using a Qiagen TissueLyser in a pre-cooled block for 1 minute at 30 Hz. The tubes were centrifuged briefly at maximum speed in a microcentrifuge to collect the powder, and then 0.5 mL of Extraction buffer (200 mM Tris-HCl, pH 7.5, 250 mM sodium chloride (NaCl), 25 mM ethylenediaminetetraacetic acid (EDTA), 0.5% (w/v) sodium dodecyl sulphate (SDS)) was added. Samples were vortexed for ~10 seconds and then centrifuged for 1 minute, again at maximum speed. The supernatant was then mixed with isopropanol at a 1:1 ratio in a fresh tube and left to precipitate at room temperature for 2 minutes, before centrifugation again, for 5 minutes at maximum speed. Each pellet was washed briefly with 70% (v/v) ethanol and left to air dry at 37°C. Once dried, the DNA pellet was resuspended in 50 μL of Milli-Q water and stored at -20°C.

2.7.2 PCR genotyping

The mutant alleles used in this study are listed in Table 2.1. All *spl1* T-DNA insertion lines were obtained from the Nottingham Arabidopsis Stock Centre (NASC). The *spl1-1* (salk_064720) mutant was obtained from the SALK population, and all

other mutants have been characterized previously (Table 2.1). Polymerase Chain Reaction (PCR) genotyping using standard reaction mixtures (10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride (KCl), 1.5 mM magnesium chloride ($MgCl_2$), 200 μM deoxyribonucleoside triphosphates (dNTPs) mix, 0.25 μM of each primer, and 0.5 units of DNA polymerase) was performed using a Biometra T3 Thermocycler to identify homozygous plants using gene-specific primers and T-DNA primers (Tables 2.1 and 2.2, Jarvis *et al.*, 1998; Kovacheva *et al.*, 2004; Basnayake *et al.*, 2011; Ling *et al.*, 2012). Derived Cleaved Amplified Polymorphic Sequences (dCAPS) analysis was employed to detect Single Nucleotide Polymorphisms (SNPs), which involved designing a primer containing one or more mismatches to the template DNA in order to generate a variable restriction site dependent on the polymorphism (Table 2.1, Neff *et al.*, 1998; Stanga *et al.*, 2009; Ling *et al.*, 2012; Woodson *et al.*, 2015). Presence of the introduced restriction site in the resultant PCR products, which depended on presence of the polymorphism of interest, was tested for by restriction enzyme digestion (Table 2.1, Neff *et al.*, 1998; Stanga *et al.*, 2009; Ling *et al.*, 2012; Woodson *et al.*, 2015).

2.7.3 Agarose gel electrophoresis

Electrophoresis was performed in tanks (Mini-Sub Cell GT, Bio-Rad) containing 0.5X Tris/Borate/EDTA (TBE) buffer at ~10 V/cm. PCR amplicons were loaded in 1X Orange G loading buffer (0.1% (w/v) Orange G, 50% (v/v) glycerol) and resolved by running on 1% (w/v) agarose gels (or 3% (w/v) for dCAPS samples) in 0.5X TBE buffer (40 mM Tris-HCl, pH 8.3, 45 mM boric acid, 1 mM EDTA) containing 1X SYBR Safe DNA gel stain (Invitrogen). The bands were visualized using GeneFlash (Syngene) or LAS-4000 (GE Healthcare) imagers.

2.7.4 RNA extraction

A pool of ten 2-week-old homozygous seedlings grown in vitro, or ~1 g of rosette leaf sample cuttings from a 4-week-old individual plant, were collected into a 2 mL centrifuge tube containing a 5 mm steel bead (Qiagen). Samples were flash frozen in liquid nitrogen before homogenizing into a powder as described in Section 2.7.1.

RNA was isolated using a Spectrum Plant Total RNA Kit (Sigma) with the remaining DNA removed by On-Column DNase I Digestion Set (Sigma). The isolated RNA was quantified using a NanoPhotometer P330 (Implen).

2.7.5 cDNA synthesis

A total of 2 μg of isolated RNA from Section 2.7.4 mixed with 1 μl of 100 μM of Oligo-d$(T)_{20}$ primer was used as the starting material for complementary DNA (cDNA) synthesis. The cDNA was synthesized using a SuperScript IV Reverse Transcriptase (Invitrogen) system, according to manufacturer's instructions. Samples were incubated at the recommended temperature using Biometra T3 Thermocycler.

2.7.6 Semi-quantitative RT-PCR

Semi-quantitative reverse transcription polymerase chain reaction (RT-PCR) was performed using primers specific for the target gene, and control gene primers for *EUKARYOTIC TRANSLATION INITIATION FACTOR 4E-1* (*eIF4E1*) (Table 2.3, (Kovacheva *et al.*, 2004)). A low number of PCR cycles was employed (23 - 28 cycles) to avoid oversaturation of product. The PCR amplicons were resolved on 1% agarose gels as described in Section 2.7.3. Band intensities were quantified using Aida Image Analyzer software (Raytest).

2.7.7 qRT-PCR

Quantitative RT-PCR (qRT-PCR) primer pairs (Table 2.4) that produce PCR products less than 100 bp were designed using OligoArchitect Online (Sigma). The PCR efficiency of the primer pairs was confirmed by qPCR analysis of serial dilutions of wild-type cDNA (Pfaffl, 2001). The analysis was performed using primers specific for the target gene, and control gene primers for *ACTIN2*.

The cDNA prepared from Section 2.7.5 was diluted to a working concentration of 5 ng/μL in a 20 μL qPCR reaction mix (1X qPCRBIO SyGreen Mix Hi-ROX (PCR Biosystems Ltd.) and 400 nM of each primer). The mixture was placed in a MicroAmp Optical 96-Well Reaction Plate (Applied Biosystems) and sealed with MicroAmp Optical Adhesive Film (Applied Biosystems). The 96-well plate was run and analyzed using a StepOnePlus Real-Time PCR System (Applied Biosystems). The settings on the instrument were programmed according to the conditions recommended in the guidelines for the qPCRBIO SyGreen Mix Hi-ROX reagent.

2.8 Bacterial work, gene cloning and plant transformation.

2.8.1 Preparation of lysogeny broth (LB) medium

Liquid LB medium was prepared by dissolving 10 g of bacto-tryptone, 5 g of yeast extract, and 5 g of NaCl in 1 L of Milli-Q water. The medium was sterilized by autoclaving at 110°C for 15 minutes. For LB agar plates, 1% (w/v) of agar was added to liquid LB medium before autoclaving. Appropriate antibiotics were added to the medium for selection of bacteria containing the plasmids of interest: 100 μg/mL ampicillin, 10 μg/mL gentamycin, 50 μg/mL kanamycin and 50 μg/mL spectinomycin.

2.8.2 Gateway cloning

All primers used for cloning are listed in Table 2.5. The forward Gateway *attB1* sequences 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTCC-3' or 5'-AAAAAGCAGGCTCC-3' and reverse Gateway *attB2* sequences 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTT-3' or 5'-AGAAAGCTGGGTT-3' to be added to the gene fragment was carried out using modified gene specific PCR primers. Full-length cDNA or coding sequences (CDS) were amplified by PCR from wild-type *Arabidopsis* first-strand cDNA, or from previously generated binary overexpression vector containing the gene of interest (Ling *et al.*, 2012). Where necessary, 1 M betaine was added into the PCR reaction mixture to reduce the annealing temperature. The RING mutants of *SPL1* and *SPL2* were generated by combined overlap-extension PCR *in vitro* (Hussain & Chong, 2016). A construct to knockdown *SPL2* expression by an artificial microRNA (amiRNA) approach was generated by replacing the original miR319a sequence in the pRS300 plasmid as described previously (Schwab *et al.*, 2006). The primers for the *SPL2* amiRNA construct were designed using the Web MicroRNA Designer (WMD) platform (Ossowski *et al.*, 2008). PCR amplicons were resolved by agarose gel electrophoresis and bands were purified using GenElute Gel Extraction Kit (Sigma). All gene cloning steps were carried out using the Gateway cloning system (Invitrogen) according to the manufacturer's instructions.

2.8.3 Bacterial transformation and plasmid isolation

Transformation of 25 μL of chemically competent *Escherichia coli* (α-Select Silver Efficiency Competent Cells, Bioline) was carried out with 2.5 μL of the Gateway plasmid reaction using the following heat-shock method. The mixture was

left to incubate on ice for 30 minutes, placed in a 42°C water bath for 45 seconds, and then placed back on ice for 2 minutes. Transformation reactions were diluted with 1 mL of LB medium and left to incubate on a 37°C shaker for 1 hour. After incubation, 100 μL of the transformation mixture was plated on LB agar plates containing the appropriate antibiotic for selection. The colonies were allowed to grow overnight at 37°C.

On the next day, PCR was performed on a number of colonies growing on the LB agar plate to test for transformants carrying the cloned gene. Colonies were picked using sterile pipette tips into prepared PCR reaction mixtures containing primers specific to the cloned gene (Section 2.7.2). The colonies that tested positive for the cloned genes were each used to inoculate 5 mL LB medium containing antibiotic and grown up overnight at 37°C with shaking. The cells were pelleted by centrifugation at maximum speed for 1 minute in a microcentrifuge, and plasmid DNA was isolated using an Isolate II Plasmid Mini Kit (Bioline). All donor vectors were verified by Sanger DNA sequencing (Source BioScience).

Sequences to be used for the generation of transgenic overexpression lines were cloned into binary overexpression vector pB2GW7 (Karimi *et al.*, 2002). To generate transgenic plants for protein pulldown experiments by immunoprecipitation (IP), the *SPL2* CDS was cloned into the pK7YWG2 binary vector (Karimi *et al.*, 2002). Sequences for protein expression in a bacterial system (*SPL1cyt*, *SPL1cyt-H308Y*, *SPL2cyt* and *SPL2cyt-H348Y*) were cloned into N-terminal GST fusion protein expression vector pDest-565 (Addgene).

2.8.4 Plant transformation

The transgenic plants used in this study were generated by *Agrobacterium*-mediated transformation by the floral dip method (Clough & Bent, 1998). The constructs in the pB2GW7 vectors were first transformed into *Agrobacterium tumefaciens* GV3101 as follows using freeze-thaw method (Hofgen & Willmitzer, 1988). About 1 μg of plasmid DNA was added to 100 μL of home-made *Agrobacterium* competent cells (Hofgen & Willmitzer, 1988) in a 1.5 mL centrifuge tube. The transformation mixture was frozen in liquid nitrogen and allowed to thaw at 37°C for ~5 minutes. Once thawed, the transformation mixture was diluted with 1 mL of liquid LB medium and left to incubate on a shaker at 28°C for 2-4 hours, prior to centrifugation at 5000 rpm for 30 seconds in a microcentrifuge to pellet the cells. The supernatant was discarded by decanting into a waste bottle, leaving approximately 100 μL of LB behind. The cells were resuspended in the remaining solution and plated on an LB agar plate containing the appropriate antibiotics (Section 2.2.1) and left to grow at 28°C for 2 days.

A colony was then selected and grown up in liquid LB containing antibiotics, overnight at 28 °C with shaking at 240 rpm. The cells were harvested by centrifugation at 5000 rpm for 10 minutes in a microcentrifuge, and then resuspended in 5% (w/v) sucrose to a final optical density at a wavelength of 600 nm (OD_{600}) reading of 0.8.

One week before dipping, the primary inflorescences of the *Arabidopsis* plants to be transformed were clipped to encourage the growth of multiple secondary inflorescences. Mature siliques and fully-opened flowers were removed prior to dipping. *Arabidopsis* plants were dipped for 20-30 seconds into the 5% (w/v) sucrose *Agrobacterium* solution supplemented with 0.05% (v/v) silwet L-77. Dipped plants

were allowed to recover in the greenhouse under high humidity by covering them with a propagator lid for one day, and then left to mature uncovered thereafter.

Typically, around 10 T_1 lines were analysed per construct. Single-locus lines showing a 3:1 ratio of antibiotic resistance to antibiotic sensitivity in the T_2 generation were selected, and those with good overexpression or knockdown of the target gene were carried forward until they were homozygous in the T_3 generation.

2.9 Analysis of proteins from plants

2.9.1 Protein extraction

Seedlings grown *in vitro* for 2 weeks, or rosette leaves from 4-week-old plants grown on soil, were collected and flash frozen in liquid nitrogen. Samples were homogenized into a powder with the aid of metal beads as in Section 2.7.1. For every extraction, 1 mL of protein extraction buffer (PEB) was prepared from a 1X stock solution (100 mM Tris-HCl, pH 6.8, 10% (v/v) glycerol, 0.5% (w/v) SDS, 0.1% (v/v) Triton X-100, 5 mM EDTA) containing fresh 10 mM dithiothreitol (DTT) and 10 μL Protease Inhibitor Cocktail (PIC; Sigma). After homogenization of the plant material, 200 μL of cold PEB was added to the powder and the mixture was vortexed for 30 seconds. Samples were centrifuged at maximum speed for 10 minutes at 4°C in a microcentrifuge to remove the insoluble material, and the supernatant was transferred into a fresh 1.5 mL centrifuge tube for analysis.

The protein concentration of each extract was quantified by using a Bradford Assay. A linear standard curve of 1, 2, 4, 8 and 10 μg/μl was generated using Bovine Serum Albumin (BSA) as the protein reference. Each protein and dye mixture was prepared in a total volume of 1 mL, consisting of 1 μL protein sample, 799 μL Milli-Q

water, and 200 μL Protein Assay dye (Bio-Rad). Absorbance was measured at OD_{595} using a NanoPhotometer P330 (Implen).

2.9.2 SDS-PAGE

The protein samples extracted in Section 2.9.1 were analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Various percentage concentrations (10-15%) of the SDS-PAGE gel were prepared by adjusting the volume of acrylamide solution accordingly. For example, to prepare a 10% acrylamide gel, 10% (v/v) of Protogel (Geneflow) was mixed with 1X Resolving buffer (0.375 M Tris-HCl, pH 8.8, 0.1% (w/v) SDS), and then 0.5% (w/v) ammonium persulphate (APS) and 0.05% (v/v) Tetramethylethylenediamine (TEMED) were added to promote polymerization. This mixture was loaded into an SDS-PAGE casting assembly (Mini Protean III, Bio-Rad) to a height equivalent to ~75% of that of the assembly, and overlaid with water to level the surface and ensure thorough polymerization. The gel was left to solidify for ~1 hour at room temperature. Once hardened, the water was poured out and replaced with high-acrylamide stacking gel (5% (v/v) Protogel, 1X Stacking buffer (0.125 M Tris-HCl pH 6.8, 0.1% SDS), 0.5% APS, 0.05% TEMED). A well-forming comb was inserted immediately after pouring, and the gel was allowed to set for ~15 minutes at room temperature.

Prior to loading into the wells, protein samples (typically 1-5 μL containing 10-20 ng of protein) were each mixed with an equal volume of 2X protein loading buffer (0.1 M Tris-HCl, pH 6.8, 2% (w/v) SDS, 20% (v/v) glycerol, 0.02% (w/v) bromophenol blue) supplemented with fresh 0.1 M DTT, and boiled at 100°C for 5 minutes.

Gels were run in Tris-glycine (TGS) buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS) at 180 V (~30 V/cm) until the dye front reached the bottom.

2.9.3 Coomassie staining

To visualize the separated protein samples, SDS-PAGE gels were incubated in InstantBlue Protein Stain (Expedeon) for 15 minutes at room temperature with gentle shaking.

2.9.4 Immunoblotting

For immunoblotting experiments, the separated proteins in a polyacrylamide gel were transferred to Amersham Protran 0.2 μm NC Nitrocellulose Blotting Membrane (GE Healthcare) by electroblotting using a Mini Trans-Blot Module (Bio-Rad). Transfer was done in ice-cold transfer buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS, 20% (v/v) methanol) at 100 V for 1 hour (or for 1.5 hours with gels that were thicker than 0.75 mm).

Once transfer was completed, the membrane was blocked for 30 minutes at room temperature by gentle shaking in blotto comprising Tris-buffered saline (TBS)-Tween (10 mM Tris-HCl pH 8, 150 mM NaCl, 0.05% (v/v) Tween 20) containing 8% (w/v) skimmed milk powder. After blocking, the membrane was incubated with primary antibody diluted in 15 mL blotto overnight on a shaker at 4°C. On the next day, the membrane was rinsed with TBS-Tween and washed two times with gentle shaking at room temperature, for 5 minutes each, with TBS-Tween. After washing, the membrane was incubated with a secondary antibody selected according to the organism in which the primary antibody was raised (typically goat anti-rabbit IgG-HRP; Abcam), diluted in blotto (1:10,000 – 20,000), for 1 hour on a shaker at room temperature. The membrane was again rinsed with TBS-Tween and washed three times with gentle shaking at room temperature, for 5 minutes each, again with TBS-Tween. Chemiluminescence was detected using an EZ-Chemiluminescence

Detection Kit for HRP (Geneflow) and recorded using an LAS-4000 imager (GE Healthcare). Band intensities were quantified using Aida Image Analyzer software (Raytest).

2.10 Protein import into isolated chloroplasts

2.10.1 Preparation of radiolabelled preproteins

The substrate used for the *in vitro* chloroplast protein import assays was the precursor of the small subunit of Rubisco (pSSU). *E. coli* DH5α cells carrying a cDNA plasmid clone encoding pSSU (*Arabidopsis RBCS1A*; At1g67090) were grown from a glycerol stock, and the plasmid was purified by Isolate II Plasmid Mini Kit (Bioline). The sequence encoding pSSU was amplified by PCR using universal M13 forward (5'-TGTAAAACGACGGCCAGT-3') and reverse (5'CAGGAAACAGCTATGACC-3') primers, and the amplicon was purified by GenElute Gel Extraction Kit (Sigma). Radiolabelled protein was prepared by using a rabbit reticulocyte *in vitro* translation system (TNT T7 Quick for PCR DNA (Promega)) in conjunction with ^{35}S-methionine (Perkin-Elmer), as described previously (Aronsson & Jarvis, 2011). The mixture was left to incubate for 90 minutes in a 30°C water bath for synthesis.

2.10.2 Chloroplast isolation and protein import

Chloroplasts were isolated from 2-week-old *in vitro* grown plants as described previously (Flores-Pérez & Jarvis, 2017). All steps were performed in 4°C including centrifugation. The machine used for centrifugation in this isolation step was a Beckman Coulter Avanti JXN-26 centrifuge.

The seedlings were homogenized in ice-cold 1X chloroplast isolation buffer (CIB) prepared by using sterile deionized water to diluting a 2X CIB stock (0.6 M sorbitol, 10 mM $MgCl_2$, 10 mM ethylene glycol tetraacetic acid (EGTA), 10 mM ethylenediaminetetraacetic acid (EDTA), 20 mM sodium hydrogen carbonate ($NaHCO_3$), 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES); adjust pH 8.0 with KOH). The homogenate was filtered through a double layer of Miracloth into a 250 mL centrifugation tube. Centrifugation was performed at 1000 x *g* for 5 minutes with the brake on using JLA-16.250 rotor. Intact chloroplasts were separated using a pre-prepared Percoll gradient (13 mL Percoll medium (GE Healthcare), 13 mL 2X CIB and 5 mg glutathione; all components were mixed and centrifuged in a JA-25.50 fixed angle rotor at 43,000 x *g* for 30 minutes with the brake off). The intact chloroplast was then collected and rinsed in 20-25 mL of HEPES-$MgSO_4$-sorbitol (HMS) buffer (50 mM HEPES-NaOH, pH 8.0, 3 mM magnesium sulphate ($MgSO_4$), 0.3 M sorbitol) to wash off the Percoll before centrifugation in a JS-13.1 swinging bucket rotor at 1000 x *g* for 5 minutes with the brake on. The integrity and yield of isolated chloroplasts was assessed under a phase-contrast microscope using a counting chamber.

Protein import was performed according to a published protocol (Aronsson & Jarvis, 2011) with minor modifications. The total volume of each import reaction depended on the number of required time-points: 150 μL of import reaction was assembled per time point (e.g., for three time-points, 450 μL was required) in a clear 2 mL centrifuge tube. For each time-point, the working reactions were made up of 1X HMS buffer, 20 mM potassium gluconate (K-gluc), 10 mM sodium bicarbonate ($NaHCO_3$), 0.2% (v/v) BSA, 5 mM MgATP, 10 mM methionine, 2.6 mM L-cysteine, 5 μL radiolabeled pSSU, and chloroplast suspension containing 10^7 chloroplasts. For

my experiments, three time-point (3, 6 and 9 minutes) import reactions were conducted for each genotype. Reactions were performed in white light (100 μmol m^{-2} s^{-1}) in a circulating water bath set at 25°C. At each time-point, the reaction was stopped by taking out a 140 μL aliquot (after gently resuspending the chloroplasts) and placing it in an equal volume of ice-cold stop buffer (50 mM EDTA dissolved in 1X HMS buffer).

Samples were centrifuged at 10,000 rpm for 1 minute at room temperature in a microcentrifuge to recover the chloroplasts. After discarding the supernatant, the chloroplasts were then resuspended in 20 μL 2X protein loading buffer. A sample of the radiolabeled pSSU equivalent to 10% of the volume used in each time-point was also prepared in 5.5 μL 2X protein loading buffer. All samples were boiled at 100°C for 5 minutes, and then 6 μL of each sample was resolved by 15% SDS-PAGE (Section 2.9.2). The gel was stained as in Section 2.9.3 and dried on a piece of 3M Whatman filter paper using a Bio-Rad 583 gel dryer at 80°C for 2 hours. The amount of radioactivity associated with the precursor and mature forms of the imported substrate was detected by phosphorimaging using a Storm 860 Molecular Imager (Amersham Biosciences). Band intensities were quantified using Aida Image Analyzer software (Raytest).

2.11 Protoplast transient assays and immunoprecipitation

All centrifugation steps were performed using 6441 swinging buckets in a Sorvall Legend RT at 4°C unless stated.

2.11.1 Protoplast preparation from *Arabidopsis thaliana*

Protoplast isolation was carried out using the 'Tape-*Arabidopsis* Sandwich' method (Wu *et al.*, 2009) using the rosette leaves of 4-week-old plants. Cut leaves were fixed on their adaxial side to a strip of Time tape (3M), while the abaxial side was fixed to a strip of Magic tape (Scotch). The lower epidermal cell layer was then peeled off carefully by pulling away the Magic tape. While still adhered to the Time tape, the peeled leaves (30 leaves per experiment) were transferred to a Petri dish containing 10 mL of enzyme solution (1% (w/v) cellulase Onozuka R-10 (Duchefa), 0.25% (w/v) macerozyme R-10 (Duchefa), 0.4 M mannitol, 10 mM calcium chloride ($CaCl_2$), 20 mM potassium chloride (KCl), 0.1% (w/v) BSA, 20 mM MES; pH adjusted to 5.7 with KOH) and left to digest on a rotating shaker at 50 rpm for 2 hours at room temperature.

Following digestion, the enzyme solution containing released protoplasts was transferred into a 50 mL Falcon tube and subjected to centrifugation at 60 x *g* for 5 minutes. The protoplasts were washed twice in pre-chilled 25 mL W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 5 mM glucose, 2 mM MES; pH adjusted to 5.7 with KOH) and left to incubate on ice for 30 minutes. After incubation, the protoplasts were harvested by centrifugation at 40 x *g* for 4 minutes and resuspended in a 1.5 mL of room temperature MMg solution (0.4 M mannitol, 15 mM $MgCl_2$, 5 mM MES; pH adjusted to 5.7 with KOH).

2.11.2 Protoplast transfection

All protoplast transfection procedures in this section were carried out at room temperature. In a 50 mL Falcon tube, 600 μL of protoplasts in MMg solution was mixed by pipetting in gently 30 μg of plasmid DNA and 660 μL of PEG4000 solution

(59% (w/v) PEG4000, 0.59 M mannitol, 0.15 M calcium nitrate ($Ca(NO_3)_2$)). The mixture was left to incubate for 5 minutes.

After incubation, the following series of W5 solution volumes was added consecutively to the mixture with gentle mixing, every 5 minutes: 1.2 mL, 2.4 mL and 4.8 mL. The transformed protoplasts were harvested by centrifugation at 40 x *g* for 3 minutes at room temperature, and then washed twice in 9 mL of W5 solution, with re-centrifugation between each wash, before finally resuspending them in 3.6 mL of fresh W5 solution. The protoplasts were transferred onto a 6 wells Nunclon round cell culture plate (Fisher Scientific) and left to incubate overnight in the dark under standard plant growth conditions.

2.11.3 Protein pulldown by immunoprecipitation

On the next day, 5 μM of the proteasome inhibitor bortezomib was added to the protoplast suspension which was then left to incubate gently on a shaker for 4 hours at room temperature. The protoplast suspension was then transferred into a 50 mL Falcon tube and subjected to 2 minutes centrifugation at 100 x *g* to harvest the cells. The pelleted protoplasts were lysed with 500 μL Lysis buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 1 mM EDTA, 1% (v/v) Triton-X100, 5 μM bortezomib, 0.5% (v/v) PIC) and left to incubate gently on a rotating mixer at 4°C for 30 minutes. After lysis, the lysate was cleared by centrifuging at maximum speed in a microcentrifuge for 10 minutes at 4°C, and the supernatant was retained. A 30 μL aliquot of the supernatant was taken out as a Total Lysate (TL) sample.

The GFP-Trap MA magnetic beads (Chromotek) to be used for immunoprecipitations were equilibrated with ice-cold IP buffer (Lysis buffer without bortezomib) before use, as follows. For each sample (complete lysate from one

transfection experiment), 25 μL of beads were equilibrated with 500 μL IP buffer twice in a 1.5 mL centrifuge tube; on adding the buffer, the mixture was vortexed briefly and then the supernatant was discarded each time after retaining the beads on a MagJET Separation Rack (Thermo Fisher). The cleared lysate was then added to the equilibrated beads and left to bind for 1 hour on a rotating mixer at 4°C. The beads were then magnetically pelleted and washed four times; each wash step with 500 μL Wash buffer (Lysis buffer without bortezomib and PIC) for 5 minutes. Bound proteins were eluted from the magnetic beads by adding 60 μL of 2X protein loading buffer and boiling at 95°C for 10 minutes. The beads were separated from the eluted protein by using the MagJET Separation Rack (Thermo Fisher), and the Eluted Protein samples were then analysed by immunoblotting as described in Section 2.9.4.

Enrichment of the YFP fusion proteins was detected using Anti-GFP antibody (Sigma) by comparing the band intensities from the Total Lysate and Eluted Protein samples.

2.12 *In vitro* autoubiquitination assays

2.12.1 Protein expression in bacteria

The plasmids constructed and isolated in Section 2.8.2 were transformed into *Escherichia coli* Rosetta (DE3) cells (Novagen). Positive transformants were selected by growing on ampicillin antibiotic-containing LB agar selection plates. Bacterial strains for the expression of Free GST, GST-SP1cyt and GST-SP1cyt-C330A were grown from a glycerol stock and have been described before (Ling *et al.*, 2012). In each case, a single colony was inoculated into 300 mL of liquid LB medium supplemented with ampicillin and left to grow at 37°C with shaking at 240

rpm until the OD_{600} reached 0.6-0.8. Expression of fusion proteins was induced by adding 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) before further shaking at 21°C for 16 hours. The cells were harvested by centrifugation at 4000 rpm for 20 minutes using 6441 swinging-bucket in a Sorvall Legend RT centrifuge, and then resuspended in 20 mL ice-cold Lysis buffer (25 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.1% (v/v) Triton X-100, EDTA-free protease inhibitor cocktail tablet (Roche) according to the manufacturer's guideline). Cells were lysed by adding 20 mg of lysozyme (Sigma) and incubating on ice for 20 minutes, before subjecting them to 15 rounds of sonication using a 705 Sonic Dismembrator (Fisher Scientific), each round comprising a 10 second burst of sonication at setting 3 with a 20 second interval. Following sonication, 1 mL of 20% Triton X-100 was added to each tube and left to rotate at 4°C for 1 hour. The lysate was centrifuged at 12,000 x *g* for 30 minutes at 4°C using a JA-25.50 rotor in a Beckman Coulter Avanti JXN-26 centrifuge, and then the supernatant was collected.

2.12.2 Purification of recombinant proteins

Prior to the purification of the GST fusion proteins, 500 μL of glutathione-agarose (Sigma) was equilibrated with at least 6-bed volumes of Lysis buffer (Section 2.12.1) for 10 minutes at 4°C. Protein purifications were carried out in Econo-Pac Chromatography Columns (Bio-Rad). The cell lysate supernatant prepared in Section 2.12.1 was incubated with the equilibrated glutathione-agarose beads on a rotating mixer overnight at 4°C. The supernatant was then discarded by allowing it to flow through the column under gravity, and the retained beads were washed with ice-cold wash buffer (25 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.1% (v/v)

Triton X-100) a total of four times, with each washing step involving ~1 minute on the rotating mixer followed by buffer removal under gravity.

The GST proteins were eluted from the beads using 50 mM reduced glutathione (Duchefa Biochemie) dissoved in elution buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% (v/v) Triton X-100). The eluted GST proteins were recovered by allowing it to flow through the column under gravity. The reduced glutathione in the eluted sample was then diluted by buffer exchange; the sample was first concentrated by centrifugation at 4°C in a microcentrifuge through a Vivaspin 500 ultrafiltration spin column (Sartorius) according to the manufacturer's instruction, and then with the same volume of Elution buffer was added. The concentration of the purified proteins were estimated by running on SDS-PAGE gel (Section 2.9.2) along with known BSA standards (0.5 and 1 μg) before freezing in 20% glycerol at -20°C.

2.12.3 Ubiquitination assay

A 1 μg sample of each purified GST and GST-tagged E3 ligase was incubated with 100 ng of human E1 (Merck), 200 ng of 6xHis-tagged AtUBC8 E2 (Ling *et al.*, 2012), and 6 of μg HA-ubiquitin (Boston Biochemical) in a 30 μL reaction containing 50 mM Tris-HCl, pH 7.5, 60 mM NaCl, 1 mM DTT, 0.05 mM zinc chloride ($ZnCl_2$) and 4 mM MgATP; the reaction was conducted at 30°C for 2 hours. After incubation, the reaction was stopped by adding an equal volume of 2X protein loading buffer. The samples were analyzed by immunoblotting using anti-HA antibody (Sigma) to detect ubiquitinated species.

To conduct ubiquitination experiments containing both GST-SP1cyt and GST-SPL1cyt, the two E3 ligases were mixed in different ratios (1:2, 1:4 and 1:8) in the same 30 μL reaction.

2.13 Forward genetic screen

Mutagenesis was carried out using the chemical mutagen ethyl methanesulphonate (EMS) inside an Atmos bag (Sigma) within a fume hood, following a previously described protocol (Leyser *et al.*, 1992). Approximately 20,000 *sp1-1 ppi1-1* (*sp243 BC3 5*) seeds were subjected to mutagenesis. The seeds were pre-imbibed in a 0.1% (w/v) potassium chloride solution overnight at room temperature, and then soaked for 3 hours at room temperature with occasional mixing in a solution comprising: 0.1 M sodium phosphate, pH 5, 5% (v/v) dimethyl sulphoxide, and 50 mM EMS. The EMS was deactivated by washing the seeds twice in 50 mL of 100 mM sodium thiosulphate, and then once in 50 mL of distilled water, with each wash lasting 15 minutes. The mutagenized M_1 seeds were sown on 100 half trays (15.5 x 21 x 5.5 cm) of soil, with the plants on each half of the same tray designated A and B; thus, a total of 100 pools of M_2 seeds were generated, termed 1A-50A and 1B-50B. Due to time constraints and the availability of space to grow the plants, only the A pools were screened. The M_2 screening was for plants which had a greener visible phenotype than the *sp1 ppi1* progenitor. The M_2 seeds were germinated on MS medium in large (120 x 120 mm) Petri dishes, and the plants were grown under standard growth conditions for 10-14 days prior to screening. Seedlings with the desired phenotype were transferred to soil and allowed to mature to produce the M_3 seeds. The phenotype in the M_3 was confirmed by crossing to *sp1 ppi1* to determine the inheritance (dominant or recessive) and the segregation ratio in the F_2. Plants which appeared greener but which also showed developmental defects such as dwarfism or sterility were not carried forward for analysis.

2.14 Mapping by sequencing

2.14.1 DNA preparation for Next Generation Sequencing (NGS)

The mutant lines of interest were backcrossed to the non-mutagenized progenitor, *sp1 ppi1*. For each line, around 50-100 plants in the segregating F_2 generation which clearly displayed the relevant mutant phenotype were selected and pooled for DNA extraction using a DNeasy Plant Mini Kit (Qiagen). The DNA from a population of a non-mutagenized parent plants was also extracted using the same procedure. The extracted DNA was concentrated by ethanol precipitation: 0.1 volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of ice-cold absolute ethanol were added to each DNA sample, and they were then left at -20°C for 1 hour before centrifugation at 14,000 rpm and 4°C for 15 minutes in a microcentrifuge. The pelleted DNA was dissolved in a suitable volume of the Qiagen DNeasy Plant Mini Kit elution buffer and the quality of the DNA was confirmed by running it on a 1% agarose gel. A Qubit 2.0 Fluorometer (Invitrogen) was used to measure the DNA concentration according to the manufacturer's instructions.

Library preparation and sequencing were conducted at the Oxford Genomics Centre at the Wellcome Centre for Human Genetics. Approximately 1-5 μg genomic DNA per sample at a minimum concentration of 29.4 ng/μl was used in sequencing library preparation. Individual barcoded Illumina TruSeq DNA libraries were generated for each genotype. The four samples were then sequenced on one lane of Illumina HiSeq 4000 with 150 bases paired-end reads.

2.14.2 Mapping

The sequenced genomes from the pooled backcrossed F_2 mutants, and from the non-mutagenized parents, were analysed using SHOREmap (Schneeberger *et*

al., 2009). The graph output from the analysis was recreated and annotated in R (The R Foundation).

2.15 Tables

Mutant allele	Nature of mutation	Notes for genotyping
spl1-1	T-DNA insertion	Table 2.2
spl1-2 (*dal2-1*)	T-DNA insertion	(Basnayake *et al.*, 2011)
spl1-3 (*dal2-2*)	T-DNA insertion	(Basnayake *et al.*, 2011)
sp1-1	Point mutation	(Ling *et al.*, 2012)
sp1-3	T-DNA insertion	(Ling *et al.*, 2012)
ppi1-1	T-DNA insertion	(Jarvis *et al.*, 1998)
ppi2-3 (*fts1*)	Point mutation	(Woodson *et al.*, 2015)
toc75-III-3	Point mutation	(Stanga *et al.*, 2009)
tic40-4	T-DNA insertion	(Kovacheva *et al.*, 2004)
hsp93-V-1	T-DNA insertion	(Kovacheva *et al.*, 2004)

Table 2.1 *Arabidopsis* mutant alleles used in this study

Mutant allele	Primer name	Sequence (5' to 3')*
spl1-1	SPL1-ORF-F	AAAAAGCAGGCTCCATGATACATTTGGCTGGATT
	SPL1-ORF-nonstop-R	AGAAAGCTGGGTTATGGCGGTAAATTTTCAAAAC
	Lba1	TGGTTCACGTAGTGGGCCATCG

Table 2.2 Primers used for genotyping the *spl1-1* mutant

Primer name	Sequence (5' to 3')*
SPL1-F	AAAAAGCAGGCTCCATGATACATTTGGCTGGATT
SPL1-R	AGAAAGCTGGGTTATGGCGGTAAATTTTCAAAAC
SPL2-F	AAAAAGCAGGCTCCATGTCCTCGCCGGAGCGTG
SPL2-R	AGAAAGCTGGGTTCTAAGAGTAATATACACGCATAGAT
eIF4e1-F	AAACAATGGCGGTAGAAGACACTC
eIF4e1-R	AAGATTTGAGAGGTTTCAAGCGGTGTAAG

Table 2.3 Primers used in RT-PCR experiments

Primer name	Sequence (5' to 3')
Q-SPL2-F	ATGACCAAGGACAAGATGA
Q-SPL2-R	ATGCCAACAGACACAATG
Q-Actin2-F	TCAGATGCCCAGAAGTCTTGTTCC
Q-Actin2-R	AAGAGCGGAAGAAGATGA

Table 2.4 Primers used in qRT-PCR experiments

Primer name	Sequence (5' to 3')*
SPL1-ORF-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGATACATTTGGCTG GATT
SPL1-stop-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAATGGCGGTAAATTTT C
SPL1-cyt-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGTGATTGAATATATTCT A
SPL1-H308Y-F	GAGTGTGGTTATATGTGCTG
SPL1-H308Y-R	CAGCACATATAACCACACTC
SPL1-E305P-F	TTGTTCCGTGTGGTCATA
SPL1-E305P-R	TATGACCACACGGAACAA
SPL1-E327-R	TTGTTGTATTCGTCGCCG
SPL1-E327stop-R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAATGGCGGTAAATTTT CAAAACTTGTTGTATTCGTCGCCG
SPL2-CDNA-F	AAAAAGCAGGCTCCACTTGTCCGTGTGACCG
SPL2-CDNA-R	AGAAAGCTGGGTTATCGGTTACAAAATTTCTTCC
SPL2-ORF-F	AAAAAGCAGGCTCCATGTCCTCGCCGGAGCGTG
SPL2-CDS-R	AGAAAGCTGGGTTCTAAGAGTAATATACACGCATAGAT
SPL2-H348Y-F	CCTGCGTTTATTCCCTGTGGATATGTAGTATG
SPL2-H348Y-R	CACATCGCCTGCAACATACTACATATCCACAG
SPL2-ORC-R	AGAAAGCTGGGTTTTGAGGCACATGAGAAGACC
SPL2-cyt-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGCTGCTGTCAGGACCT GGAA
SPL2amR1-I miR-s	GATACTAAGAGTAATATACGCGCTCTCTCTTTTGTATTCC
SPL2amR1-II miR-a	GAGCGCGTATATTACTCTTAGTATCAAAGAGAATCAATGA
SPL2amR1-III miR*s	GAGCACGTATATTACACTTAGTTTCACAGGTCGTGATATG
SPL2amR1-IV miR*a	GAAACTAAGTGTAATATACGTGCTCTACATATATATTCCT

Table 2.5 Primers used to generate various constructs and point mutations

*Nucleotides shown in red corresponds to linker sequences or mutation sites.

Chapter 3

Results I

CHARACTERIZATION OF THE *SP1-LIKE* GENES

Abstract

A forward genetic experiment in *Arabidopsis thaliana* identified a novel gene *SP1* that encodes for a RING E3 ligase, a component of the ubiquitin-proteasome system (UPS). This discovery demonstrated the involvement of the UPS in determining the fate of the plastid via turnover of the TOC protein. Upon further investigation, two homologues of *SP1* were identified in the same plant species. These homologues (SPL1 and SPL2) share a similar topology and sequence identity with SP1, and have been shown to be similarly localized in the chloroplast envelope. The published genetic evidence did not support functional redundancy between *SPL1* and *SP1*, and no data were available for *SPL2*. In this chapter, I describe the characterization of both of the homologues in detail. I began by employing reverse genetics to obtain *spl* mutants, which were then analysed for their effects on the chlorotic *Arabidopsis toc* mutant *ppi1*. Mutants of *spl1* did not show any obvious effects on the *ppi1* phenotype when grown under standard growth conditions. However, knockdown (KD) of *SPL2* gene expression did cause weak suppression of the chlorotic *ppi1* phenotype. The *SPL2-KD ppi1* chloroplasts were more developed than those in *ppi1*, and this recovery was linked to an improvement in their ability to carry out the process of protein import. Further genetic analysis with other *toc* mutants revealed that *SPL2* may be promiscuous in function and could have additional substrates other than the TOC receptors. Nonetheless, a strong association was observed between SPL2 and the TOC proteins by immunoprecipitation. The results presented here provided clues concerning the roles of the *SPLs* in chloroplast biogenesis which aided the further investigations described in the following results chapter.

CHAPTER 3:

CHARACTERIZATION OF THE *SP1-LIKE* GENES

3.1 Introduction

It is well known that chloroplasts have evolved an intricate mechanism to import almost all (>90%) of the proteins required for their development, functions and homeostasis, which involves in two protein complexes known as the translocons at the outer/inner envelope membranes of chloroplasts (TOC/TIC) (Jarvis & López-Juez, 2013; Paila *et al.*, 2015). As their names suggest, these complexes are located in the double envelope membrane of the chloroplast. Although the exact molecular mechanism of import is unclear, new evidence suggests that the two complexes exist as a stable supercomplex (Chen & Li, 2017) or at least work in close cooperativity (Richardson *et al.*, 2018) during transit peptide recognition and translocation. The TOC and TIC complexes work in unison, and together form what is known as the chloroplast protein import machinery (Strittmatter *et al.*, 2010; Paila *et al.*, 2015).

The discovery and identification of the *SP1* gene opened up a new exciting area of plastid biology, in particular concerning how the protein import machinery is regulated post-translationally (Ling *et al.*, 2012). The *SP1* gene encodes a chloroplast RING-type ubiquitin E3 ligase, which is a component of the UPS protein degradation pathway. Genetic and biochemical evidence suggest that SP1, together with its cognate E2 conjugase, primarily act on the TOC protein complex by tagging them with ubiquitin moieties for degradation by the cytosolic 26S proteasome (Ling *et al.*, 2012). Interestingly, two homologues of SP1 were identified and reported in

the study. These are SP1-Like1 (SPL1) and SP1-Like2 (SPL2). Both homologues showed amino acid sequence homology to SP1, with SPL1 being the most closely related. The SPL1 and SPL2 proteins were predicted to share a similar topological arrangement with SP1, and were demonstrated to have chloroplast envelope localization, like SP1, via YFP fusion-protein analysis using confocal microscopy in protoplasts (Ling *et al.*, 2012).

A preliminary investigation of the function of the SPL1 E3 ligase was conducted by overexpressing the *SPL1* gene in the *sp1 ppi1* double mutant, to assess for complementation of *sp1* (Ling *et al.*, 2012). Overexpression of *SPL1* was not able to complement the *sp1* mutation in the *ppi1* background as confirmed by measuring the chlorophyll content. This suggested that *SPL1* has a distinctive role, possibly unrelated to the regulation of the chloroplast protein import machinery. No data were presented on the possible involvement of *SPL2* in this regulatory pathway.

Therefore the aim of the work described in this chapter was to characterize the two homologues in greater detail, by employing reverse genetics and other methods, in order to shed a light on their roles in the context of chloroplast biogenesis.

3.2 Results

3.2.1 Overexpression of *SPL2* partially complements the *sp1* mutation

In order to gain an insight into the potential role of *SPL2* in regulating the chloroplast protein import machinery, the *SPL2* cDNA was cloned downstream of the constitutive cauliflower mosaic virus (CaMV) *35S* promoter in the binary vector pB2GW7. The construct was transformed into the *sp1 ppi1* mutant (and henceforth known as *SPL2-OX sp1-3 ppi1*) to test for phenotypic complementation. These steps were performed by Dr Qihua Ling. The resultant T_1 seeds were screened for positive transformants by germinating on phosphinotricin-containing MS media. At least 10 T_1 lines were analyzed followed by selection for 3:1 ratio lines in the T_2 generation, indicating single T-DNA insertion. Lastly, homozygous lines displaying 100% resistance on phosphinotricin-containing MS media in the T_3 were selected for further analysis.

The transgenic plants appeared slightly paler than *sp1-3 ppi1* but were still greener than *ppi1* (Figure 3.1 A). This observation was confirmed by measuring the chlorophyll contents of the plants, which placed the *SPL2-OX sp1-3 ppi1* in between *sp1-3 ppi1* and *ppi1* (Figure 3.1 B). Semi-quantitative RT-PCR analysis using primers that amplify the full-length *SPL2* cDNA confirmed a ~14-fold increase in the level of *SPL2* transcript in the *SPL2-OX sp1-3 ppi1* plants (Figure 3.1 C and D).

This initial finding suggested that *SPL2* may be partially redundant with *sp1* in function, owing to the observed incomplete complementation.

3.2.2 Identification of *spl1* mutant lines

To further understand the function of *SPL1* in *Arabidopsis*, several T-DNA mutants were identified and characterized. Two of the mutants were of the Salk

population (salk_064720 and salk_024744) and one was from the Wisconsin population (WiscDsLox383C6). The salk_064720 was the first mutant to be identified and therefore will be known as *spl1-1*. Both salk_024744 (*dal2-1*) and WiscDsLox383C6 (*dal2-2*) have been characterized before in a different study (Basnayake *et al.*, 2011), and so will be renamed here as *spl1-2* and *spl1-3*, respectively (Figure 3.2 A).

Semi-quantitative RT-PCR analysis using primers that amplify the full-length cDNA of the *SPL1* gene showed no amplification in the *spl1-1* mutant (Figure 3.2 B). This line is, therefore, a new null mutant of *SPL1*, similar to *spl1-2* and *spl1-3* as reported previously (Basnayake *et al.*, 2011). These steps were performed by Dr Weihua Huang. The *spl1* knockout mutant plants showed no observable change in their visible phenotype under standard growth conditions (Figure 3.2 D); the plants appeared exactly like the Col-0 wild type. The three *spl1* alleles were phenotypically identical, and so data are presented only for *spl1-1* as a representative allele.

3.2.3 Generation of *SPL2* knockdown line by artificial microRNA

There were no suitable T-DNA insertion lines available for *SPL2* and so knockdown by artificial microRNA (amiRNA) was adopted (Schwab *et al.*, 2006; Ossowski *et al.*, 2008) to alter the native *SPL2* gene expression. The amiRNA was introduced into wild-type *Arabidopsis* by floral dipping. The positive transformants were screened by germinating the seeds on MS media supplemented with phosphinotricin. Homozygous plant with single T-DNA insertion were chosen for further analysis. This transgenic plant line will be identified as *SPL2-KD*. Gene expression analysis by qRT-PCR showed the native *SPL2* transcript level was significantly reduced, to ~20% relative to the wild type, in the *SPL2-KD* line (Figure

3.2 C). As was found with the *spl1* mutants (and with *sp1* previously (Ling *et al.*, 2012)), depletion of the *SPL2* transcript did not cause any observable changes in the visible phenotype of the plants under standard growth conditions (Figure 3.2 D).

It was noted that only one good *SPL2-KD* line was obtained in this part of the study. This may have limitations for a robust physiological analysis such as non-specific effect to other developmentally related genes.

3.2.4 Genetic interaction analysis with *ppi1*

The suppression mediated by the *sp1* mutation was identified and characterized extensively in the *ppi1* mutant (Ling *et al.*, 2012). For this reason, and given the homologous relationship of the SPLs with SP1, it made sense to start by looking at the genetic interaction of the mutants of the homologues with the same plastid protein import mutant (*ppi1*), to seek clues concerning the roles of the homologues in chloroplast biogenesis.

The *spl1-1* mutation was introduced by cross-pollination into the *ppi1* background. Successful pollination in the F_1 was confirmed by PCR based test using gene specific primers and a primer specific for the left-border of the T-DNA (Table 2.2, (Jarvis *et al.*, 1998). The identified double *spl1-1 ppi1* heterozygote mutants were further propagated until double-homozygous for the mutant alleles were reached, and their genotypes again confirmed by PCR-based tests. A similar approach was applied with the *SPL2-KD* line, by crossing the transgene into the *ppi1* background; the resulting double-homozygous line will be identified here as *SPL2-KD ppi1 #1*. In addition, another line (termed *SPL2-KD ppi1 #2*) was generated by direct transformation of the *SPL2-KD* construct into *ppi1* plants. For both of these lines, the extent of *SPL2* transcript knockdown in the *ppi1* mutant background was

assessed by qRT-PCR. In both cases, the expression of the native *SPL2* transcript was successfully reduced to ~25% relative to the *ppi1* mutant (Figure 3.3 A).

The *spl1-1 ppi1* double mutant showed no obvious differences in plant morphology compared to *ppi1*, whereas both of the *SPL2-KD ppi1* lines exhibited mild suppression of the chlorotic phenotype associated with *ppi1* (Figure 3.3 B and C). The phenotype of the *spl1-1 ppi1* was further confirmed with a different knockout allele of *spl1* (Figure 3.4 A). Just like *spl1-1 ppi1*, the *spl1-2 ppi1* double mutants showed no significant difference in their visible phenotype or their chlorophyll content relative to *ppi1* (Figure 3.4 B).

Both *SPL2-KD ppi1* lines appeared slightly greener than *ppi1*, but less green than *sp1 ppi1*, and this was reflected in the chlorophyll measurements (Figure 3.3 D). This observation was consistently made with plants at different developmental stages, whether grown *in vitro* (Figure 3.3 B) or grown on soil (Figure 3.3 C).

In the report by Ling *et al.*, (2012), a TEM analysis revealed the suppression mediated by the *sp1* mutation in the *ppi1* background resulted in an improvement in chloroplast size and ultrastructural organization by TEM. A similar approach was conducted to study the effect of *SPL2-KD ppi1* on the chloroplast development. Rosette leaves of 35-day old plants were sent to Dr Errin Johnson at the Sir William Dunn School of Pathology, University of Oxford for processing, and imaging was performed on Technai 12 TEM. The analysis on the double-mutant *SPL2-KD ppi1* plants exhibited an overall larger chloroplast cross-sectional area than *ppi1*, comparable to that of *sp1 ppi1* (Figure 3.5 A and B). The is no effect on chloroplast ultrastructure in the *spl1-1 ppi1* double mutant (Figure 3.5 A and B), in line with the visible phenotype and chlorophyll data (Figure 3.3 B, C and D).

These findings suggest that *SPL2* has a similar function to *SP1* in regulating the process of chloroplast protein import and chloroplast biogenesis. They further suggest that *SPL1*, however, may not share the same role as *SP1*, and so its function remains unclear.

3.2.5 *SPL2* regulates the chloroplast protein import machinery

Ling *et al.* (2012) showed that mutation of *sp1* in the *ppi1* background stabilizes the TOC complex of the protein import machinery, due to a reduction in protein turnover by the UPS. As a consequence, the levels of the TOC proteins were elevated in the *sp1* mutant; this was especially true for Toc75. These TOC protein abundance changes were linked to an improvement in the protein import capacity of the double-mutant chloroplasts relative to those of the pale *ppi1* mutant, resulting in the phenotypic recovery of the plants.

The observations made as a result of the genetic interaction studies in Section 3.2.4 suggest that the knockdown of *SPL2* may affect the levels of TOC proteins, just like in *sp1 ppi1*. Therefore, an immunoblot analysis of the protein import apparatus components was conducted. The results revealed significant changes in the level of Toc75 in the *SPL2-KD ppi1*, relative to *ppi1*, although the effect was not as large as in *sp1 ppi1* (Figure 3.6 A and B). For *spl1 ppi1* double mutant, however, only the level of Toc159 was slightly increased to the same level as *sp1 ppi1 and SPL2-KD ppi1* (Figure 3.6 B) but was insufficient to cause any phenotypic changes (Figure 3.3 B and Figure 3.4 A). The level of Toc75 remained unchanged in the *spl1 ppi1* double mutant compared to the *ppi1* single mutant. As observed previously for the *sp1 ppi1* (Ling *et al.*, 2012), there were no significant changes to the levels of TIC components of the chloroplast protein import machinery

in either of the *spl* double mutants, suggesting that the effects of the homologues on the TOC components are specific (Figure 3.6 B).

To determine if the elevated levels of TOC proteins in the *SPL2-KD ppi1* transgenic line have an effect on protein import into chloroplasts, *in vitro* import assays were conducted. The chloroplasts of 2-week-old wild-type, *ppi1* and *SPL2-KD ppi1* plants were isolated using a Percoll gradient, and then incubated in import assay buffer with radiolabelled Rubisco small subunit precursor protein over a time-course of 9 minutes. As expected, protein import into *ppi1* chloroplasts was substantially retarded relative to that into wild-type chloroplasts. Interestingly, the rate of import into *SPL2-KD ppi1* chloroplasts was partially recovered, relative to the *ppi1* single mutant (Figure 3.7 A and B). This finding demonstrated an improvement in chloroplast protein import capacity in *SPL2-KD ppi1*, relative to *ppi1*, which we may attribute to the stabilization of the TOC complexes (Figure 3.6); as a result, larger sized chloroplasts develop (Figure 3.5) leading to a greener plant phenotype (Figure 3.3).

3.2.6 Genetic analysis of *SPL2-KD* with other chloroplast protein import mutants

The immunoblot analysis of *SPL2-KD ppi1* in Figure 3.6 showed a slight increase in the levels of Toc159 and Toc75, relative to *ppi1*,with no changes to the TIC components. This suggested a specific functional link between SPL2 and the TOC components. To further test this hypothesis, the *SPL2 amiRNA* transgene was crossed into various chloroplast protein import mutant backgrounds (affecting both TOC and TIC), namely *ppi2-3* (a hypomorphic allele of *toc159* originally reported as *fts1* (Woodson *et al.*, 2015), but will be known as *ppi2-3* here following an

established nomenclature (Bauer *et al.*, 2000)), *toc75-III-3* (Stanga *et al.*, 2009; Huang *et al.*, 2011), *tic40-4* and *hsp93-V-1* (Kovacheva *et al.*, 2004).

In contrast with what was observed with *ppi1*, where suppression was detected, the *SPL2-KD ppi2-3* double mutant was slightly smaller in size than the *ppi2-3* single mutant, and appeared slightly more chlorotic (Figure 3.8 A and B). In spite of this, immunoblot analysis revealed that the level of Toc75 protein was elevated in the double mutant (Figure 3.8 C and D); while the level of Toc159 remained unaffected in the double mutant (Figure C and D), this was possibly due to the hypomorphic nature of the *ppi2-3* mutation (a Q1472stop nonsense mutation in the M domain, (Woodson *et al.*, 2015)). The other newly-identified *toc* double mutant, *SPL2-KD toc75-III-3*, also showed no suppression of the *toc* mutant visible phenotype, and this was confirmed by making chlorophyll measurements (Figure 3.9 A and B). The reason for the failure of *SPL2-KD* to suppress these two additional *toc* mutations is unclear, but it may indicate an unanticipated degree of specificity of the SPL2 component, or reflect technical details such as the weakness or hypomorphic nature of the *toc* mutations.

As expected, the phenotypes of the TIC-associated mutants (*tic40-4* and *hsp93-V-1*) remained unchanged when the *SPL2 amiRNA* transgene was introduced (Figure 3.9 C and E), and both observations were confirmed by measuring the chlorophyll contents of the double mutants (Figure 3.9 D and F). These data support the notion that the suppression effect of *SPL2-KD* on *ppi1* is specific.

3.2.7 SPL2 E3 ligase interacts directly with the TOC proteins

The data presented so far have suggested a close functional link between SPL2 and TOC proteins, analogous to the previously-reported link between SP1 and

TOC proteins (Ling *et al.*, 2012). To investigate the possibility that TOC proteins may be substrates of the SPL2 E3 ligase, immunoprecipitation (IP) experiments were performed using YFP-tagged SPL2 transiently expressed in *Arabidopsis* protoplasts. To this end, the *SPL2* coding region was cloned upstream of sequence encoding the YFP tag in the binary vector pK7YWG2, and the resulting vector was used to transfect protoplasts. Two control transfection experiments were set up in parallel, using either water (as a negative control) or plasmid encoding SP1-YFP based on the vector p2GWY7 (Karimi *et al.*, 2007; Ling *et al.*, 2012). Following lysis, solubilization and anti-YFP IP, bound protein complexes were eluted and denatured in 2X protein loading buffer prior to analysis by SDS-PAGE and immunoblotting. Both SP1-YFP and SPL2-YFP were enriched in the eluted fraction compared to the total lysate, indicating successful IP. The results obtained for the SP1-YFP control were as expected (Ling *et al.*, 2012), and those for SPL2-YFP were very similar, demonstrating a direct interaction of SPL2 with Toc159, Toc75 and Toc33 (Figure 3.10). No interactions were observed with non-TOC components of the chloroplast protein transport machinery (Tic110 and STIC1/ALB4), demonstrating the specificity of the detected TOC protein interactions. Similarly, no precipitation of TOC proteins was observed in the negative control experiment conducted using water.

3.3 Discussion

The involvement of the SP1 E3 ligase and hence the UPS in regulating the physiological state of the chloroplast protein import machinery has been established and characterized in detail using *Arabidopsis thaliana* (Ling *et al.*, 2012). In that study, two other chloroplast-localized RING E3 ligases were identified as the homologues of SP1, in the same species.

An initial analysis of *SPL1* (testing its ability to complement the *sp1* mutation) did not support functional redundancy to *SP1* (Ling *et al.*, 2012). To extend this analysis, I employed a reverse genetics approach to characterize the gene's function in greater detail. As a result, a gene knockout of *SPL1* in the Toc33-deficient *ppi1* mutant was identified; in this double mutant, *spl1* did not cause any perturbation of chloroplast development or the plant's phenotype, relative to *ppi1* (Figure 3.3, 3.4 and 3.5). This was in accordance with the previously-published results (Ling *et al.*, 2012). Interestingly, there was an increase in the level of the Toc159 receptor in *spl1 ppi1*, similar to that seen in *sp1 ppi1* (Figure 3.6), although apparently this was not enough to elicit a change in the phenotype of the plants. Overall, my results in this chapter, together with the published results of SPL1 (Ling *et al.*, 2012), supported the hypothesis that SPL1 may have different targets from SP1, and roles other than as a regulator of the chloroplast protein import machinery under standard growth condition.

The analysis on *SPL2*, however, yielded a very different outcome. Complementation analysis of the *sp1 ppi1* double mutant based on overexpression of the native *SPL2* gene gave positive results, as the resulting plants appeared slightly paler than the *sp1 ppi1* progenitor but still greener than the *ppi1* single mutant (Figure 3.1). This partial complementation result suggested an involvement of

SPL2 in protein import machinery regulation, similar at least to some extent to the role of *SP1*. This hypothesis was further supported by the suppression of the *ppi1* phenotype (albeit rather weakly) upon reduction of the expression level of the native *SPL2* gene (Figure 3.3). The *SPL2-KD ppi1* double mutant exhibited enlarged chloroplasts (Figure 3.5), relative to *ppi1*, which can most likely be attributed to the observed increase in TOC protein levels (Figure 3.6) and the consequent improvement in the ability of the organelles to import proteins from the cytosol (Figure 3.7), which is necessary for the plastid development (Jarvis & López-Juez, 2013). It was noted that the phenotypic suppression of *ppi1* mutant by *SPL2-KD* is much weaker than that seen in *sp1 ppi1*. This could be due to incomplete knockout of the *SPL2* gene, as approximately 20-30% of the transcript still remained (although no information is available on the abundance of the SPL2 E3 ligase protein ultimately produced). Alternatively, the weaker suppression may reflect the fact that a functional *SP1* gene was present in the *SPL2-KD* line, which could well be the dominant regulator of the chloroplast protein import machinery. The fact that no suppression was observed when *SPL2* expression was reduced in other *toc* mutants (Figure 3.9) is supportive of this notion.

When the *SPL2-KD* construct was introduced into a hypomorphic mutant allele of *toc159* (*ppi2-3*), the mutant phenotype was not suppressed. Instead, an enhancement of the pale phenotype was observed (Figure 3.8). The immunoblot data showed an increase in the level of Toc75 protein in the *SPL2-KD ppi2-3* background, which is in line with what was observed in the *SPL2-KD ppi1* mutant (Figure 3.6), and with what was described previously for *sp1 ppi1* (Ling *et al.*, 2012). As a result, the reason for the enhanced chlorosis seen in this double mutant is unclear. One possibility is that it is a pleiotropic effect linked to the possibility that the

SPL2 E3 ligase has other substrates in addition to the TOC machinery. Alternatively, it may be technical peculiarity linked to a particular aspect of the *ppi2-3* missense mutation.

The specificity of the suppression mediated by *SPL2-KD* on *ppi1* was indicated by genetic interaction studies with mutants affecting to the TIC machinery of the inner membrane (i.e., *tic40* and *hsp93-V*), which showed no effects of *SPL2-KD* on the phenotype of either mutant (Figure 3.9).

The findings of my genetic experiments correlated very well with the results of the immunoprecipitation assays (Figure 3.10), which provided information on the interacting partners of the SPL2 E3 ligase. In these experiments, both SPL2 and SP1 were able to pull down components of the TOC machinery (Toc159, Toc75 and Toc33), but not components of the internal protein transport machinery (Tic110, STIC1/ALB4). In combination with the genetic results, these data strongly supported the notion that TOC components are substrates of SPL2, as they are of SP1.

Overall, the analyses presented in this chapter provided significant new insight into the role of SPL2, and its involvement in the regulatory mechanisms governing the chloroplast protein import machinery. As for SPL1, however, its exact functional role remained unclear.

3.4 Figures

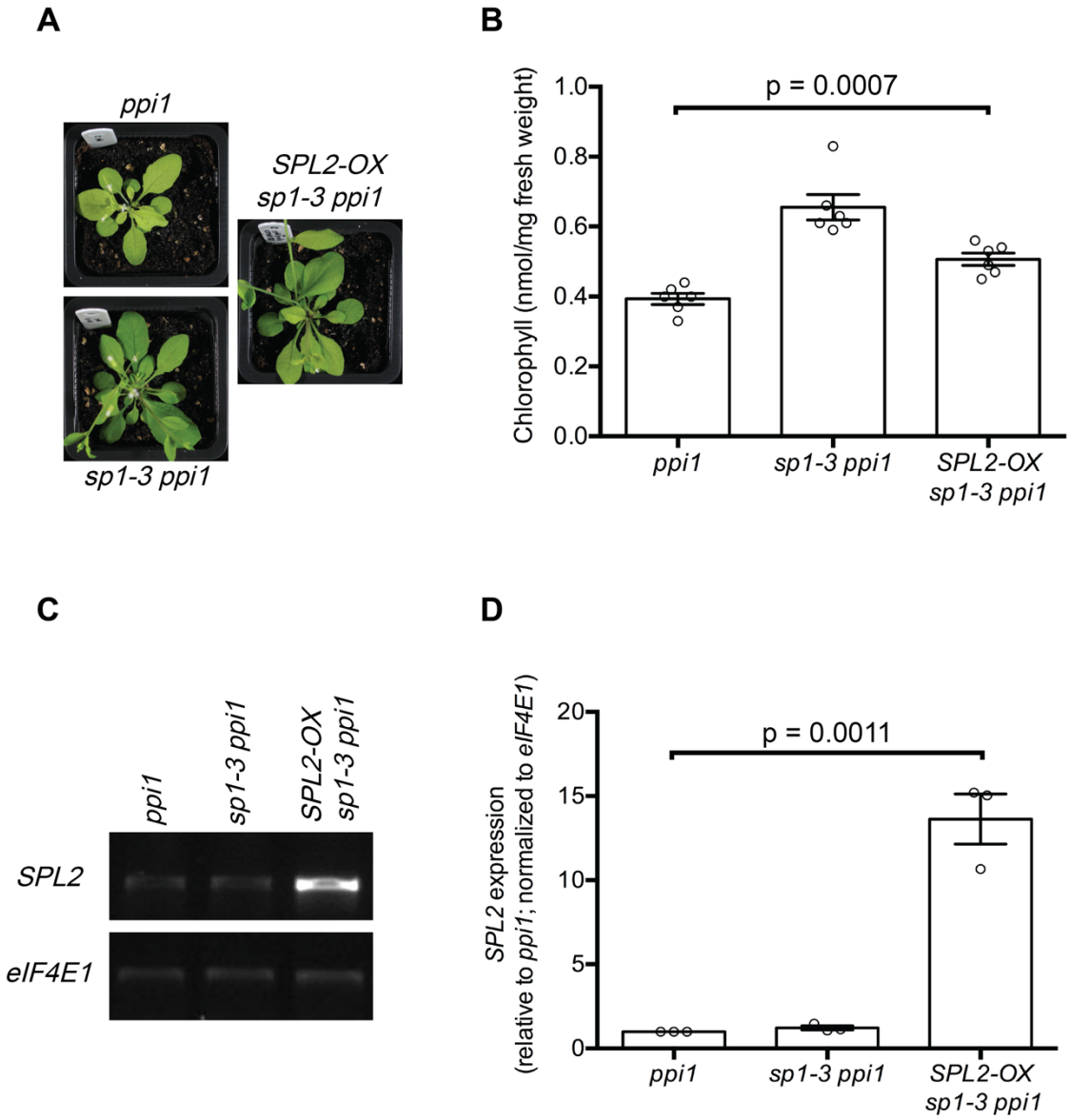


Figure 3.1 Analysis of the effect of *SPL2* overexpression on the *sp1 ppi1* suppressor phenotype. The *SPL2* CDS was cloned downstream of constitutive CaMV 35S promoter in the pB2GW7 binary vector. The resultant construct was used to transform *sp1-3 ppi1* by floral dipping using *Agrobacterium tumefaciens*. Positive transformants were screened for by germinating the seeds on MS media supplemented with phosphinothricin, and a suitable line was selected. (A) Phenotypic analysis of the selected *SPL2-OX* transformant. Plants were grown

under standard conditions on soil for 35 days before photography; representative individuals are shown. (B) Chlorophyll content was measured using a Konica-Minolta SPAD-502 meter. Meter values were converted to chlorophyll per fresh weight using an established equation. Values shown are means (n=6) with open circles indicating individual data points and error bars indicating SEM. (C, D) Semi-quantitative RT-PCR analysis of the *SPL2* expression in the selected *SPL2-OX* transgenic line. Total RNA samples were analysed by RT-PCR using gene-specific primers for *SPL2* and the reference gene *eIF4E1*. Amplicons from three independent experiments were quantified using Aida software. Values shown are means (n=3) with open circles indicating individual data points and error bars indicating SEM. The p values in B and D were derived from a Student's *t*-test of the means of the indicated genotypes.

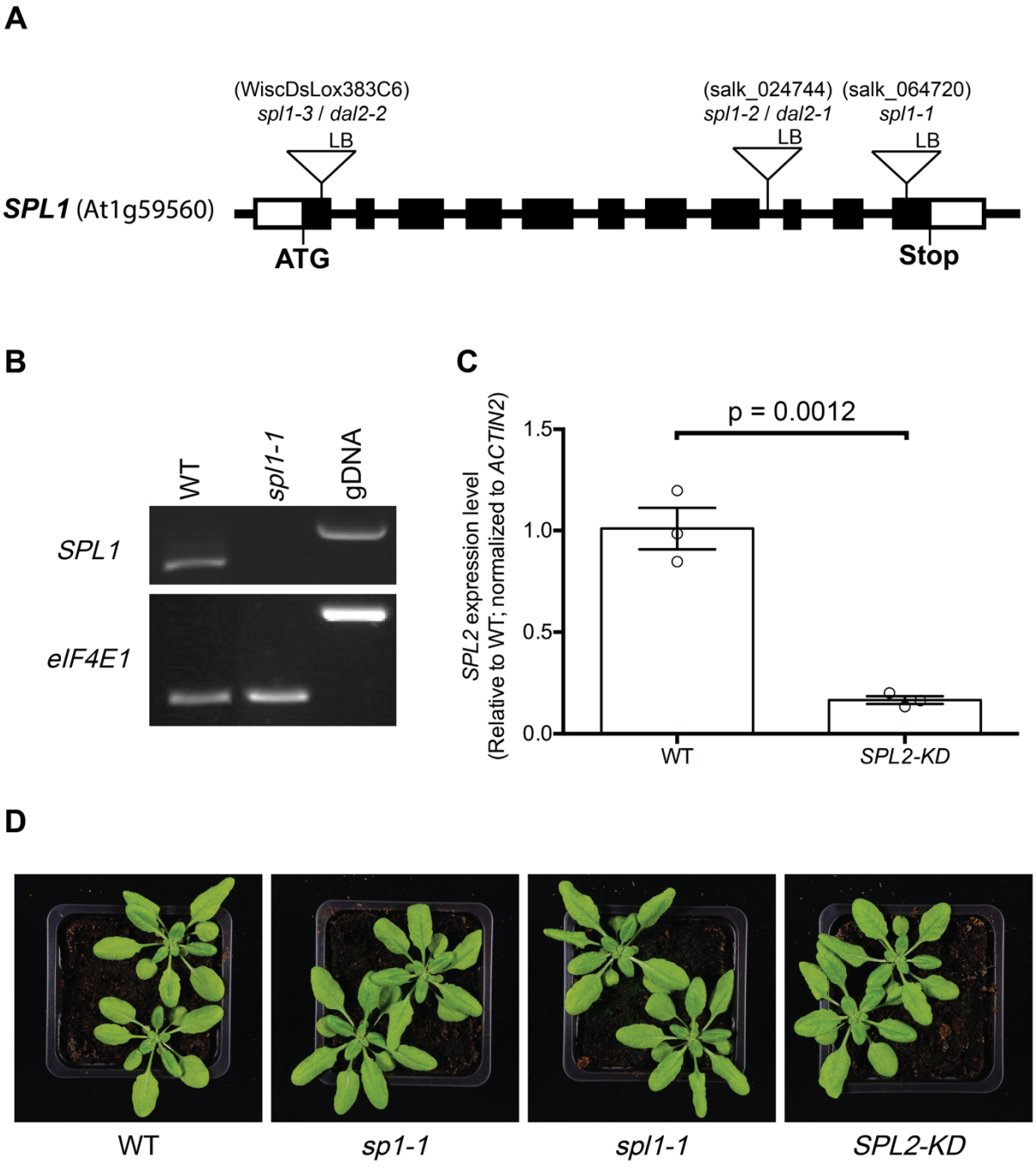


Figure 3.2 Identification and analysis of T-DNA mutant or knockdown lines to study the roles of the *SPLs* in chloroplast biogenesis. (A) Schematic representation of the *SPL1* genomic locus (gene model At1g59560) indicating the positions of the T-DNA insertion sites. Protein-coding exons = black boxes; untranslated regions = white boxes; introns = black lines between the boxes; LB = T-DNA left border sequence;

ATG = translation initiation codon; Stop = translation termination codon. (B) Analysis of *SPL1* mRNA expression by RT-PCR. Total RNA samples were isolated from the indicated genotypes and analysed by RT-PCR using gene-specific primers for *SPL1* and reference gene *eIF4E1*. Wild-type genomic DNA (gDNA) was included as a control. (C) Analysis of *SPL2* mRNA expression by qRT-PCR. Total RNA samples were analysed by qRT-PCR using gene-specific primers for *SPL2* and the reference gene *ACTIN2*. Values shown are means (n=3) with open circles indicating individual data points and error bars indicating SEM. The p values were derived from a Student's *t*-test of the means of the indicated genotypes. (D) Phenotypic analysis of wild-type, *sp1-1*, *spl1-1* and *SPL2-KD* plants. The plants were grown under standard growth conditions on soil for 28 days before photography; representative individuals are shown. The three *spl1* alleles were phenotypically identical, and so only *spl1-1* is presented as a representative allele.

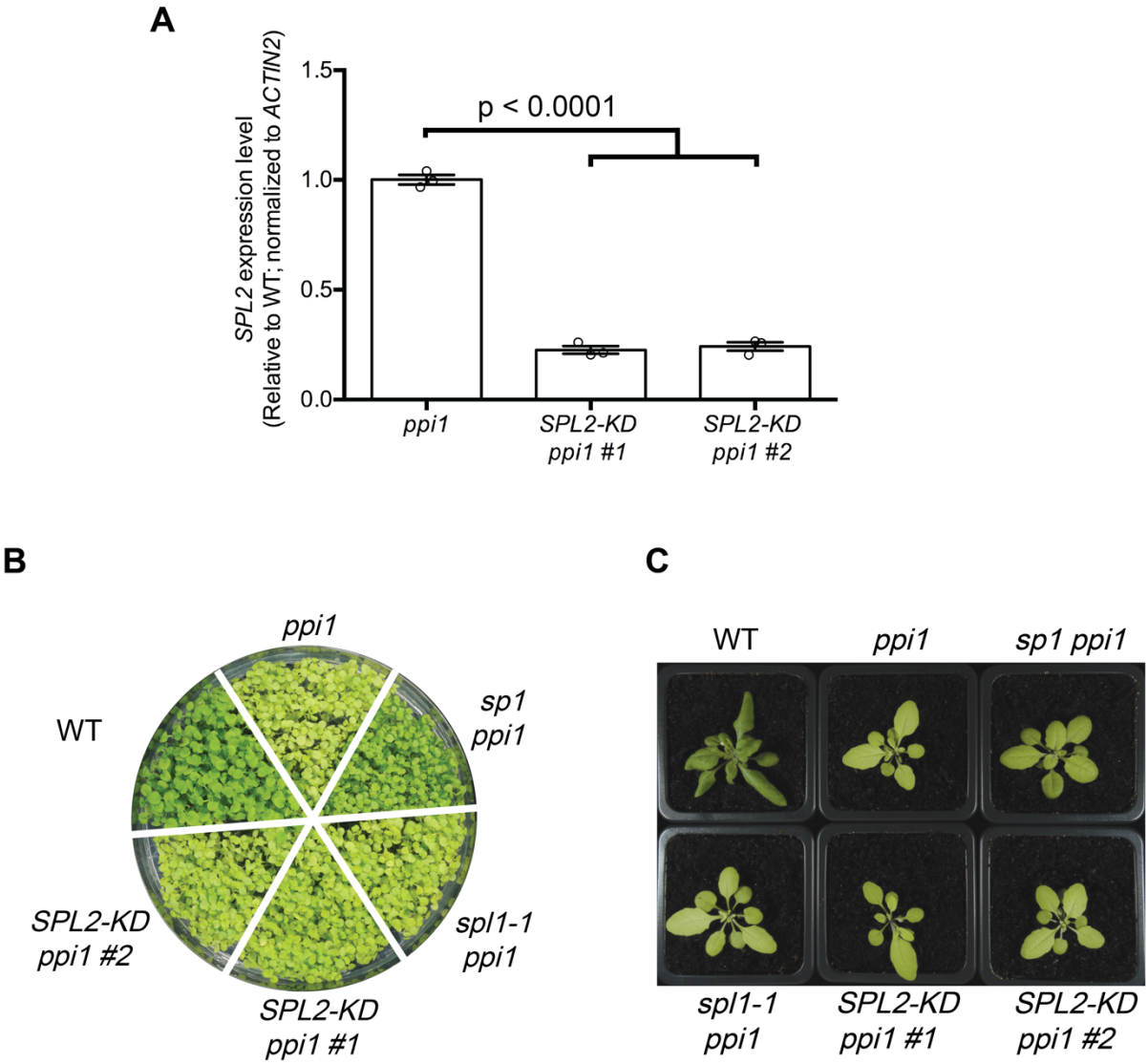

A
SPL2 expression level
(Relative to WT; normalized to ACTIN2)
1.5
1.0
0.5
0.0
p < 0.0001
ppi1
SPL2-KD
ppi1 #1
SPL2-KD
ppi1 #2
B
ppi1
WT
sp1
ppi1
SPL2-KD
ppi1 #2
spl1-1
ppi1
SPL2-KD
ppi1 #1
C
WT
ppi1
sp1 ppi1
spl1-1
ppi1
SPL2-KD
ppi1 #1
SPL2-KD
ppi1 #2
D
Chlorophyll (nmol/mg fresh weight)
1.5
1.0
0.5
0.0
p = 0.0002
p = 0.0005
WT
ppi1
sp1
ppi1
spl1-1
ppi1
SPL2-KD
ppi1 #1
SPL2-KD
ppi1 #2

Figure 3.3 Genetic analyses to assess for possible suppression of *ppi1* mutation. (A) Assessment of *SPL2* expression in knockdown lines by qRT-PCR. Total RNA samples were analysed by qRT-PCR using gene-specific primers for *SPL2* and the reference gene *ACTIN2*. Values shown are means (n=3) with open circles indicating individual data points and error bars indicating SEM. (B) Phenotypic analysis of young seedlings. Wild-type, *ppi1*, *sp1 ppi1*, *spl1-1 ppi1*, *SPL2-KD ppi1 #1* and *SPL2-KD ppi1 #2* plants were grown side-by-side on MS agar medium under standard growth conditions for two weeks before photography. (C) Phenotypic analysis of 28-day-old rosette-stage plants. Seedlings of the indicated genotypes from B were transferred onto soil and allowed to grow for an additional two weeks under standard growth conditions before photography; representative individuals are shown. (D) Chlorophyll content analysis of 10-day-old seedlings. Chlorophyll levels in plants grown on MS agar medium B were quantified photometrically following extraction in DMF. Values shown are mean (n=4) with open circles indicating individual data points and error bars indicating SEM. The p values in A and D were derived from a Student's *t*-test of the means of the indicated genotypes.

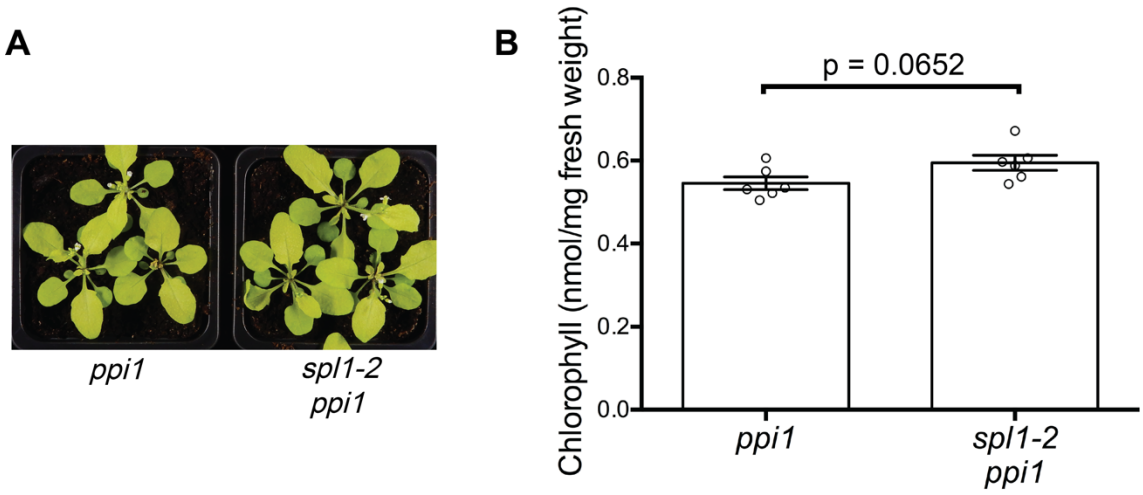


Figure 3.4 Comparison of the effects of two different *spl1* alleles on the *ppi1* mutation. (A) Phenotypic analysis of the *ppi1* and *spl1-2 ppi1* double mutants. Plants were grown under standard growth conditions on soil for 28 days before photography; representative individuals are shown. (B) Chlorophyll content analysis of the *ppi1* and *spl1-2 ppi1* double mutants. The leaves of plants similar to those shown in A were analysed using a Konica-Minolta SPAD-502 meter. Meter values were converted to chlorophyll per fresh weight using an established equation. Values shown are means (n=6) with open circles indicating individual data points and error bars indicating SEM. The p values were derived from a Student's *t*-test of the means of the indicated genotypes.

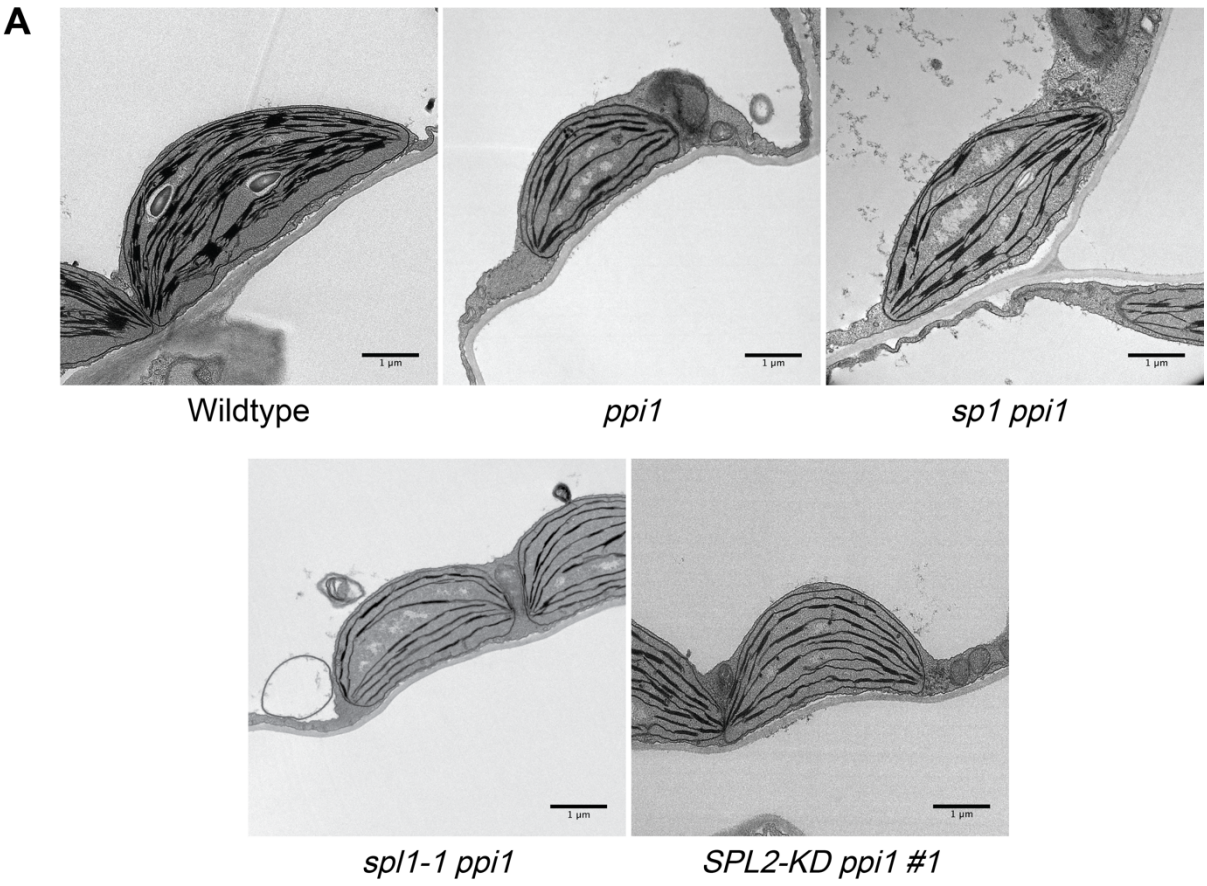


Figure 3.5 Chloroplast ultrastructural analysis by transmission electron microscopy. (A) Transmission electron micrographs of typical chloroplasts in rosette leaves of 35-day-old plants of the indicated genotypes are shown. Scale bars show 1 µm. (B)

Quantitative analysis of the micrographs represented in A. Box and whisker plots showing the calculated cross-sectional area of the chloroplasts in each genotype. The data are based on measurements from 30 chloroplasts per genotype (indicated as open circles). The central box represents values between the first and the third quartiles and the middle line represents the median value; the whiskers extend to the minimum and maximum values. The p values were derived from a Student's *t*-test of the means of the indicated genotypes.

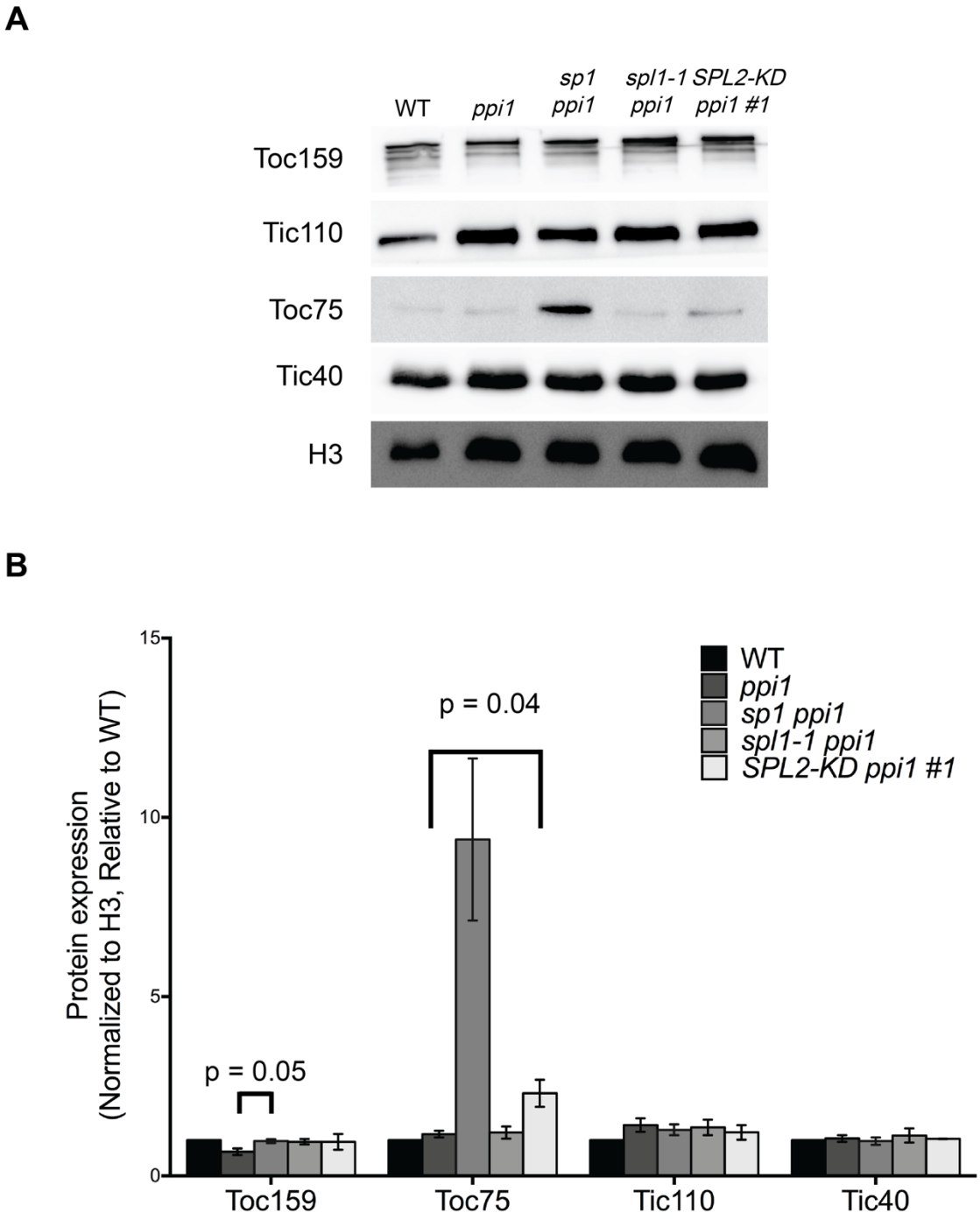


Figure 3.6 Immunoblot analysis of TOC and TIC protein levels in the *spl ppi1* double mutants. (A) Total protein extracts from two-week-old seedlings of the indicated genotypes were resolved by SDS-PAGE and transferred onto nitrocellulose membrane for detection using various antibodies directed against components of the chloroplast protein import machinery. (B) Bands of interest from three independent experiments similar the one shown in A were quantified. Values shown are means

(n=3) with error bars indicating SEM. Histone H3 was used as a loading control. The p values were derived from a Student's *t*-test of the means of the indicated genotypes.

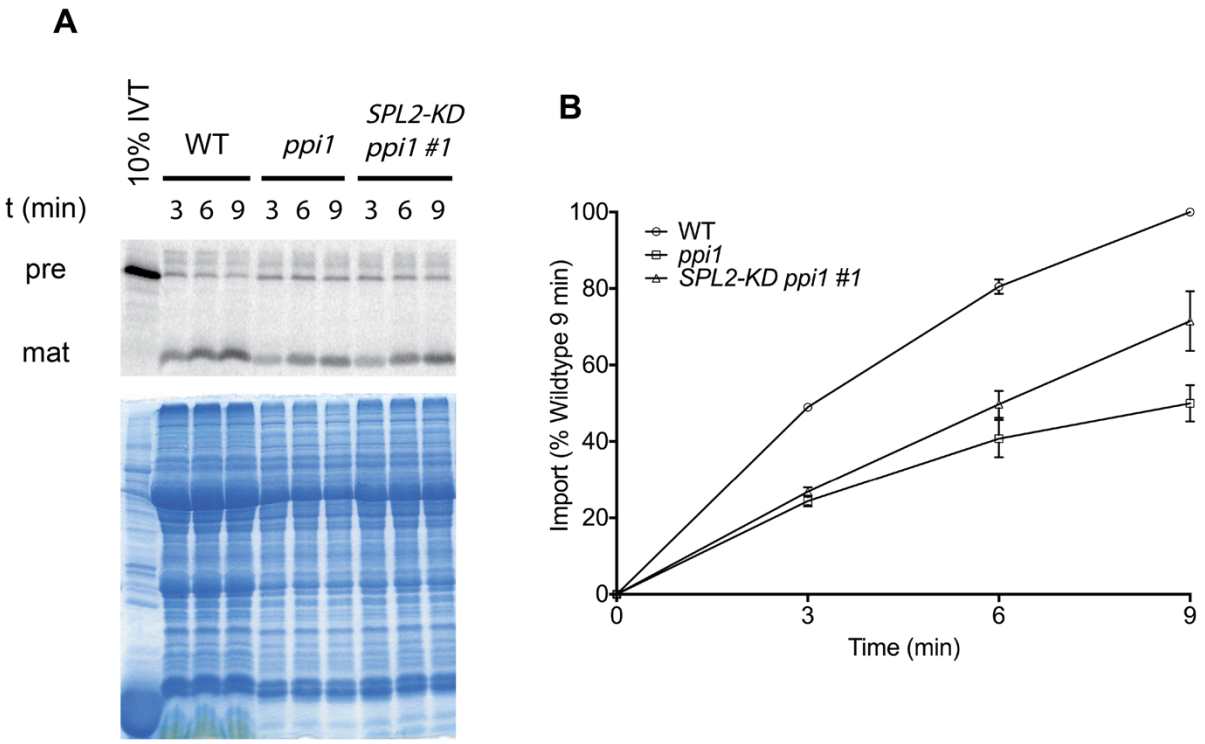


Figure 3.7 Analysis of chloroplast protein import efficiency in the *SPL2-KD ppi1* double mutant. (A) Isolated chloroplasts from 14-day-old wild-type, *ppi1* and *SPL2-KD ppi1 #1* plants were used to conduct *in vitro* protein import assays. *In vitro* translated (IVT) Rubisco small subunit precursor (pre) was used as the substrate, and the rate of import and processing to the mature (mat) form was assessed over a three point time-course (3, 6 and 9 minutes). (B) The maturation of precursors at each time point was quantified and expressed as a percentage of the amount of protein imported into wild-type chloroplasts at 9 minutes. Values shown are means of 3 independent import experiments with error bars indicating SEM.

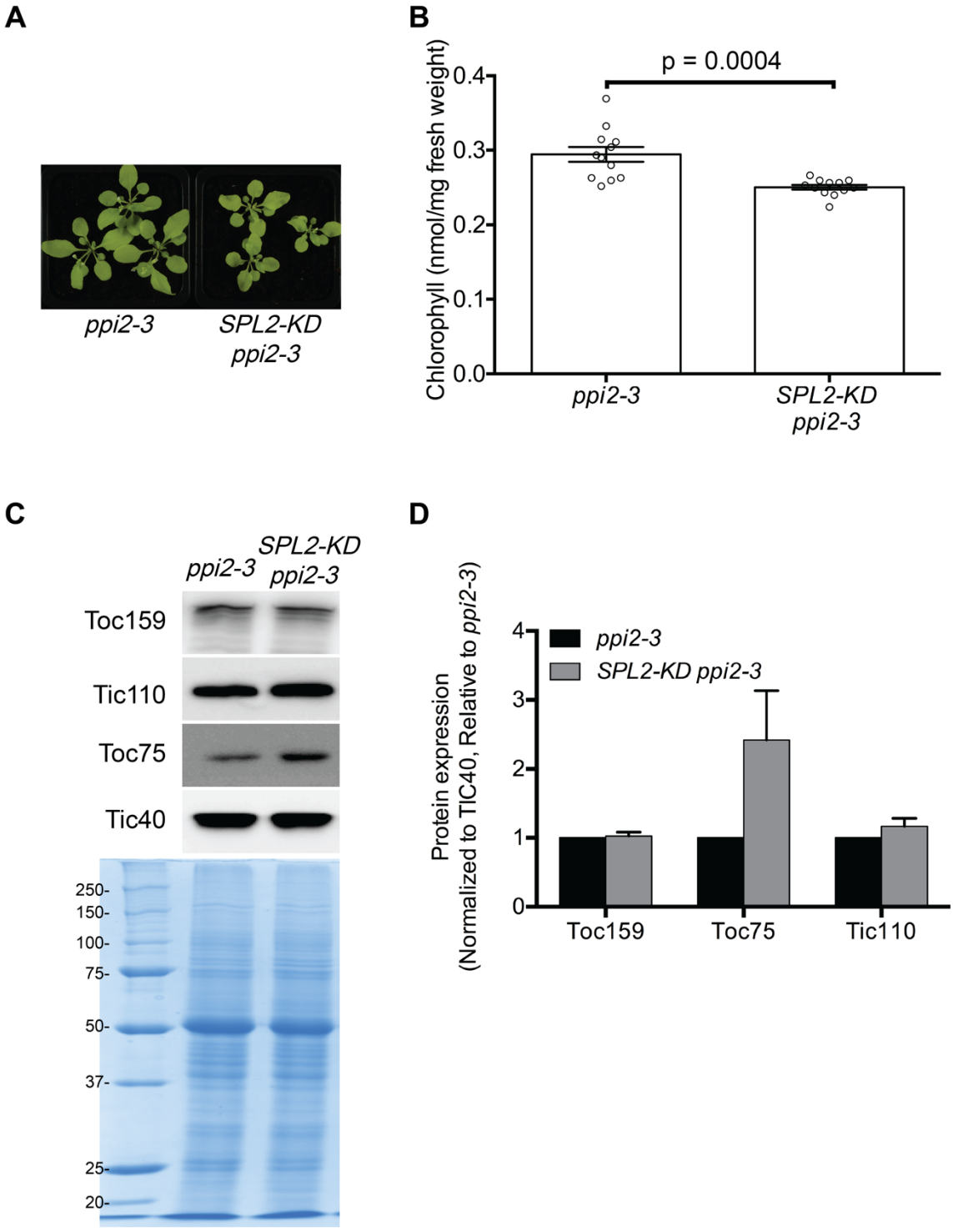


Figure 3.8 Genetic analysis of the interaction between *SPL2-KD* and the *ppi2-3* mutation. (A) Phenotypic analysis of *ppi2-3* and *SPL2-KD ppi2-3* plants. Plants were grown on soil under the standard growth conditions for 28 days before photography; representative individuals are shown. (B) Chlorophyll content analysis of plants similar to those shown in A. Measurements were made using a Konica-Minolta

SPAD-502 meter, and the meter values were converted to chlorophyll per fresh weight values using an established equation. Values shown are mean (n=12) with open circles indicating individual data points and error bars indicating SEM. The p value was derived from a Student's *t*-test of the means of the indicated genotypes. (C) Immunoblot analysis of TOC and TIC protein levels in plants similar to those shown in A. Total protein extract samples (20 μg) from each genotype were resolved by SDS-PAGE and transferred onto nitrocellulose membrane for detection using various antibodies. An equivalent Coomassie-stained gel was prepared to show equal loading of the samples. (D) Quantification of the data represented in C. Bands of interest from three independent experiments similar to the one shown in C were quantified. Values shown are means (n=3) with error bars indicating SEM. Tic40 was used as a normalization control in this case.

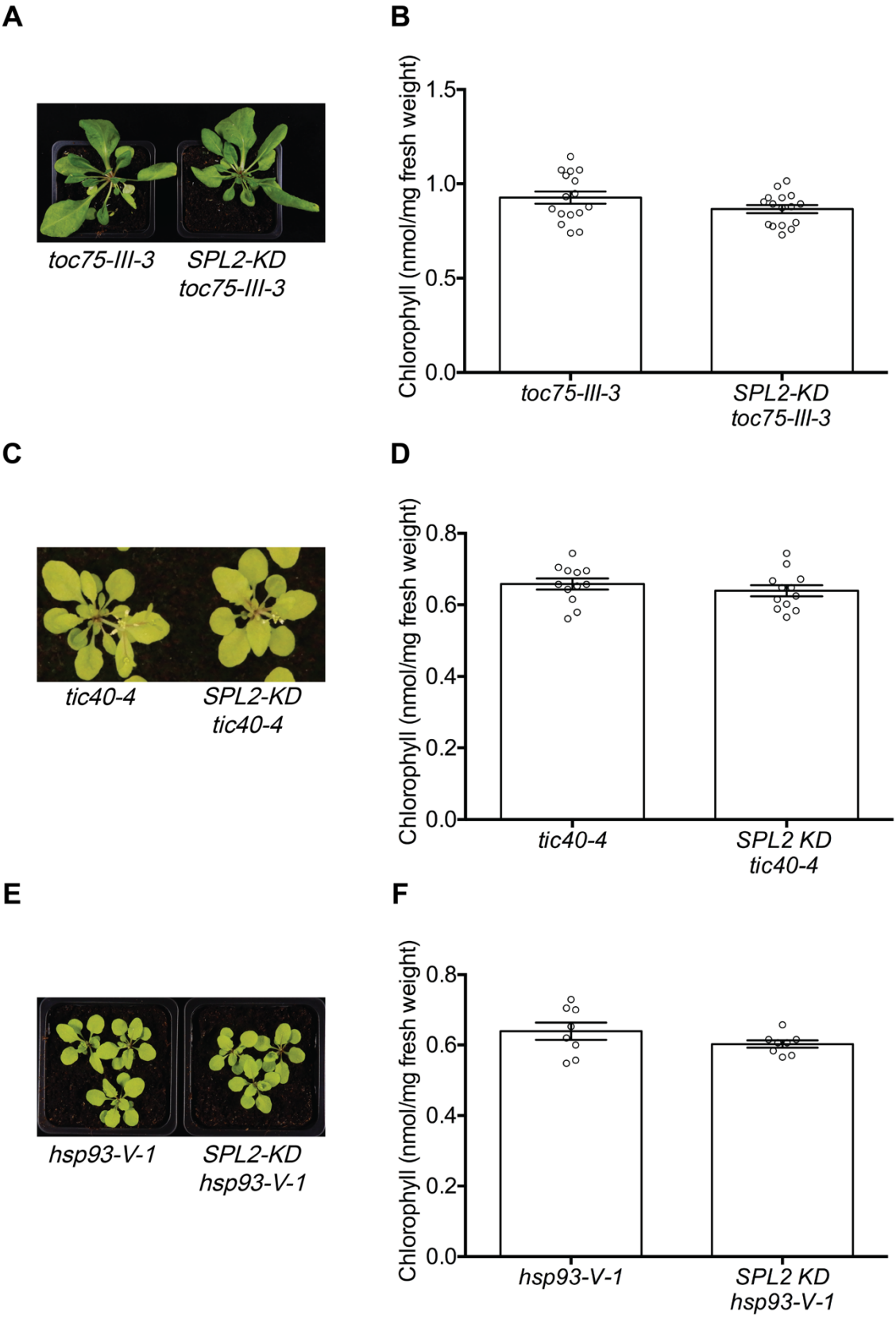


Figure 3.9 Genetic analysis of the interaction of *SPL2-KD* with other chloroplast protein import mutations. (A, C, E) Phenotypic analysis of *toc75-III-3*, *tic40-4*, *hsp93-*

V-1 and their corresponding *SPL2-KD* double mutants. Plants were grown on soil under the standard growth condition for 28 days before photography; representative plants are shown. (B, D, F) Chlorophyll content analysis of the same genotypes. The leaves of plants similar to those shown in A, C and E were analysed using a Konica-Minolta SPAD-502 meter. Meter values were converted to chlorophyll per fresh weight values using an established equation. Values shown are means with open circles indicating individual data points (n = 16 (B), 12 (D) or 8 (F)) and error bars indicating SEM.

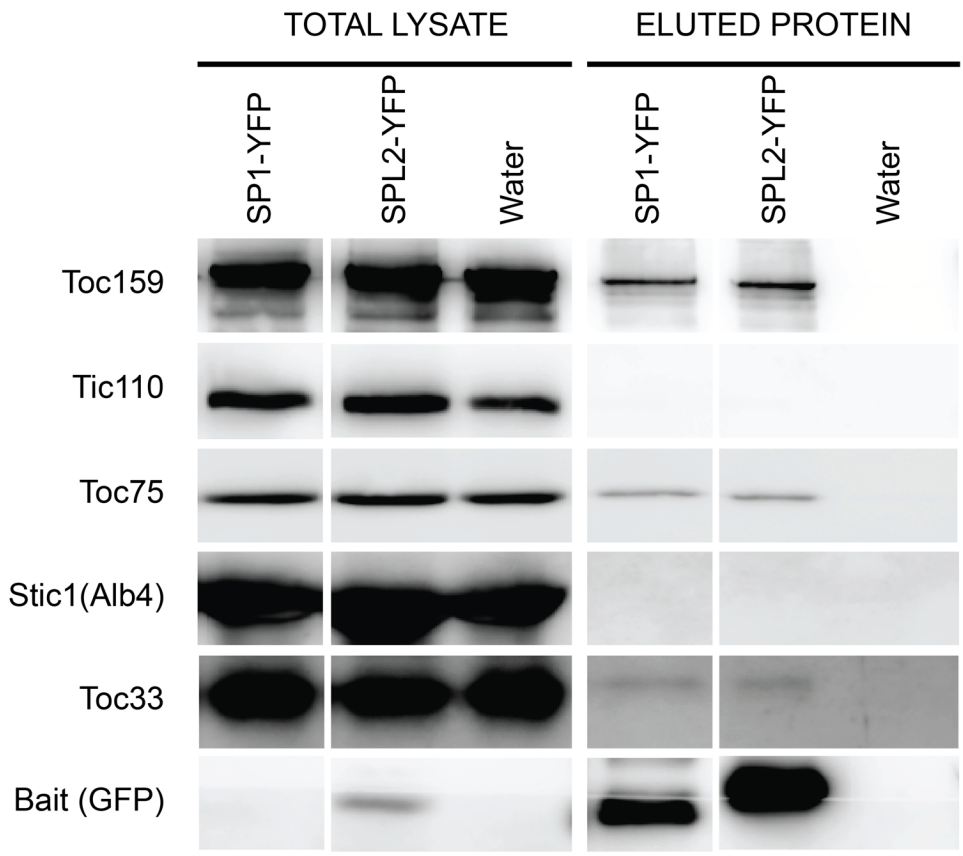


Figure 3.10 Immunoprecipitation of TOC components with YFP-tagged SPL2 from the protoplast. THE SPL2 CDS was cloned downstream of the 35S promoter in the pK7YWG2 binary vector. The resulting *SPL2-YFP* was used to transform *Arabidopsis* protoplasts. In parallel, the *SP1-YFP* construct was used as a positive control and water as a negative control, to confirm the TOC proteins does not non-specifically associate with the GFP-Trap MA magnetic beads. A 20 μl of the total lysate (before immunoprecipitation steps) and the eluted protein were analysed by immunoblotting using antibodies against: the YFP tag, to verify the enrichment of SPL2-YFP and SP1-YFP; the TOC components, to detect substrate interaction; the Tic110 and Stic1/Alb4, to act as a negative control and confirm the specificity of the detected interactions.

Chapter 4

Results II

FUNCTIONAL ANALYSES OF THE SP1-LIKE E3 LIGASES

Abstract

The process of substrate ubiquitination by the SP1 E3 ligase is attributed in part to the presence of an intact zinc finger motif within the RING finger domain. This motif is important for successful interaction with the respective E2 conjugase. In the previous chapter, mutation of the *SPL1* gene was reported to have no effect on the *ppi1* mutant phenotype despite it being described as the closest relative of *SP1*. An evolutionary analysis was therefore conducted by constructing a phylogenetic tree to infer homology relationships in the Viridiplantae. Interestingly, while SP1 and SPL2 evolved independently from a common ancestor to form monophyletic groups represented in all of the tested species, SPL1 was suggested to have arisen from a duplication event within the SP1 clade after the separation of Chlorophyta and land plants. This motivated me to investigate whether a functional ubiquitination activity was retained after the duplication event by analysing SPL1 (and SPL2) biochemically using *in vitro* ubiquitination assays. The results suggested that SPL1 is a null E3 ligase. Nonetheless, further analyses indicated that SPL1 has an important role during leaf senescence. Knockout mutant *spl1* plants exhibited accelerated senescence similar to *SP1-OX* transgenic plant lines. Although the exact molecular mechanism is unclear, public microarray data together with further *in vitro* ubiquitination results suggested that the presence of SPL1 has a negative impact on SP1 ubiquitination activity, presumably through direct competitive inhibition of binding to the E2 conjugase. In contrast, the SPL2 E3 ligase exhibited similar activity to SP1, as both proteins were able to carry out self-ubiquitination, and both influenced leaf senescence in a similar way.

CHAPTER 4:

FUNCTIONAL ANALYSES OF THE SP1-LIKE E3 LIGASES

4.1 Introduction

The process of ubiquitination relies on the sequential action of three enzymes: E1 activase, E2 conjugase and E3 ligase (Vierstra, 2009). The transfer of ubiquitin to a substrate destined for degradation by the 26S proteasome primarily involves the formation of a complex consisting of E2 conjugase, E3 ligase and the substrate although other factors may be involved such as adaptor proteins, activator proteins, and others (Deshaies & Joazeiro, 2009).

For the chloroplastic RING-type E3 ligase SP1, ubiquitination activity is dependent on the presence of a functional and intact RING finger (RNF) domain (Ling *et al.*, 2012). This domain promotes the formation of an E2:E3 RING complex which is a feature typical of the RING E3 ligase family (Lorick *et al.*, 1999; Stone *et al.*, 2005). Mutation at the conserved RING E3 metal ligand positions, in particular, can disrupt the complex assembly, preventing substrate ubiquitination from taking place (Brzovic *et al.*, 2003). The SP1 E3 ligase was demonstrated to be functional and able to carry out self-ubiquitination *in vitro* (Ling *et al.*, 2012), suggesting an ability to carry out self-regulation. Factors that might influence the activity of SP1 include developmental cues (Ling *et al.*, 2012) and the generation of reactive oxygen species (ROS) during unfavourable growth conditions (Ling & Jarvis, 2015b). The SP1 autoubiquitination data raise an interesting question as to whether the SP1 homologues behave similarly. This is pertinent because the findings presented in Chapter 3 of this thesis suggested that SPL1 is functionally distinct from SP1. This is

unexpected because SPL1 is more closely related to SP1 than SPL2, and SPL2 was demonstrated to be partially redundant with SP1. Therefore, another possible explanation is that SPL1 may be a non-functional E3 ligase or a pseudogene.

Plastids are very dynamic organelles which undergo changes in their proteomes in order to transition from one form to another during different plant developmental stages (Jarvis & López-Juez, 2013). Loss of SP1 (in *sp1* mutants) and the overexpression of SP1, in *Arabidopsis*, were both shown to have an effect on the rate of plastid type conversion, during de-etiolation and senescence (Ling *et al.*, 2012). These findings demonstrated the involvement of SP1 as a key chloroplast protein import machinery regulator, dictating the progression of plastid conversions via organellar proteome changes. As *SPL2* was observed to be partially redundant with *SP1* (Chapter 3), it would be interesting to determine whether *SPL2* is also involved in these dynamic processes of plastid biogenesis.

Therefore, in this chapter, I will describe investigations concerning the ubiquitination activity of the SP1 homologues, as well as their involvement during the plastid transition processes as was done in SP1.

4.2 Results

4.2.1 Phylogenetic analysis of the SP1 homologues

As a starting point for these studies, a phylogenetic tree was constructed to shed light on the evolutionary and homology relationships between the E3 ligases.

BLAST searches for plant species with a complete RING sequence were conducted at Phytozome, Ensembl Plants and NCBI in order to identify homologues of SP1, SPL1 and SPL2 for analysis. Reciprocal BLAST searches were also performed against the *Arabidopsis thaliana* sequences using the hits with the lowest

Expect (E) value from each species. The human mitochondrial outer-membrane protein MULAN/MAPL was included in the analysis as an outgroup; this protein was reported to be related to SP1 previously (Ling *et al.*, 2012).

SP1 orthologues were observed throughout the plant kingdom, including the Chlorophytes as the most basal species (Figure 4.1). In contrast, SPL1 orthologues were limited to the Brassicaceae family of flowering plants, while SPL2 orthologues were ubiquitous in the Viridiplantae. Unlike the situation with SP1, SPL2 was not present in any of the representative species of the Chlorophyta, and so the liverwort *Marchantia polymorpha* was observed to be the most basal species with SPL2.

The phylogenetic analysis confirmed the expected close relationship between SP1 and SPL1. The two sequences shared a common, relatively recent ancestor right after the divergence of the Chlorophyta and Embryophyta within the SP1 clade. However, the number of species represented in the SPL1 clade was greatly reduced, relative to the SP1 clade, with only members of the Brassicaceae family present. The phylogeny suggests that those species outside of the Brassicaceae family have lost SPL1. It is very unlikely for the family to gain SPL1 as the acquisition of the homologue would mean placing the relevant clade within the SP1 Brassicaceae group instead. The branch lengths within the SP1 Embryophyta clade for vascular plants were noticeably very short, indicating a very low frequency of amino acid substitutions between the species. This supports the importance of SP1 as a dedicated and highly conserved E3 ligase required for regulating the turnover of the TOC machinery during plant development (Ling *et al.*, 2012) and unfavourable growth conditions (Ling & Jarvis, 2015b). In contrast, the branch lengths in the SPL1 clade were much longer, indicating a more accelerated divergence. This may be

indicative of neofunctionalization or pseudogene formation, especially as SPL1 did not show any functional redundancy with SP1 as discussed in Chapter 3.

The SPL2 sequences formed their own monophyletic group, derived from a distant common ancestor that also gave rise to SP1 (and SPL1). The branch lengths in the SPL2 clade were much more varied than those in the SP1 clade. This implies that SPL2 has evolved at different rates in different species; suggesting greater conservation in some species and more divergence in others. In spite of that, atSPL2 still retain a partial functional redundancy to atSP1. It is possible that atSPL2 may have acquired new functions and substrates in addition to those of the TOC machinery.

It is also noteworthy in the phylogeny (Figure 4.1) that certain species like mosses in SP1 and SPL2, and the Brassicaceae in SPL1, exhibited multiple copies of the E3 ligases.

4.2.2 Mutation of two amino acid residues alters the predicted conformation of SPL1 cytosolic domain

In order to shed a light on the role of the SPL1 protein, structural prediction studies were conducted by submitting the cytosolic domain (cyt) amino acid residues of the three SP1-related *Arabidopsis thaliana* proteins to the web server iTASSER (Zhang, 2008; Roy *et al.*, 2010). The predicted SP1cyt and SPL2cyt structures both showed a more globular conformation than that of SPL1cyt, with the conserved C3HC4 RING domain partially hidden by the arc structure in each case (Figure 4.2 A). The atSPL1cyt domain showed a much more elongated structure, when viewed at 90° rotation to the left (Figure 4.2 A). As a result, the RING domain of SPL1cyt is much more exposed than that of SP1cyt and SPL2cyt.

To begin to understand the basis for the structural prediction differences described above, a multiple sequence alignment of the RNF domains of SP1-type ubiquitin E3 ligases from a number of species was constructed. This analysis revealed noticeable differences at two sites within the RNF domain of atSPL1, relative to the highly conserved SP1 sequences (Figure 4.2 B). There are negatively-charged glutamic acid residues in place of non-polar proline at position 461, and in place of positively-charged arginine at position 488 (numbers denoting the aligned sequences). When these two residues were mutated to match those in atSP1, the predicted structure of the resulting atSPL1cyt-E305P-E327R became much more globular and compact, resembling that of atSP1cyt and atSPL2cyt (Figure 4.2 A).

To assess the function of the above-described mutated SPL1, and address the hypothesis that the atSPL1 might be a non-functional E3 ligase, *35S:SPL1-E305P-E327R* transgenic lines were generated by *Agrobacterium*-mediated transformation by floral dipping of the *ppi1* mutant (Figure 4.3 A and B). However, no difference was observed between the visible phenotypes of *35S:SPL1-E305P-E327R ppi1* and *ppi1*, and this was confirmed by making chlorophyll measurements (Figure 4.3 C). This indicated that the two amino acid residue sites might not be responsible for the lack of genetic activity related to the *spl1* mutation presented in Chapter 3.

4.2.3 SPL1 is a non-functional E3 ligase *in vitro*

It was demonstrated previously that the cytosolic domain of SP1 was the minimal requirement for *in vitro* ubiquitination activity, and actually more efficient when compared to the full-length SP1 with deleted transmembrane domains (Ling *et al.*, 2012). For this reason, I employed cytosolic domains in a similar approach to

study the ubiquitination activities of SPL1 and SPL2. The cDNA sequences encoding the SPL1cyt and SPL2cyt domains were cloned into an N-terminal GST fusion protein expression vector, and then transformed into *E. coli* cells for protein expression. To generate negative controls, the conserved histidine residue of the C3HC4 RING domain was substituted to tyrosine, generating the mutant versions, SPL1cyt-H308Y and SPL2cyt-H348Y. Based on previous studies (Gao *et al.*, 2015), it is know that such mutants are inactive owing to the disruption of the RING domain structure.

All four constructs were expressed and purified alongside Free-GST, GST-SP1cyt and GST-SP1cyt-C330A (Ling *et al.*, 2012). The purified recombinant proteins were run on an SDS-PAGE with BSA protein standards to estimate their concentrations (Figure 4.4 A). From that, an equal amount (1 μg) of each recombinant protein was incubated in a ubiquitination reaction mixture containing E1, E2 and HA-tagged ubiquitin (Ub).

The autoubiquitination activity of each protein was determined by immunoblot analysis of the reactions using an antibody raised against the HA-tag, to detect the presence of the HA-tagged ubiquitin molecule. The presence of various high molecular weight bands of varying size (polyubiquitin conjugates) indicated that ubiquitination had occurred, and this was observed for both GST-SP1cyt and GST-SPL2cyt (Figure 4.4 B). As expected, the RING mutants did not show similar activity, confirming the specificity of the results and indicating the importance of the mutated residues for functionality. Most interestingly, GST-SPL1cyt did not show any ubiquitination activity, nor did the mutant version of it. This may explain the lack of phenotypic effect of the *spl1* mutations in the genetic interaction studies in Chapter

3, as well as the published negative complementation-test results upon overexpressing SPL1 in the *sp1 ppi1* mutant background (Ling *et al.*, 2012).

4.2.4 *SPL1* and *SPL2* are both important during senescence

Manipulation of *SP1* expression has an effect on the rate of developmental processes linked to plastid-type conversions including de-etiolation and senescence (Ling *et al.*, 2012). The *sp1* mutant showed a delay in these processes, whereas overexpression of *SP1* accelerated them. As *SPL2* was earlier shown to display partial similarity of function to *SP1* (Chapter 3), I wished to test whether the SP1 homologues are involved in such dynamic processes. To do this, I decided to focus my analyses on leaf senescence.

For this work, an *SPL2-OX* transgenic line in the wild-type background was obtained by outcrossing the *SPL2-OX sp1-3 ppi1* line (Chapter 3) to wild-type plants. Homozygous lines for the *SPL2-OX* transgene were identified in the F_3 generation by monitoring resistance to the herbicide phosphinotricin, and the absence of *sp1-3* and *ppi1* T-DNA inserts was confirmed by PCR genotyping. The *SPL2-OX* construct was then tested by qRT-PCR to ensure the transgene was still overexpressing in the resulting line (Figure 4.5 A).

Wild-type, *sp1-1*, *spl1-1*, *SPL2-KD*, *sp1-1 spl1-1*, *SP1-OX* and *SPL2-OX* plants were grown on soil for four weeks under standard growth conditions. All plants at this stage showed phenotypes similar to that of the wild type (Figure 4.5 B). Selected leaves of similar sizes from each genotype were then covered with aluminium foil to promote darkness-induced senescence (Schelbert *et al.*, 2009; Ling *et al.*, 2012), and the plants were left to grow for a further five days under standard growth conditions before analysis.

Following the dark treatment, the effects of the different genotypes on senescence were assessed visually. The effects of *SPL2* manipulation resembled closely those of *SP1* manipulation; i.e. *SPL2-KD* showed a delay in leaf senescence, whereas *SPL2-OX* showed acceleration of the process (Figure 4.6 A). These observations were confirmed by chlorophyll content analysis post-treatment (Figure 4.6 B), as well as by following the photochemical efficiency of photosystem II (Figure 4.6 C and D). Interestingly, senescence induction was greatly enhanced in the *spl1-1* mutant, relative to wild type, much like it was in *SP1-OX* leaves (Figure 4.6 A and B*)*. This observation was consistent across different *spl1* mutant alleles (Figure 4.6 E). Moreover, the *sp1-1 spl1-1* double mutant also exhibited rapid senescence, to a comparable degree, which indicated that the effect of the *spl1* mutation is epistatic over that of the *sp1* mutation in this context (Figure 4.6 A, B, C and D).

To develop a better understanding of the role of SPL1 in leaf senescence, I conducted an analysis of the expression of all three *Arabidopsis SP1*-related genes using public microarray data using Genevestigator (Hruz *et al.*, 2008). The data revealed a marked decrease in *SPL1* expression during leaf senescence (Figure 4.7). In contrast, the level of *SP1* expression was elevated during senescence. Overall, these trends in gene expression were consistent with the observations made in the induced senescence experiments described above, and pointed towards a role for SPL1 in antagonizing SP1 function in this developmental process.

4.2.5 SPL1 negatively influences the ubiquitination activity of SP1 *in vitro*

In light of the data from the senescence and gene expression data described in Section 4.2.4 above, it was proposed that the SPL1 protein may have an inhibitory effect on the enzymatic activity of SP1 via an unknown mechanism, such as direct

competition for the E2 conjugase. To test this hypothesis, bacterially-expressed recombinant GST-SP1cyt and GST-SPL1cyt (prepared as described earlier) were mixed in a different ratios within the *in vitro* ubiquitination assay system that was presented in Section 4.2.3. The results showed a reduction in ubiquitination activity when GST-SPL1cyt was added in excess (Figure 4.8 A), which is consistent with the hypothesis that SPL1 is a negative regulator of SP1. Next, a control experiment was conducted using GST-SP1cyt-C330A in place of GST-SPL1cyt, to confirm that the effect on ubiquitination activity was not due to the presence of an excess non-functional E3 ligase (Figure 4.8 B). The results revealed no effect of GST-SP1cyt-C330A on ubiquitination activity, suggesting that the effect of SPL1 on SP1 is quite specific.

4.2.6 Overexpression of *SPL1* abrogates the effect of *SP1* overexpression on the *ppi1* mutant phenotype

The observations from Section 4.2.5 indicated that *SPL1* has a negative effect on *SP1* ubiquitination activity *in vitro*. To assess whether SPL1 exerts a similar influence *in vivo* (as was suggested by the data in Section 4.2.4), the effects of overexpression of *SP1* and *SPL1* on the *ppi1* mutant phenotype were assessed, either in isolation or in combination in the F_1 generation derived from crosses between the lines. For this work, the *SPL1-OX* transgene was needed in the *ppi1* background; this genotype was obtained by outcrossing the *SPL1-OX sp1-3 ppi1* line (Ling *et al.*, 2012) to *ppi1* and subsequently conducting appropriate genotyping experiments in the F_2 generation, and monitoring zygosity of the transgene by following phosphinotricin antibiotic resistance.

As expected, *SP1-OX ppi1* plants were very small and chlorotic, and had lower chlorophyll concentrations than *ppi1* (Figure 4.9 A and B, Ling *et al.*, 2012). In contrast, the *SPL1-OX ppi1* plants appeared slightly greener than the *ppi1* single mutant, and this was confirmed by chlorophyll content measurements (Figure 4.9 A and B); this partial recovery was also clearly seen in 5-day-old seedlings, where the *SPL1-OX ppi1* plants exhibited significantly larger cotyledons than *ppi1* (Figure 4.9 C). Most interestingly, the phenotype of the double overexpressors of *SP1* and *SPL1* in the *ppi1* background resembled that of *SPL1-OX ppi1* plants (Figure 4.9 A and B). This indicated that overexpression of the SPL1 was interfering with the ability of *SP1-OX* to accentuate the *ppi1* mutant phenotype, and so was in line with the *in vitro* results presented in Section 4.2.5.

4.3 Discussion

The SPL1 and SPL2 components were originally reported as homologues of SP1 due to their similarity in topology and amino acid sequence identity (Ling *et al.*, 2012). All three proteins possess two transmembrane domains separated by an intermembrane space domain, and a cytosolic domain where the conserved C3HC4 RING finger domain is located (Ling *et al.*, 2012). However, little was known about their true evolutionary relationships, appreciation of which will aid the understanding of their proposed roles as other regulators of chloroplast biogenesis. Some studies have attempted to construct a phylogeny for these E3 ligases, but these analyses were not very comprehensive or rigorous (Basnayake *et al.*, 2011; Pan & Hu, 2017). For example, analysis was focused on only two different species in the angiosperms: *Arabidopsis thaliana* and *Oryza sativa*; and therefore the amount of information that could be extracted from the phylogeny was very limited. To overcome this limitation,

I conducted extensive BLAST searches for homologues from species across the Viridiplantae, and used the sequences to construct the phylogeny presented in Figure 4.1. Interestingly, whilst the SP1 and SPL2 sequences were each observed to form a discrete monophyletic group, SPL1 was predicted to have arisen as a result of duplication event within the SP1 clade right after the separation of the Chlorophyta and the land plants.

Based on the phylogenetic analysis, it was proposed that the SPL1 E3 ligase might have evolved and acquired a new function that is different to SP1; this notion is supported by the longer branch lengths in the SPL1 clade than in the corresponding SP1 Brassicaceae clade. To investigate this theory, a structural prediction analysis was carried out on the cytosolic domain of the *A. thaliana* E3 ligases (Figure 4.2). The atSP1cyt and atSPL2cyt domains were predicted to have similar globular protein conformations, whereas atSPL1cyt had a more elongated structure, when viewed from a particular angle. Inspection of a multiple sequence alignment of the RNF domains revealed at least two amino acid residues in the atSPL1cyt RING domain that have notably different properties to those present in atSP1cyt and atSPL2cyt (Figure 4.2). Mutation of these two residues to match those present in atSP1 altered the structural model for atSPL1cyt to a more globular shape, similar to that for atSP1cyt and atSPL2cyt. It was therefore hypothesized that these two residues are important in controlling functional differences between SPL1 and SP1. However, when full-length SPL1 overexpression constructs harbouring these two mutations were introduced into the *ppi1* background, the transgenic plants did not show enhancement of the *ppi1* pale phenotype, as is observed upon overexpression of SP1 in *ppi1* (Figure 4.3, Ling *et al.*, 2012).

The SP1 E3 ligase has been demonstrated to have autoubiquitination activity *in vitro*, with the disruption in the conserved C3HC4 RING domain rendering it non-functional (Ling *et al.*, 2012). The autoubiquitination of SPL2 (under a different name, NERF) was tested in a different study (Gao *et al.*, 2015), using full-length SPL2 and RING-mutant versions of it; SPL2 showed robust autoubiquitination, whereas the RING mutant lacked the activity. However, SPL1 had not previously been tested in this way for autoubiquitination activity *in vitro*.

In accordance with the published results, my *in vitro* analysis showed that atSPL2cyt has intrinsic autoubiquitination activity *in vitro* that is similar to that of atSP1cyt, with RING mutants in both cases lacking activity (Figure 4.4). Notwithstanding how closely related to SP1 the SPL1 protein is, atSPL1cyt did not possess the ability to carry out autoubiquitination, as shown by the absence of polyubiquitinated smears in the immunoblot results. One possible interpretation of this is that atSPL1 is a pseudogene, as it does share homology and a similar gene structure with its paralogue, atSP1 (Ling *et al.*, 2012; Tutar, 2012). However, the atSPL1 protein is predicted to be only 5 amino acids shorter than atSP1 (arguing against an accumulation of mutations owing to a lack of selection pressure, which one might expect of a pseudogene), and public microarray data from Genevestigator indicate that atSPL1 expression, while lower than that of atSP1, is quite robust (Figure 4.7).

On investigating the process of leaf senescence, it was observed that the *spl1* mutants display an unanticipated phenotype. The *spl1* mutants exhibited accelerated senescence, similar to *SP1-OX* plants (Figure 4.6). This observation was consistent across all mutant alleles of *spl1* (Figure 4.6). According to the gene expression data generated by Genevestigator (Figure 4.7), the level of *SP1* expression increases

during plant senescence, while at the same time the expression of *SPL1* decreases. This trend fits well with the observation of accelerated senescence seen in plants overexpressing *SP1* or lacking *SPL1*. This discovery suggests that *SPL1* may be a functional pseudogene which may interfere with the parental gene function (that of *SP1*), a phenomenon which has been reported in several studies in cancer biology (Poliseno, 2012; Xiao-Jie *et al.*, 2015). Thus, an experiment was carried out to test whether the presence of SPL1 interferes with the ubiquitination activity of SP1. The *in vitro* results suggested that this is indeed the case (Figure 4.8), especially as the presence of control atSP1cyt-C330A mutant protein did not affect the ubiquitination activity of SP1. Extending this analysis to a living system, *ppi1* plants were engineered to overexpress *SP1* or *SPL1* or both, to test whether SPL1 expression interferes with SP1 activity *in vivo* (Figure 4.9). Interestingly, the *SPL1-OX ppi1* plants exhibited a slightly suppressed phenotype, revealed an effect of *SPL1-OX* that is the opposite of that of *SP1-OX*. Most importantly, the double overexpressor *SP1-OX SPL1-OX ppi1* plants exhibited a similar phenotype to *SPL1-OX ppi1* plants, indicating that SPL1 had exerted a dominant inhibitory effect on SP1.

A similar dominant effect was observed in the *sp1 spl1* double knockout mutants during senescence (Figure 4.6): the delayed onset of senescence phenotype of *sp1* was abrogated by the *spl1* mutation. However, mechanism in this case is unclear, as *sp1* is a null mutant and so there would be no SP1 activity to be modified by the presence/absence of SPL1. In this case, it was proposed that the absence of SPL1 might have affected SPL2 function, as manipulation of *SPL2* gene expression resulted in a similar phenotypes to that of *SP1*; i.e., knockdown of *SPL2* delayed the transition process, whereas overexpression accelerates it (Figure 4.6). With this hypothesis in mind, it would be very informative to assess the effect of

SPL1 on SPL2 *in vitro* ubiquitination activity, in further assays similar to those described in Figure 4.8.

This interesting discovery of *SPL1* as a putative functional pseudogenic protein and negative regulator of the SP1 E3 ligase sets an exciting new paradigm in the area of regulatory mechanisms governing chloroplast biogenesis, protein import, and SP1 function. This regulatory system might act as a complement to the known system of SP1 autoinhibition by self-ubiquitination (Ling *et al.*, 2012). Although the exact mechanism of SP1 regulation by SPL1 is not known at this stage, the data are consistent with SPL1 acting as a direct competitive inhibitor for binding to the E2 conjugase. This model is suggested by the results presented in Figures 4.8 and 4.9, and by information on gene expression during senescence (Figure 4.7). Moreover, the findings on *SPL2* presented in this chapter strengthen the proposal that its role is as a second regulator of the chloroplast protein import machinery.

4.4 Figures

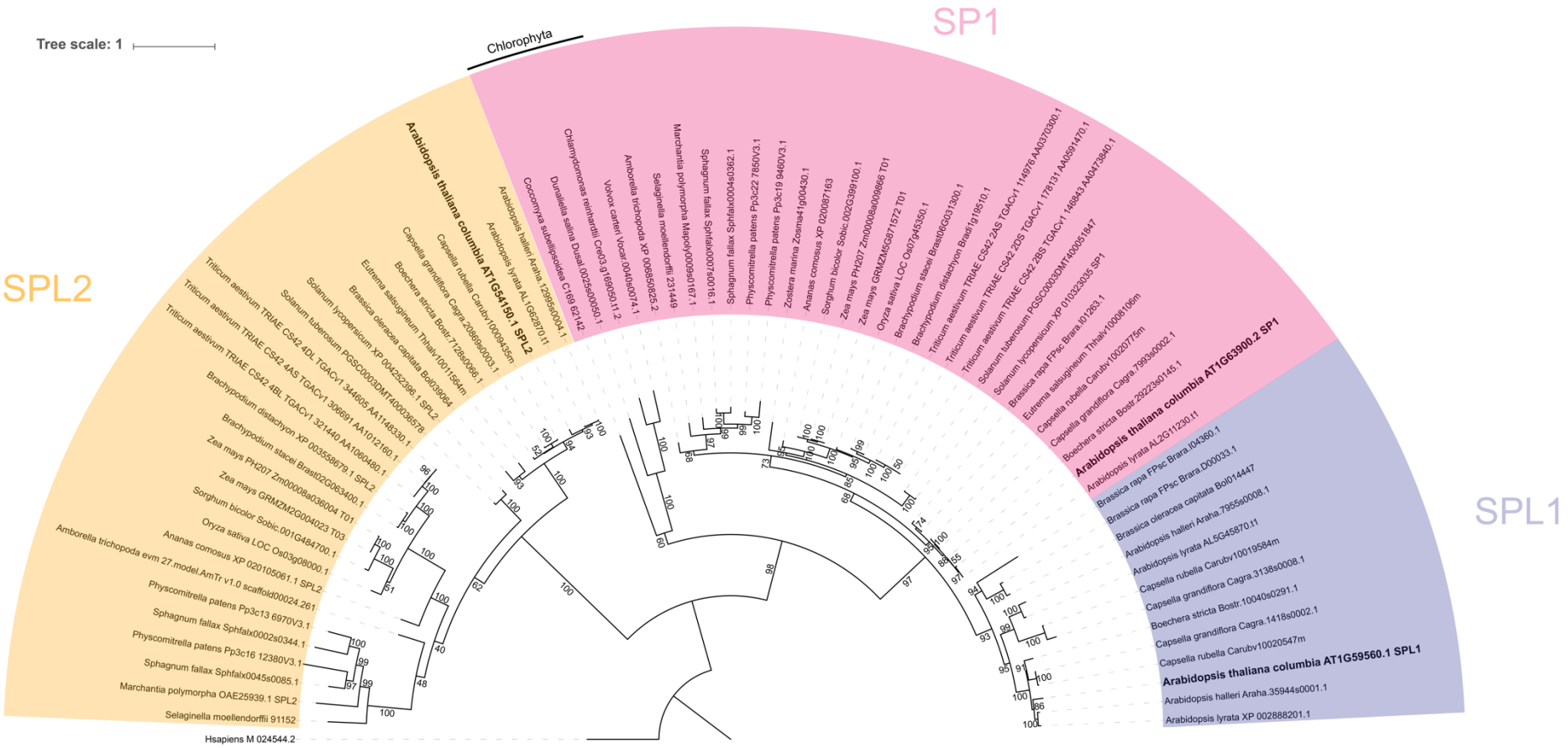


Figure 4.1 A bootstrapped maximum likelihood phylogenetic tree of SP1, SPL1 and SPL2 E3 ligases in Viridiplantae. Amino acid sequences of SP1, SPL1 and SPL2 orthologues were obtained by performing BLAST searches at Phytozome, Ensembl Plants and

NCBI. Alignment of the sequences were performed using MergeAlign (Katoh *et al.*, 2005; Collingridge & Kelly, 2012). The tree was inferred using W-IQ-TREE (Trifinopoulos *et al.*, 2016) with 1000 ultrafast bootstrap replicates and visualized in iTOL (https://itol.embl.de/). Scale bar indicates average number of substitutions per site. Percentage ultrafast bootstrap values are indicated at internal nodes. The species belonging to SP1 are coloured in pink, SPL1 in violet and SPL2 in orange. *Arabidopsis thaliana* E3 ligases are highlighted in bold text.

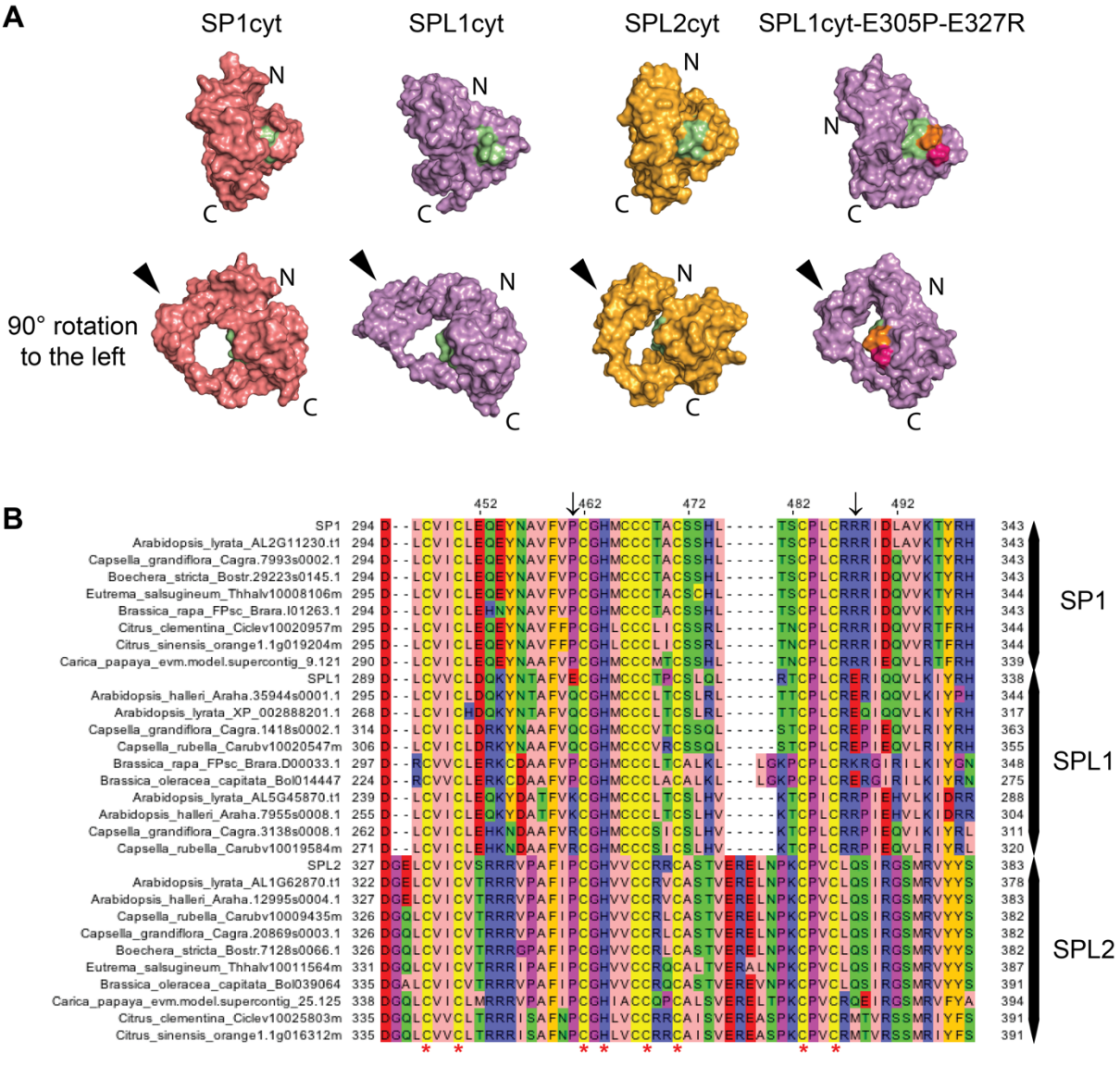


Figure 4.2 Structural and sequence analyses of the SP1/SPL cytosolic domains. (A) Predicted structures of *A. thaliana* SP1/SPL cytosolic domains. Structures were predicted by submitting the relevant amino acid sequences to the web server I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The 3D models were visualized using PyMol (https://pymol.org/2/). The green colour indicates the conserved C3HC4 RNF domain. The magenta and orange colours in SPL1cyt-E305P-E327R indicate the E305P and E327R mutated sites (see running text for details). Black arrow indicates the arc structure, N indicates the region where the cytosolic domain is attached to the transmembrane domain and C indicates the C-

terminal end. (B) Multiple sequence alignment of the SP1, SPL1 and SPL2 C3HC4 RNF domains from selected plant species (indicated to the left). The alignment was constructed using MergeAlign, and visualized using Jalview (www.jalview.org). *A. thaliana* sequences are labelled as SP1, SPL1 and SPL2. The critical conserved cysteine and histidine residues of the C3HC4 motif are indicated with asterisks. The arrow at the top indicates the amino acid of interests for atSPL1cyt-E305P-E327R studies.

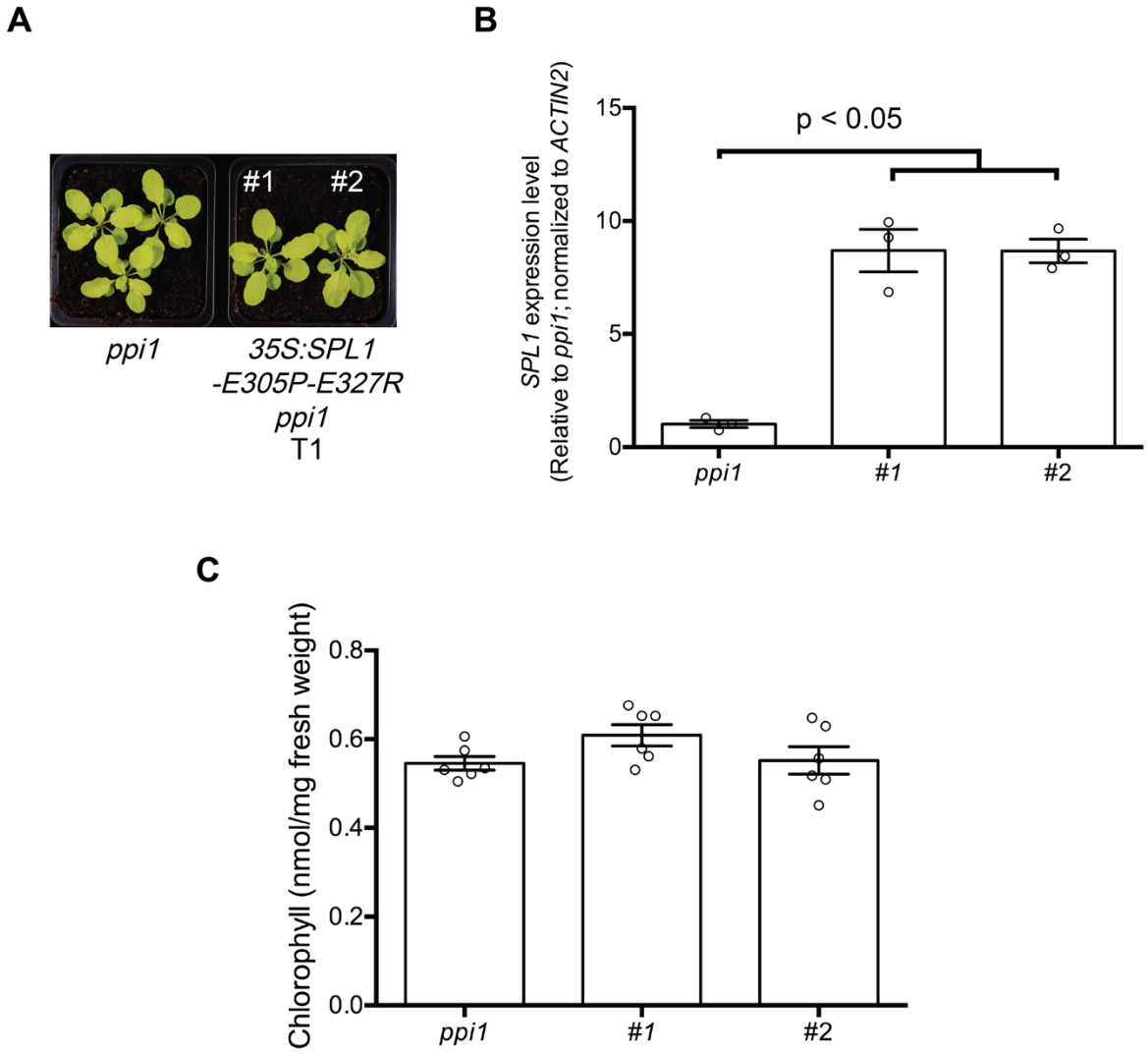


Figure 4.3 *In vivo* analysis of *SPL1* with mutations at two residues in the RING domain. (A) Phenotypic analysis of *ppi1* and *35S:SPL1-E305P-E327R ppi1* plants. Transformed T_1 generation plants were selected by growing for two weeks on MS media supplemented with phosphinothricin and then transferred to soil and allowed to grow under standard growth conditions for an additional two weeks. Representative plants were photographed. (B) Analysis of *SPL1* mRNA expression in the *35S:SPL1-E305P-E327R ppi1* plants. Total RNA samples of the 28-day-old plants shown in A were analysed by qRT-PCR using gene-specific primers for *SPL1* and the reference gene *ACTIN2*. Values shown are means (n=3) with open circles indicating individual data points and error bars indicating SEM. The p values were

derived from a Student's *t*-test of the means of the indicated genotypes. (C) Chlorophyll content analysis of the *35S:SPL1-E305P-E327R ppi1* plants. The 28-day-old plants shown in A were analysed using a Konica-Minolta SPAD-502 meter. Meter values were converted to chlorophyll amounts per unit fresh weight using an established equation. Values shown are means (n=6) with open circles indicating individual data points and error bars indicating SEM.

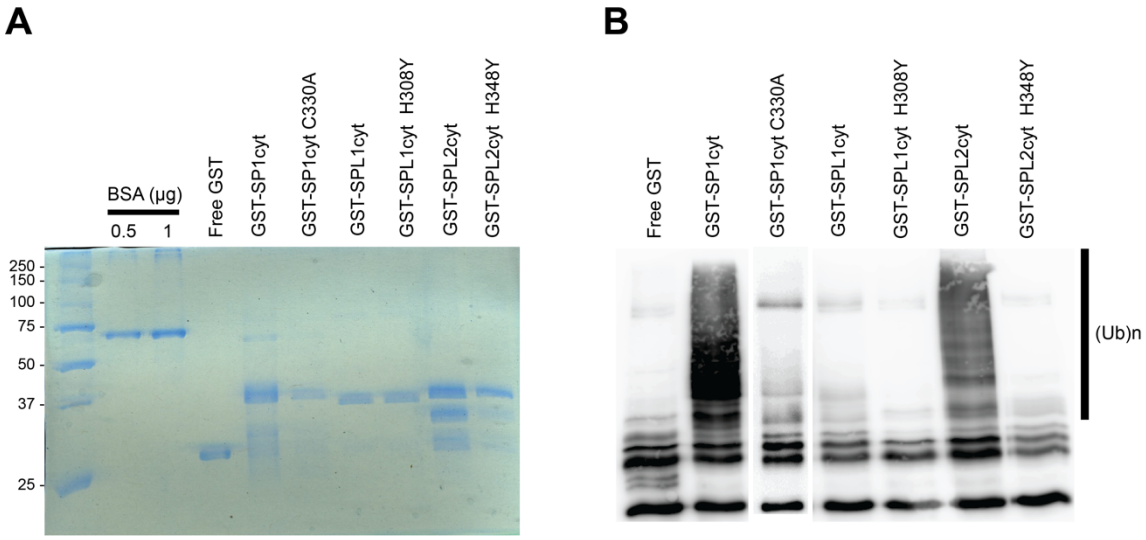


Figure 4.4. *In vitro* analysis of autoubiquitination activity of the SP1/SPL proteins and their RING mutant variants. (A) Bacterially expressed and purified Free GST and GST fusions of the indicated SP1/SPL cytosolic domains were run on SDS-PAGE to asses purity and concentration. The concentrations of the purified proteins were estimated by comparison with known amounts of bovine serum albumin standards. (B) Analysis on the ubiquitination activities of the purified proteins *in vitro*. A fixed amount of each proteins (1 μg) from A was incubated in a standard reaction containing all other components required for ubiquitination (E1, E2 and ubiquitin), and then the reaction products were resolved by immunoblotting using anti-HA. Successful autoubiquitination is indicated by the presence of high molecular weight bands of varying size, indicated by the bar labelled (Ub)n.

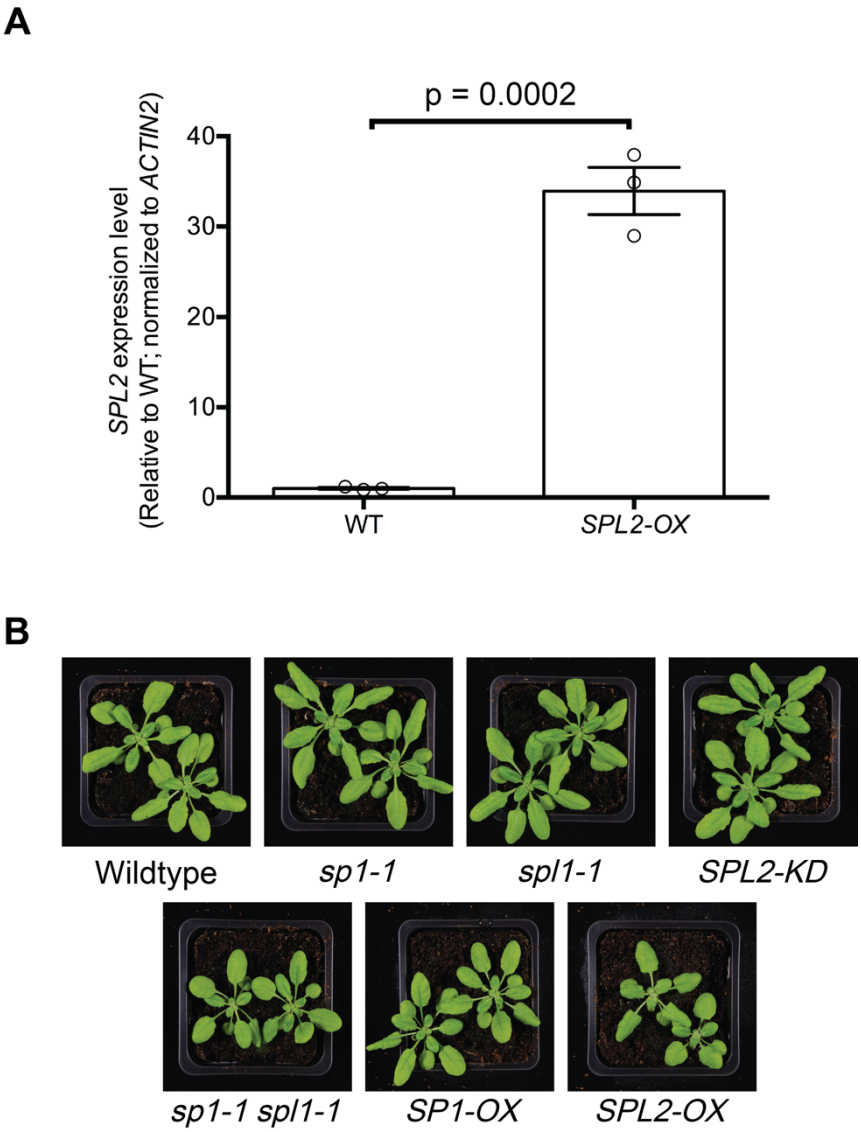


Figure 4.5 Establishment and phenotypic analysis of the SP1/SPL mutant and overexpressor lines to be used in the leaf senescence studies. (A) Plants of the indicated genotypes were grown on soil under standard growth condition for 21 days before photography. Representative plants are shown. No phenotypic differences between the genotypes were apparent under these conditions. (B) Analysis of *SPL2* mRNA expression in wild-type and *SPL2-OX* plants. Total RNA samples isolated from 21 day-old plants similar to those shown in A were analysed by qRT-PCR using gene-specific primers for *SPL1* and the reference gene *ACTIN2*. The analysis confirmed the continued overexpression of *SPL2* in the *SPL2-OX* line, which was

generated by outcrossing from the *sp1-3 ppi1* mutant background (see running text). Values shown are means (n=3) with open circles indicating individual data points and error bars indicating SEM. The p value was derived from a Student's *t*-test of the means of the indicated genotypes.

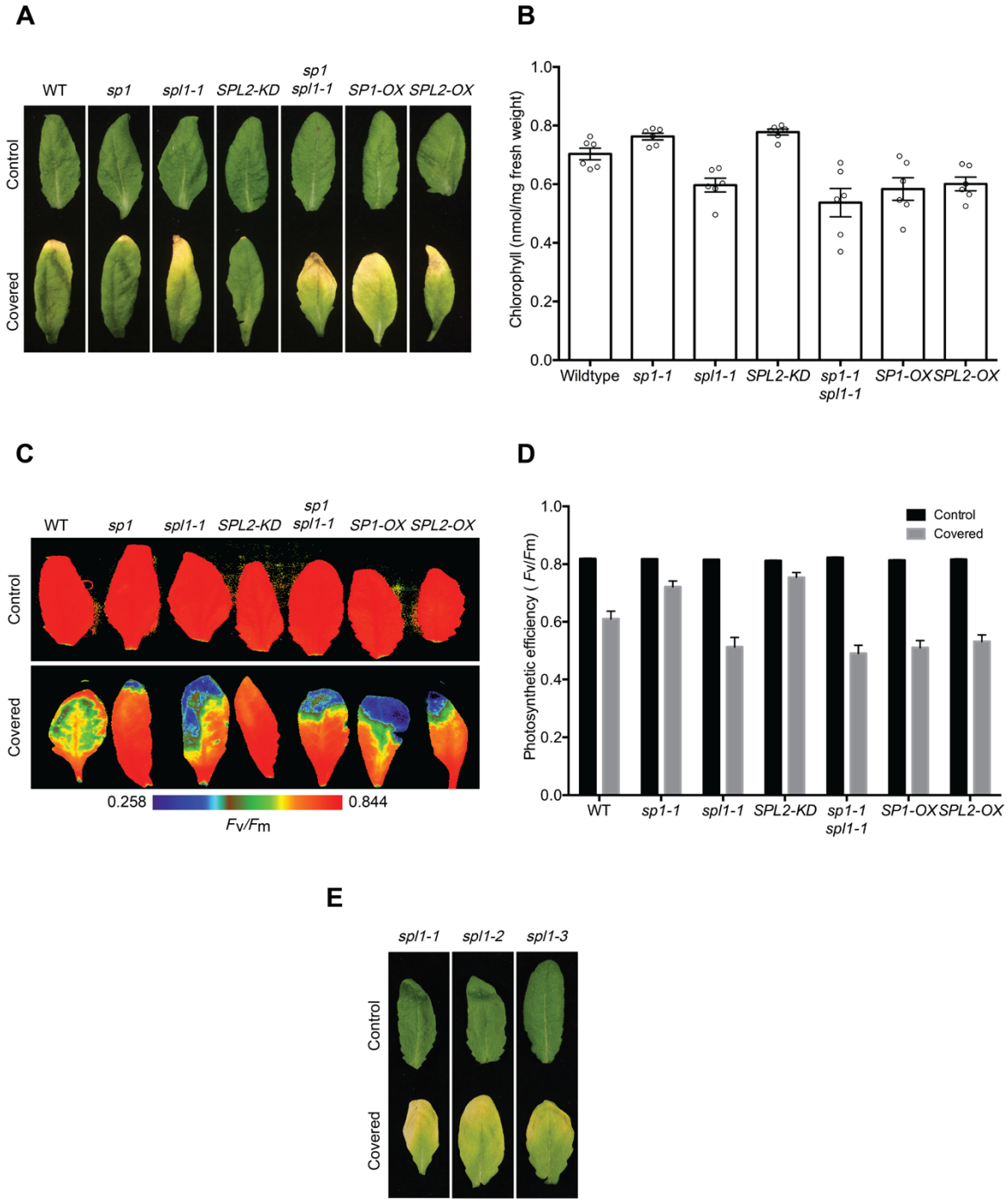


Figure 4.6 Analysis of leaf senescence in the different SP1/SPL mutant and overexpressor plant lines. (A) Visual analysis of the extent of senescence. Selected, equivalent rosette leaves of 28-day-old plants grown under standard conditions, for each genotype, were induced to senesce by covering them with aluminium foil (Covered); control leaves were left uncovered. Leaves that were representative of

the 6 analysed per genotype per condition were excised at the end of the treatment and photographed. (B) The chlorophyll contents of the senescent leaves described in A were measured using a Konica-Minolta SPAD-502 meter, as an indicator the extent of senescence. Meter values were converted to chlorophyll amounts per unit fresh weight using an established equation. Values shown are means (n=6) with open circles indicating individual data points and error bars indicating SEM. Student's *t*-test was used to test differences between the wild type and the mutants; the means were statistically different at $p < 0.05$. (C and D) The photochemical efficiency (F_v/F_m) of photosystem II in the control and senescent leaves described in A was measured, as a further indicator the extent of senescence. Values shown are means (n=6) and error bars indicate SEM. Student's *t*-test was used to test differences between the wild type and the mutants. The means were statistically different at $p < 0.05$. (E) Visual analysis of the extent of senescence in different mutant alleles of *spl1*. Procedures were as described in A.

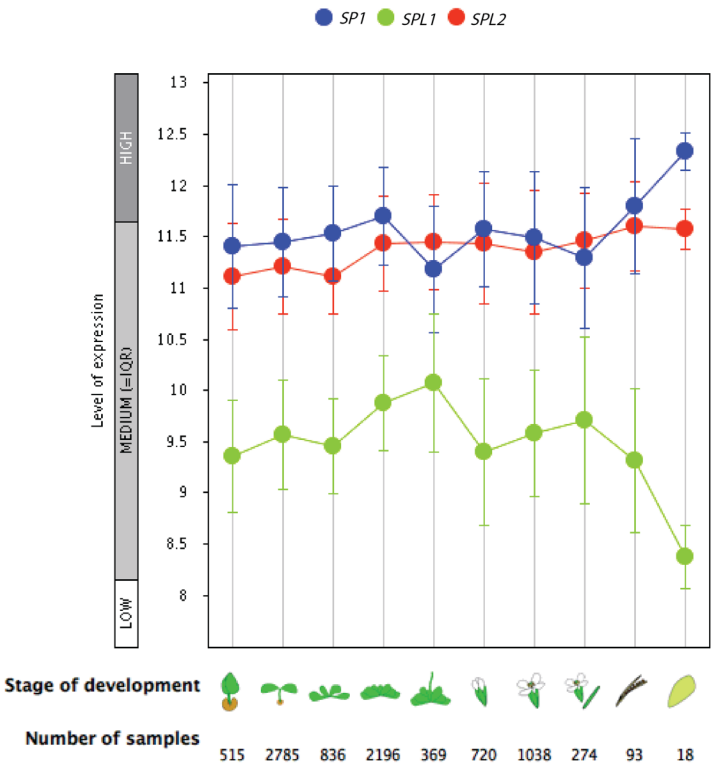


Figure 4.7 Expression of the *SP1*, *SPL1* and *SPL2* genes at different developmental stages. Gene expression data derived from microarray experiments were retrieved using the Genevestigator V3 analysis tool (https://genevestigator.com/gv/). Values shown are means and error bars indicate SEM; the total number of samples used to derive each data point is indicated. Stages of development from left to right are defined as follows: germinating seed, seedling, rosette, developed rosette, bolting, young flower, developed flower, flower and siliques, mature siliques and senescence. Typical ranges of low, medium and high expression are shown; medium is defined as the interquartile range (IQR).

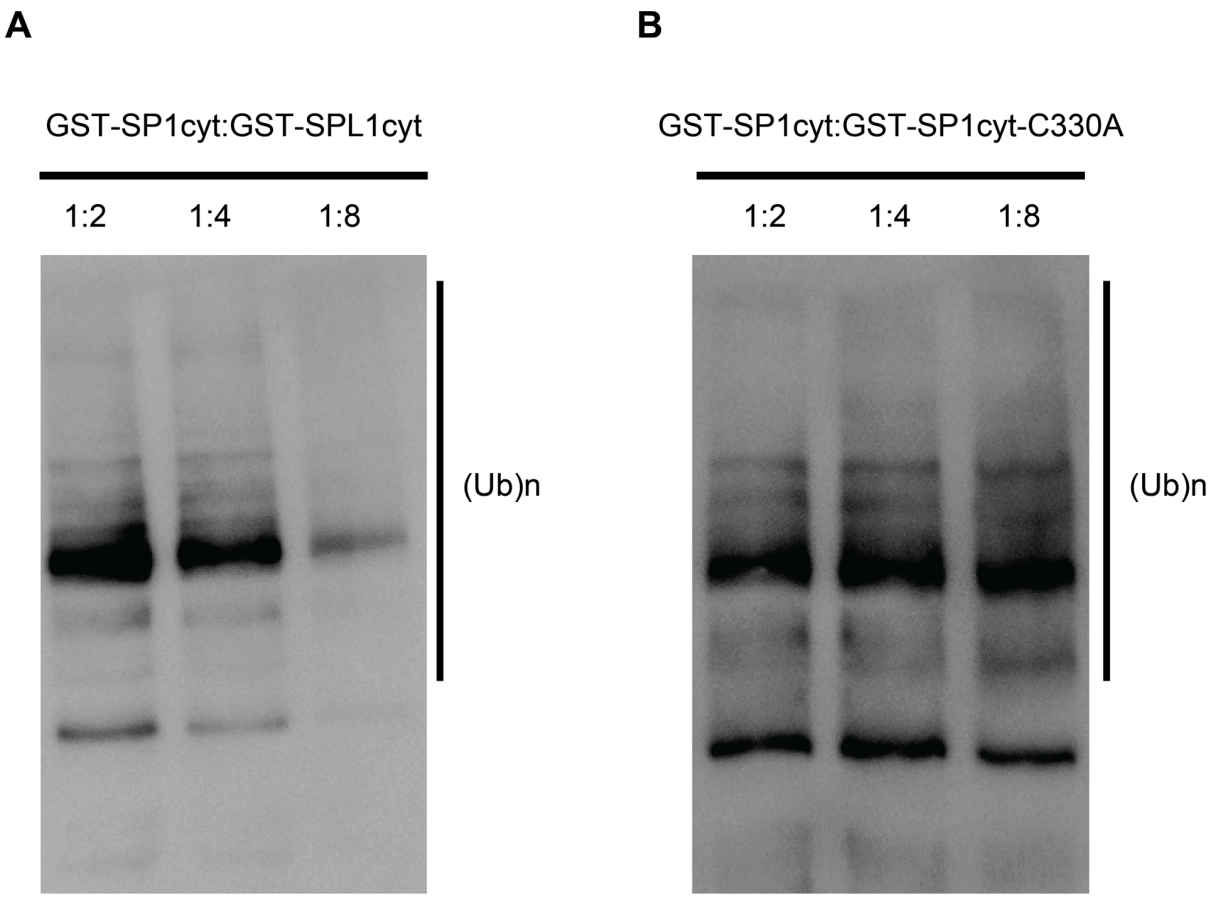


Figure 4.8 Analysis of the effect of SPL1 on SP1 autoubiquitination activity *in vitro*. (A) The indicated GST fusion proteins were bacterially expressed and purified as described in Figure 4.4. The GST-SPL1cyt protein was mixed in different ratios with 1 μg of GST-SP1cyt as indicated, and the mixtures were employed in ubiquitination assays; following incubation, the different reactions were resolved by SDS-PAGE and analysed by immunoblotting as described in Figure 4.4. Successful autoubiquitination is indicated by the presence of high molecular weight bands of varying size, indicated by the bar labelled (Ub)n. (B) A similar experiment was conducted using the SP1 RING mutant, GST-SP1cyt-C330A, instead of GST-SPL1cyt, as a control.

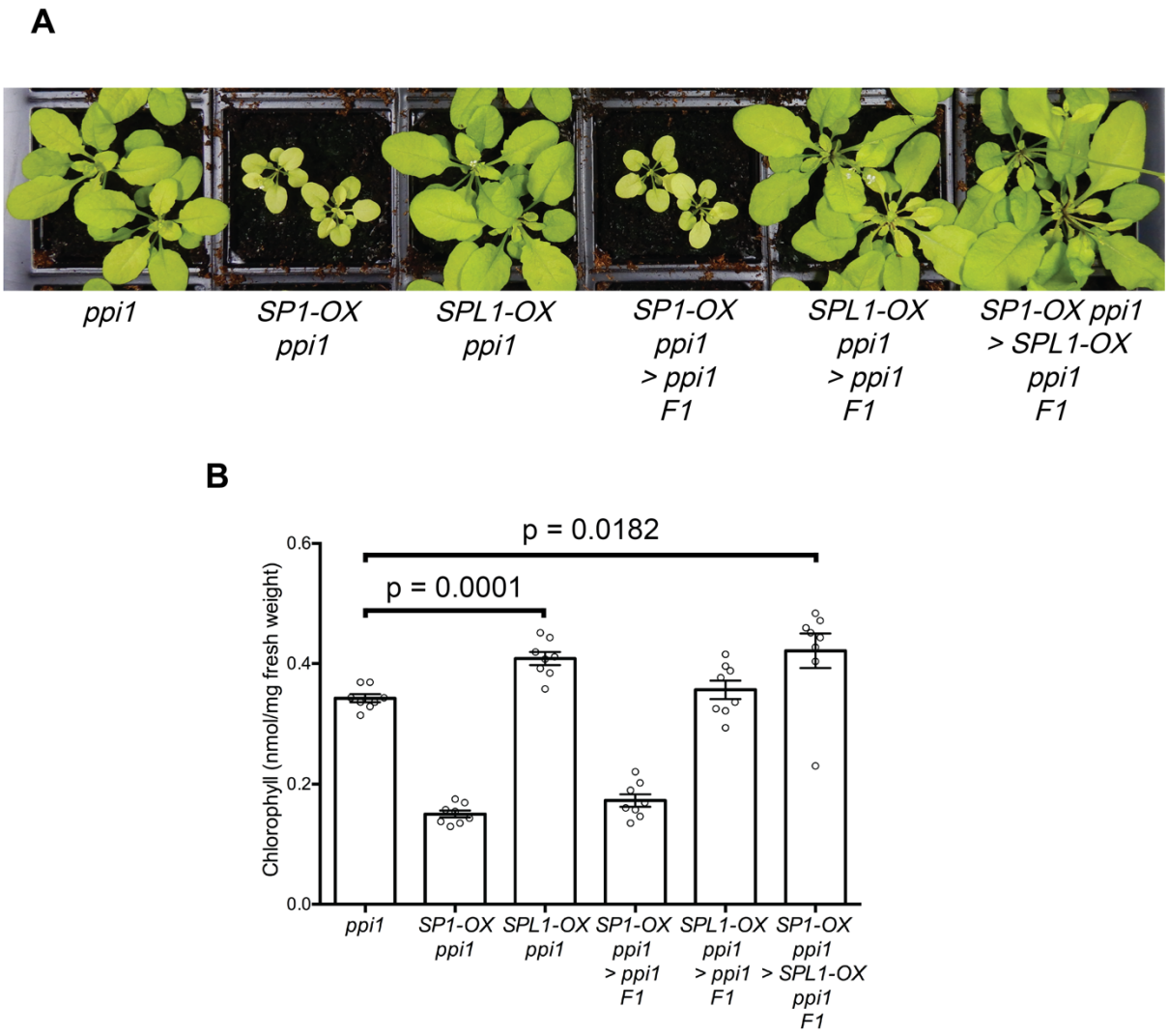


Figure 4.9 Analysis of the effects of *SPL1* overexpression on the phenotypes caused by the *ppi1* and *SP1-OX ppi1* genotypes. (A) Visible phenotype analysis of the plants. Plants of the indicated genotypes were grown on soil under standard

growth conditions for 28 days before photography; representative plants are shown. (B) Chlorophyll content analysis. Equivalent rosette leaves of 28-day-old plants similar to those shown in A were measured using a Konica-Minolta SPAD-502 meter. Meter values were converted to chlorophyll amounts per unit fresh weight using an established equation. Values shown are means (n=8) with open circles indicating individual data points and error bars indicating SEM. The p values were derived from a Student's *t*-test of the means of the indicated genotypes. (C) Visible phenotype analysis of young seedlings. Plants of the indicated genotypes were grown on MS medium under standard growth conditions for 5 days prior to photography.

Chapter 5

Results III

NEW SCREEN FOR *sp1* ENHANCERS

Abstract

The identification of the *sp1* gene was done through a forward genetic screen following EMS mutagenesis. In this chapter, I will describe a similar mutagenesis-based experimental set-up to search for enhancers of *sp1*. An estimated 20,000 *sp1 ppi1* seeds were used as starting material for the mutagenesis (M_1 seeds). Due to time constraints and availability of space to grow the plants, only half of the resulting M_2 seeds were screened. The screen identified three *sp1 enhanced chlorophyll* (*spec*) mutant lines that are relatively larger and greener than *sp1 ppi1*. Next-Generation Sequencing analysis of populations segregating for the *spec* mutations, coupled with computational analysis, was conducted to map the causative genetic lesions in the genomes of the mutant plants. The results yielded various candidates encoding proteins that are predicted to be chloroplast localized, as well as several encoding non-chloroplast-resident proteins. Two of the candidate genes have been reported in different studies to have an effect on the process of chloroplast protein import and/or chloroplast biogenesis, although the exact molecular mechanisms remain unclear. Even though the exact genes responsible for the *spec* mutant phenotypes are not known at this stage, this work has identified some interesting mutants, and corresponding candidate genes, for future analysis and characterization work.

CHAPTER 5:

NEW SCREEN FOR *sp1* ENHANCERS

5.1 Introduction

The discovery of *SP1* has led us to a tremendous amount of information on aspects of chloroplast biology, especially concerning the regulation of the TOC machinery during developmental processes (Ling *et al.*, 2012). A follow-up study revealed that *SP1* is also important for plant survivability during unfavourable growth conditions by limiting the import and activity of photosynthetic proteins and thus avoiding the generation of harmful reactive oxygen species (ROS) (Ling & Jarvis, 2015b). However, we are still lacking understanding of how the SP1 E3 ligase activity is regulated, other than that it involves autoregulation by self-ubiquitination (Ling *et al.*, 2012). We are also lacking information on which E2 conjugases are responsible for ubiquitin transfer onto the substrate in this system. It is noteworthy that there is evidence demonstrating the capability of E2 conjugases to direct the polyubiquitin chain to specific sites for activity regulation (David *et al.*, 2010; Stewart *et al.*, 2016). Moreover, it is very likely that other non-SP1-related components act on plastids, such as the components of other post-translational modification processes including protein kinases, SUMO (Small Ubiquitin-like Modifier) system proteins, deubiquitylating enzymes (DUBs), and others.

In order to further explore this area of plastid biology, a new forward genetic screen for enhancers of *sp1 ppi1* was conducted, in a similar manner to the screen that identified *sp1*. The aim was to search for new components in the SP1 pathway, and/or new factors that regulate plastid protein import. The genetic screen was

carried out following mutagenesis with ethyl methanesulfonate (EMS), an alkylating agent that randomly generates chemical modifications of nucleotides throughout the genome, causing base substitution from G/C to A/T (Maple & Møller, 2007). This mutagenesis technique enables searches for both loss- and gain-of-function mutants, and may also identify the functionally important residues in the protein encoded by the mutagenized gene. Another advantage to using this technique, as compared to other methods such as insertional mutagenesis using the *Agrobacterial* transfer DNA (T-DNA), is the non-occurrence of chromosomal rearrangements that may complicate the screening process.

5.2 Results

5.2.1 Forward genetic screen

An estimated 20,000 *sp1 ppi1* seeds were subjected to EMS mutagenesis. To aid screening, all of the EMS-mutagenized M_1 seeds were roughly divided into 100 groups, each containing approximately 200 seeds. The seeds were then grown in trays containing soil and placed in the greenhouse under standard growth conditions. In total, 50 trays were used, and these were each split into two half-trays, giving two groups labelled as A and B; i.e., there are 50 half-trays in A, and 50 half-trays in B. The M_2 seeds were harvested by pooling the plants in each half-tray together, giving a total of 100 pools.

The screening of the M_2 generation was done by germinating the seeds *in vitro* on MS medium. This method was done to aid new mutant phenotype identification. In my screen, plants that appeared greener than *sp1 ppi1* (in other words, those with mutations that enhanced the *sp1* phenotype) were selected and propagated on the soil to grow into maturity. In total, 226 putative mutants of interest

were identified, from the screening of Group A pools only. However, a large proportion of the putative mutant lines were lost due to sterility. The Group B M2 pools were not screened due to time and space constraints, and due to the fact that a sufficient number of mutants was identified from screening the Group A pools.

The aim of this mutant screen was to identify enhancer(s) of *sp1* in the *ppi1* mutant background, i.e., mutants of similar or large size with a clearly greener phenotype than *sp1 ppi1*. Of the 226 putative mutants originally identified, 9 failed to reach maturity and set seeds, and a further 37 did not display a heritable enhancer phenotype upon re-screening in the M3 generation. Of the remainder, 123 mutants were excluded on the basis that they expressed developmental defects, such dwarfism that can indirectly lead to a greener phenotype (Peng *et al.*, 1997), while 54 were not prioritized owing to the weakness of their enhancer phenotypes. In end, the screen of Group A yielded three putative mutants of interest: 1A-1, 37A-5 and 40A-1 (Figure 5.1 A). All three mutants exhibited elevated chlorophyll levels that were intermediate between those of *sp1 ppi1* and wild-type plants (Figure 5.1 B). These three triple mutants were renamed as *spec* (*sp1 enhanced chlorophyll*) mutants; specifically, *spec1*, *spec2* and *spec3*, respectively (Figure 5.1 A and B).

5.2.2 Generation of the *spec* mutant mapping populations and genomic datasets

Mapping of the *spec* mutant gene candidates in the *sp1 ppi1* background was done by Next Generation Sequencing (NGS) combined with computational analysis using the software package SHOREmap v3.6 (Schneeberger *et al.*, 2009; Sun & Schneeberger, 2015). This mapping-by-sequencing approach requires sequence information from an isogenic mapping population as well as from the non-

mutagenized progenitor (line ID: sp243 BC 3 5). The isogenic mapping population for each *spec* mutant was generated by backcrossing (BC) the corresponding triple mutant to *sp1 ppi1* (sp243 BC 3 5). The F_1 generations were allowed to self-pollinate to generate segregating F_2 populations. In the F_2 generations, the plants clearly exhibiting the *spec* mutant phenotype were selected and pooled (a minimum of 50 samples per mutant were selected (James *et al.*, 2013)) and used for DNA extraction. The mapping population sizes were as follows: 100 individuals for *spec1 sp1 ppi1*, and 50 individuals each for *spec2 sp1 ppi1* and *spec3 sp1 ppi1*.

The DNA extracted from the three pooled mapping populations, and from the non-mutagenized progenitor, were subjected library preparation and sequencing on a single lane of an Illumina HiSeq 4000 at the Wellcome Centre for Human Genomics, Oxford. Paired-end reads of 150 bases to a sequencing depth of 160x were generated.

5.2.3 Computational analysis of the sequence data to identify candidate *spec* loci

SHOREmapping analysis was performed on the Binary Alignment Map (BAM) files of the pooled individuals and parental line. The reference sequence of *Arabidopsis thaliana* (TAIR10_chr_all.fas) and gene annotation (TAIR10_GFF3_genes.gff) from The Arabidopsis Information Resource (TAIR; www.arabidopsis.org) were also used for the analysis. The aim of this computational analysis was to identify loci wherein single-nucleotide polymorphisms (SNPs) between the mapping populations and the progenitor were heavily skewed (>90%) away from the progenitor genotype, implying linkage to the *spec* mutant phenotype.

5.2.3.1 Candidate loci for *spec1*

For *spec1 sp1 ppi1*, a cluster of SNPs showing very strong linkage (>99%) to the enhancer phenotype was observed on chromosome 2, at a genomic position between 1.6 Mb and 2.2 Mb (Figure 5.2 A). It is highly likely that the lesion responsible for the *spec1* phenotype lies in this region. A closer look at this region revealed three genes that are directly affected by the mutations, with the remainder of the SNPs lying in intronic and intergenic regions (Figure 5.2 B and Table 5.1). The mutations in two affected genes (At2g05140 and At2g05642) showed an allele frequency of 1, while those in a third gene (At2g05170) showed an allele frequency of 0.99. The mutations at At2g05140 and At2g05170 each cause an amino acid substitution from leucine (L) to phenylalanine (F), residues which have similar hydrophobic side chain properties; this suggested that these two genes are unlikely candidates. The third gene, At2g05642, is most likely to be the causal, *spec1* locus because the original glutamic acid (E) residue was substituted to a lysine (K) residue by the corresponding mutation; this is a switch from a negatively-charged amino acid to positively-charged amino acid, and is more likely to be linked to change in protein function.

5.2.3.2 Candidate loci for *spec2*

The causal mutation for *spec2* is also likely to be on chromosome 2; the mutations on chromosome 1, 3 and 5 showing SNPs with frequencies over 75% resided in intronic and intergenic regions. In this case, the detected linkage was much weaker, possibly owing to difficulties in phenotypically identifying homozygous mutants in the mapping population. A number of SNPs showing linkage (>75%) to the enhancer phenotype was observed at genomic positions between 4 Mb and 16

Mb (Figure 5.3 A). Again, several of these mutations resided in intronic and intergenic regions, but in three cases the mutations affected the coding region of a candidate gene, and with similar allele frequency values (Figure 5.3 B and Table 5.2). These three genes are At2g20010, At2g36800 and At2g37550. The At2g20010 mutation causes a conservative L-to-F substitution, while the At2g36800 mutation causes the replacement of a polar proline (P) residue by a non-polar L residue. However, the third gene, At2g37550, may be the most likely candidate as the mutation in this case results in a large change in amino acid properties, from a positively-charged arginine (R) residue to cysteine (C).

5.2.3.3 Candidate loci for *spec3*

For *spec3 sp1 ppi1*, a cluster of SNPs showing linkage (>75%) to the enhancer phenotype was observed on chromosome 3, at genomic positions between 0 Mb and 15 Mb (Figure 5.4 A). Again, several of these mutations resided in intronic and intergenic regions, and so could be excluded as likely candidates. In the region of interest, there were a total of six candidate genes each bearing a non-synonymous mutation with an allele frequency greater than 0.75 (Figure 5.4 B and Table 5.3). One of the candidate genes, At3g19570, has a nonsense mutation that is predicted to result in premature termination of translation in the exon 1. This is the highest likely candidate as the termination occurs in the 1st exon, at a site very near to the translation initiation site. The other gene candidates has missense mutations that does not result in a significant change of amino acid properties, and are therefore least likely to cause the enhancer phenotype.

The detailed properties of the candidate *SPEC* genes as provided by the TAIR website are summarized in Table 5.4.

5.3 Discussion

The introduction and analysis of genetic variation by mutagenesis and forward genetic screening is a powerful tool for identifying genes involved in biological processes (Woodson *et al.*, 2015; Bédard *et al.*, 2017). In this chapter, I described the identification of three new enhancers of *sp1* following EMS mutagenesis; all of these mutants are phenotypically greener than their progenitor, the *sp1 ppi1* double mutant (Figure 5.1). These enhancer mutant lines were named *spec*, and numbered according to their order of identification.

While forward genetics is undoubtedly a very powerful approach, the process of identifying the causal mutation in the genome of each new mutant was in the past limited by the complex, time-consuming and laborious method of traditional map-based cloning (Jander *et al.*, 2002). The rapid advancement of whole-genome sequencing technologies and their relatively low cost has allowed us to dramatically speed up the identification process, by applying computational analysis to sequence data derived from pooled mapping populations (Schneeberger, 2014). In this chapter, I employed the software package SHOREmap (http://1001genomes.org/software/shoremap.html) to rapidly identify candidate causal mutations for the three newly-identified *sp1* enhancer mutants, through analysis of deep sequencing data from a large pool of recombinants in each case (Schneeberger *et al.*, 2009; Sun & Schneeberger, 2015).

The mapping-by-sequencing analysis for *spec1* narrowed down the analysis to three candidate genes (Figure 5.2). The products of two of these genes are

predicted by TAIR to be components of the chloroplast (Figure 5.2, Table 5.1 and Table 5.4). According to TAIR, At2g05140 encodes for a "phosphoribosylaminoimidazole carboxylase family protein / AIR carboxylase family protein", which is an enzyme involved in a catalytic process during the late stages of purine biosynthesis (Patey & Shaw, 1973; Rowe *et al.*, 1978). A related mutant affecting the purine biosynthesis pathway, *cia1* (*chloroplast import apparatus 1*), was isolated in a different study, demonstrating the negative effect of lesions in this metabolic process on chloroplast protein import and biogenesis (Hung, 2004). The *CIA1* gene encodes glutamine phosphoribosyl pyrophosphate amidotransferase 2 (ATase2), one of the enzymes responsible for the early stages of purine biosynthesis. The process of protein import into the chloroplasts requires the purine nucleotides, ATP and GTP (Jarvis, 2008). However, supplementation of import reactions with ATP and GTP only slightly alleviated the import defect of the *cia1* mutant; a possible explanation for this result is that there is an alteration in the amount of the ATP/ADP translocator of the inner envelope membrane in *cia1*, and that this influences ATP availability to the import machinery in the stroma. Given that the *cia1* mutation causes a negative effect on protein import, it seems unlikely that the *spec1* phenotype (which presumably is linked to improved import capacity) is caused by a lesion in the same pathway. Also, the amino acid residue change in this gene is from leucine to phenylalanine, which does not represent a significant change in side chain properties, except for the addition of an aromatic ring. A different class of mutant related to purine turnover, *altered chloroplast import2* (*aci2*), was also reported to have an effect on the chloroplast pre-protein import efficiency (Zhong *et al.*, 2010).

This leaves At2g05642 as the final candidate for the *spec1* locus, which is the other gene proposed to encode a chloroplast component. There is very little information available concerning this gene, but it is proposed to have nucleic acid-binding property due to the presence of oligonucleotide/oligosaccharide-binding (OB)-fold-like domain (Table 5.4, Flynn & Zou, 2010). It is worth noting that the amino acid substitution in this case is a significant change, from glutamic acid to lysine; i.e., a change from a negatively-charged residue to a positively-charged residue. Although OB-fold proteins are mostly reported as members of DNA damage response, they are also recognized to have a protein-protein binding activity for cell surface recognition in human pathogens or even forms a catalytic site in yeast, archaea and bacteria (Arcus, 2002; Ginalski *et al.*, 2004). Considering chloroplasts are descendants of an ancient prokaryotic organism, this OB-fold protein could have similar catalytic or protein recognition role in the organelle that might have an influence on the protein import activity.

The allele frequency values for the SHOREmap analysis on *spec2* and *spec3* were not as high as those for *spec1*, which made the selection of candidates a little difficult (Figure 5.3 and Figure 5.4). This might be due to the lower number of a individuals included in the respective mapping populations (50 plants each instead of 100 plants in the case of *spec1*). The possibility of having contaminants (individuals that are not actually homozygous for the mutation of interest) in the pooled DNA sample can massively reduce the reliability of the analysis, and this is especially problematic with smaller mapping populations. Weaker phenotypes that are difficult to reliably phenotype (such as *spec2* and *spec3*) are therefore more difficult to analyse. To overcome this issue, a cut-off allele frequency value of 0.75 with high-quality reads was chosen for the analysis.

The analysis of *spec2* yielded two candidate genes with predicted chloroplast localization (Figure 5.3, Table 5.2 and Table 5.4). One of these putative chloroplast proteins is At2g20010, which is an uncharacterized protein with a domain of unknown function (DUF). Although this might be an interesting candidate to study, the nature of amino acid mutation, from leucine to phenylalanine, is such that it is not predicted to have a big effect on protein function. The other putative chloroplast protein is At2g36800, which is a deoxynivalenol (DON)-glucosyltransferase, a component that is proposed to act in the homeostasis of brassinosteroid hormones (Table 5.4). Although the nature of mutation in this case (causing a proline-to-leucine substitution) suggests that this gene is the most likely candidate for *spec2*, it seems unlikely that such a lesion would have a direct effect on protein import. At the outset of this study, it was realized that hormone defects causing dwarfism have the potential to give a greener phenotype (Peng *et al.*, 1997), thereby giving the impression of suppression. However, this was kept in mind throughout the analysis, and none of the selected mutants (including *spec2*) displayed such developmental defects. The third, non-chloroplast-predicted candidate for *spec2* carries a lesion causing a very strong difference in amino acid properties (an arginine-to-cysteine substitution), although it is unclear at this stage of how disruption of this gene could affect chloroplast development. This protein has been reported to have protein trafficking activities in plant cells, in particular, the Golgi complex and the endoplasmic reticulum (Min *et al.*, 2007).

As for *spec3*, one of the candidate genes carries a nonsense mutation (Figure 5.4, Table 5.3, and Table 5.4). The relevant gene, At3g19570, encodes SCO3 (SNOWY COTYLEDON3), a peroxisome-localized protein that is apparently required for normal chloroplast biogenesis; mutant *sco3* plants have impaired chloroplast and

etioplast development in the cotyledons (Albrecht *et al.*, 2010). Even though the molecular mechanism of SCO3's involvement in chloroplast biogenesis is unclear, it was demonstrated that perturbation of the microtubule cytoskeleton phenocopies the effect of *sco3* on plastid morphology. However, given that *sco3* exerts a negative effect on plastid biogenesis, in a developmentally restricted way (i.e., in the cotyledons), it seems unlikely that a mutation at the same locus would give rise to improvements in greening throughout plant development as seen in *spec3*. In addition, the characteristic pale-cotyledon phenotype of *sco3* was not apparent in the *spec3* mutant. The other likely candidates (At3g17510, At3g21450 and At3g25840) in the list for *spec3* are protein kinases, which have the ability to regulate diverse cellular processes via posttranslational modification of protein substrates (Cheng *et al.*, 2011). A recent report identified a regulatory kinase at the outer chloroplast membrane 1 (KOC1) that stably associates with the TOC complex (Zufferey *et al.*, 2017). Mutation of this kinase results in impaired protein import efficiency *in vitro* and plants harbouring the *koc1* mutation has a high mortality rate during plastid transition process in de-etiolation. Therefore, having protein kinases as a candidate for *spec3* is very plausible.

Although the discussion in this chapter concerning the identity of the *spec* genes is necessarily speculative, owing to a lack of time to conduct the necessary confirmatory analysis (e.g., obtaining T-DNA mutants for each of the candidate genes, and then crossing them with *sp1 ppi1* in order to assess for recapitulation of the *spec* phenotypes), I was able to determine a short list of reasonable and interesting candidates in each case, through next-generation sequencing combined with computational analysis. The next step in this project will be to confirm which of the candidates in Table 5.4 are the genes that cause the greener phenotypes of the

spec mutants. As noted above, the best way to achieve this will be by using reverse-genetic methods, obtaining T-DNA mutants from the *Arabidopsis* stocks centres and testing whether they can reproduce the enhancer phenotypes. An alternative approach would be transgenic complementation analysis, which would involve introducing the non-mutated candidate gene sequences into the corresponding *spec* mutants and testing for abrogation of the enhancer phenotypes. It will also be important to perform more backcrosses, to clean up the background of each *spec* mutant by removing unrelated mutations; combined with PCR genotyping assays, this might also help to confirm the identity of the *spec* loci.

5.4 Figures

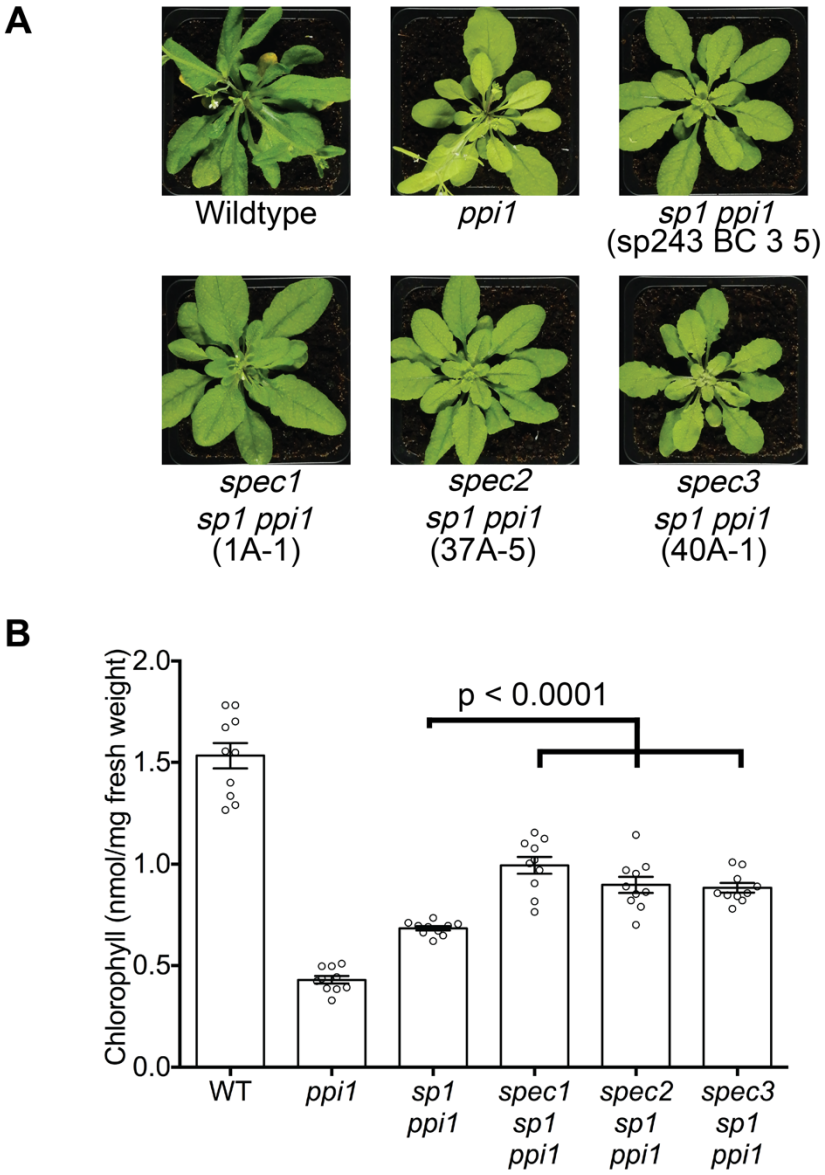


Figure 5.1 The selected *spec* mutant lines identified from EMS mutagenesis screen. (A) Phenotypic analysis of wild-type, *ppi1*, *sp1 ppi1* (sp243 BC 3 5) and homozygous M_4 *spec* mutant plants. The plants were grown under standard growth conditions on soil for 35 days before photography; representative individuals are shown. (B) Chlorophyll content was measured using a Konica-Minolta SPAD-502 meter. Meter values were converted to chlorophyll per fresh weight using an established equation. Values shown are means (n=10) with open circles indicating

individual data points and error bars indicating SEM. The p values were derived from a Student's *t*-test of the means of the indicated genotypes.

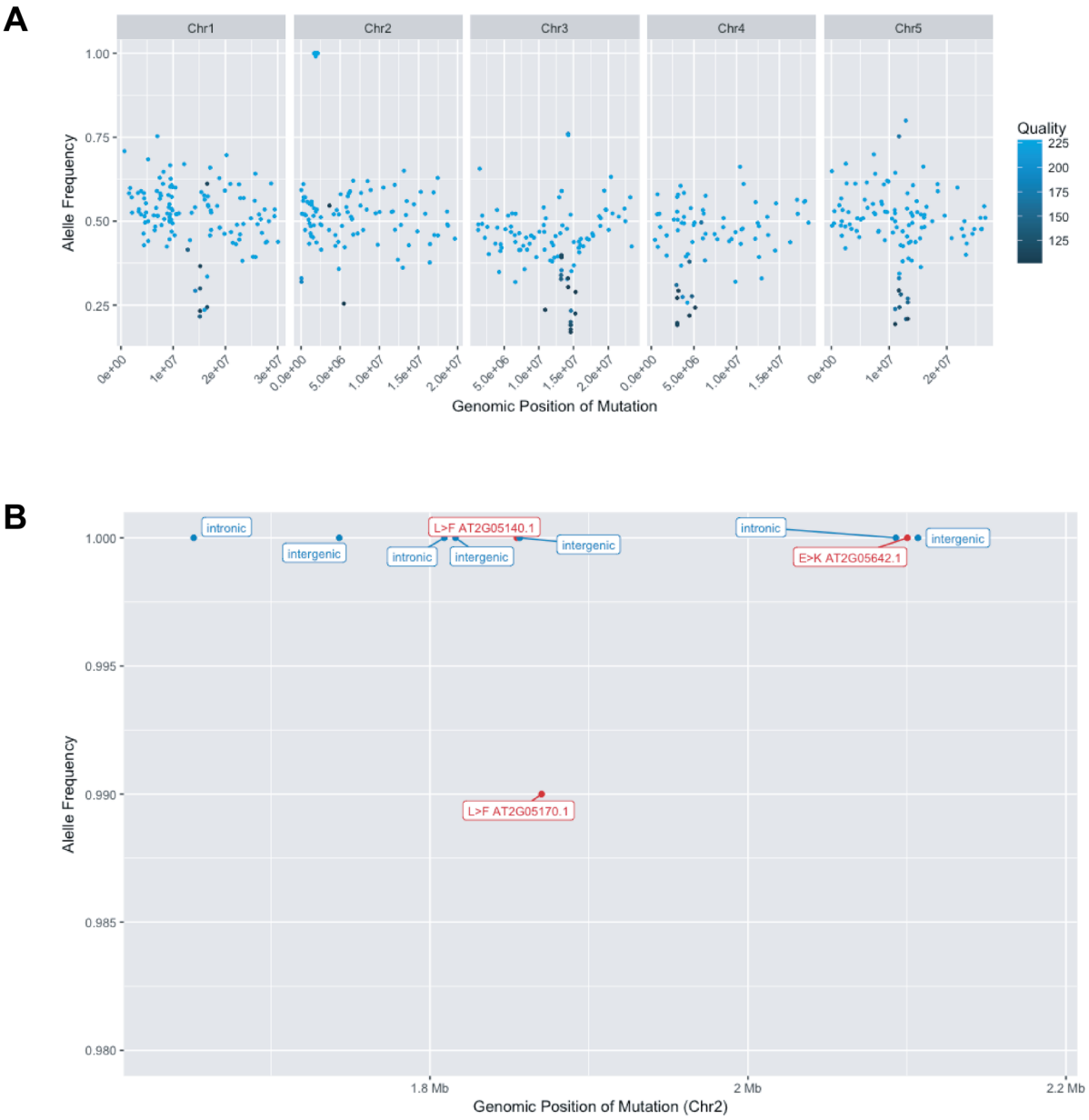


Figure 5.2 Mapping-by-sequencing analysis of *spec1 sp1 ppi1.* (A) Graphical representation using R (https://www.r-project.org/) of allele frequency estimations in a backcrossed mapping population. The points indicate individual mutation markers and are colour-coded in a gradient of blue (high quality) to black (low quality); this represents the quality of the relevant reads from the Illumina sequencing. (B) Graphical representation as above of the region between position 1.6 Mb and 2.2 Mb of chromosome 2, which is suggested to be the location of the causal mutation site.

The red labels indicate high priority nonsynonymous mutations, and the blue labels indicate low priority intergenic and intronic mutations.

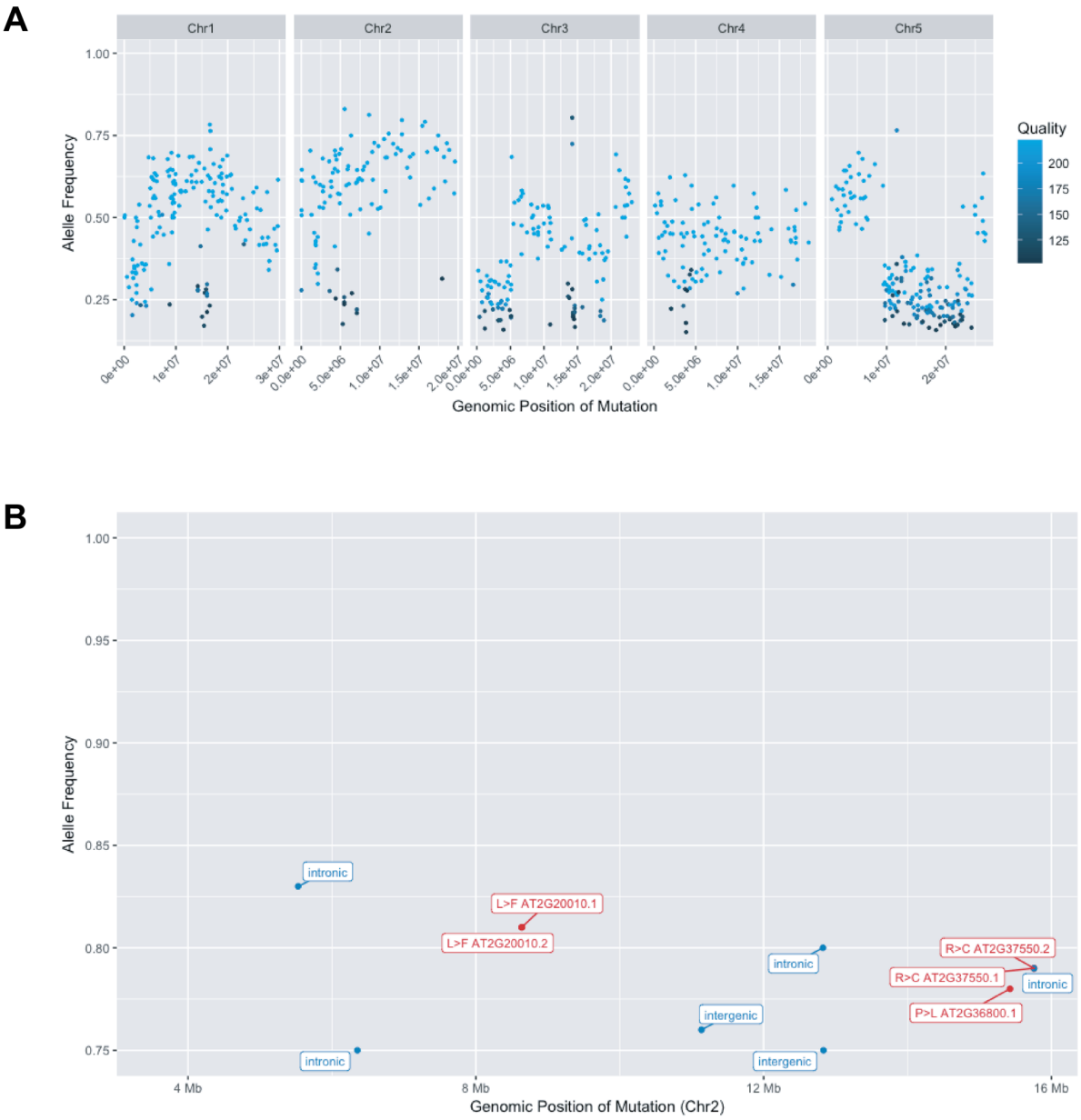


Figure 5.3 Mapping-by-sequencing analysis of *spec2 sp1 ppi1*. (A) Graphical representation using R (https://www.r-project.org) of allele frequency estimations in a backcrossed mapping population. The points indicate individual mutation markers and are colour coded in a gradient of blue (high quality) to black (low quality); this represents the quality of the relevant reads from the Illumina sequencing. (B) Graphical representation as above of the region between position 4 Mb and 16 Mb of chromosome 2, which is suggested to be the location of the causal mutation site.

The red labels indicate high priority nonsynonymous mutations, and the blue labels indicate low priority intergenic and intronic mutations. The two red labels at a point At2g20010 and At2g37550 indicate the presence of multiple gene models in TAIR.

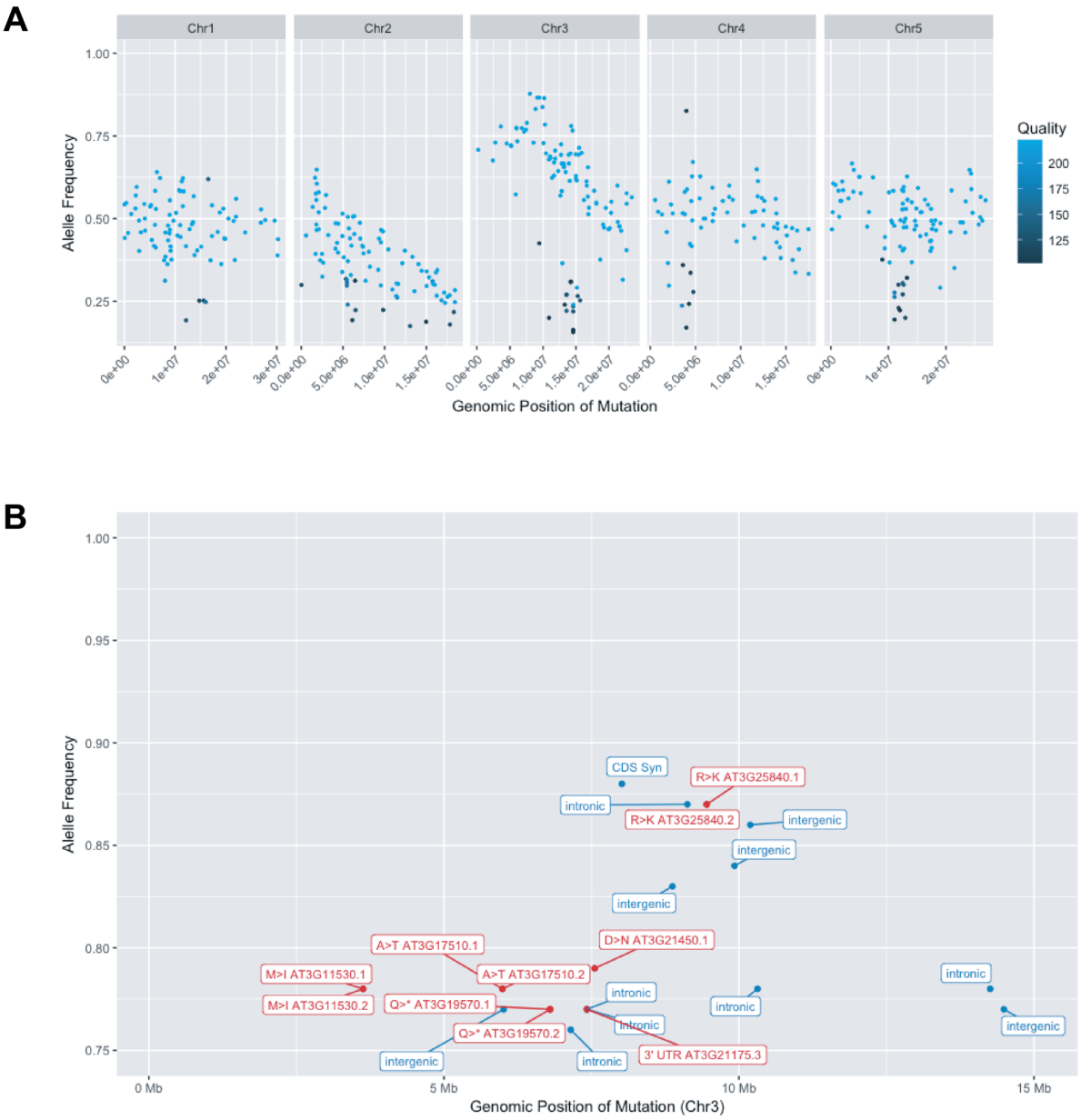


Figure 5.4 Mapping-by-sequencing analysis of *spec3 sp1 ppi1*. (A) Graphical representation using R (https://www.r-project.org) of allele frequency estimations in a backcross mapping population. The points indicate individual mutation markers and are colour coded in a gradient of blue (high quality) to black (low quality); this represents the quality of the relevant reads from Illumina sequencing. (B) Graphical representation as above of the region between position 0 Mb and 15 Mb of chromosome 2, which is suggested to be the location of the causal mutation site.

The red labels indicate high priority nonsynonymous mutations, and the blue labels indicate low priority intergenic, intronic and synonymous mutations. The two red labels at a point At3g11530, At3g17510 and At3g19570 indicate the presence of multiple gene models in TAIR.

5.5 Tables

Genomic Position (Chromosome 2)	Allele Frequency	Reference base	Mutant base	Read Quality	Nature of Mutation	Amino Acid Substitution	Genetic Locus
1651436	1.00	C	T	225	intronic	-	At2g04710
1742983	1.00	C	T	225	intergenic	-	-
1809006	1.00	C	T	225	intronic	-	At2g05080
1816117	1.00	C	T	225	intergenic	-	-
1854696	1.00	C	T	225	nonsynonymous	L > F	At2g05140
1856480	1.00	C	T	225	intergenic	-	-
1870230	0.99	C	T	228	nonsynonymous	L > F	At2g05170
2093048	1.00	C	T	225	intronic	-	At2g05640
2100248	1.00	C	T	225	nonsynonymous	E > K	At2g05642
2106924	1.00	C	T	225	intergenic	-	-

Table 5.1 Information on the mutation sites from the *spec1* mapping analysis presented in Figure 5.2 B.

Genomic Position (Chromosome 2)	Allele Frequency	Reference base	Mutant base	Read Quality	Nature of Mutation	Amino Acid Substitution	Genetic Locus
5529194	0.83	G	A	220	intronic	-	At2g13330
6353170	0.75	G	A	221	intronic	-	At2g14796
8638351	0.81	G	A	220	nonsynonymous	L > F	At2g20010
11135618	0.76	G	A	221	intergenic	-	-
12824631	0.80	G	A	221	intronic	-	At2g30050
12831703	0.75	G	A	221	intergenic	-	-
15424394	0.78	G	A	221	nonsynonymous	P > L	At2g36800
15757318	0.79	G	A	221	nonsynonymous	R > C	At2g37550
					intronic	-	At2g37555

Table 5.2 Information on the mutation sites from the *spec2* mapping analysis presented in Figure 5.3 B. There are two genetic locus at genomic position 15757318 due to overlapping genes.

Genomic Position (Chromosome 3)	Allele Frequency	Reference base	Mutant base	Read Quality	Nature of Mutation	Amino Acid Substitution	Genetic Locus
3629883	0.78	C	T	221	nonsynonymous	M > I	At3g11530
5991194	0.78	C	T	221	nonsynonymous	A > T	At3g17510
6012629	0.77	C	T	221	intergenic	-	-
6797862	0.77	C	T	221	nonsense	Q > *	At3g19570
7148617	0.76	C	T	221	intronic	-	At3g20475
7423805	0.77	C	T	221	intronic 3' UTR	-	At3g21175
7554901	0.79	C	T	221	nonsynonymous	D > N	At3g21450
8016378	0.88	C	T	219	CDS synonymous	-	At3g22660
8872593	0.83	C	T	220	intergenic	-	-
9125672	0.87	C	T	219	intronic	-	At3g25040
9453780	0.87	C	T	219	nonsynonymous	R > K	At3g25840
9924758	0.84	G	A	220	intergenic	-	-
10191328	0.86	G	A	219	intergenic	-	-
10315564	0.78	G	A	221	intronic	-	At3g27820
14257645	0.78	G	A	221	intronic	-	At3g42065
14489259	0.77	C	T	221	intergenic	-	-

Table 5.3 Information on the mutation sites from the *spec3* mapping analysis presented in Figure 5.4 B.

Mutant	Genetic Locus	Mutation	Description	Biological Process	Cellular Component
spec 1 sp1 ppi1	At2g05140	L > F	phosphoribosylaminoimidazole carboxylase family protein / AIR carboxylase family protein	'de novo' IMP biosynthetic process	chloroplast, phosphoribosylaminoimidazole carboxylase complex
	At2g05170	L > F	Homologous to yeast VPS11. Forms a complex with VCL1 and AtVPS33. Involved in vacuolar biogenesis. The mRNA is cell-to-cell mobile	endosome organization, intracellular protein transport, lysosome organization, regulation of SNARE complex assembly, vacuole organization, vesicle docking involved in exocytosis	CORVET complex, HOPS complex, endosome, late endosome membrane, nucleus, plant-type vacuole membrane, protein-containing complex
	At2g05642	E > K	Nucleic acid-binding, OB-fold-like protein	N/A	chloroplast
spec2 sp1 ppi1	At2g20010	L > F	Gls protein (DUF810)	N/A	chloroplast, mitochondrion
	At2g36800	P > L	Encodes a DON-Glucosyltransferase. The UGT73C5 glucosylates both brassinolide and castasterone in the 23-O position. The enzyme is presumably involved in the homeostasis of those steroid hormones hence regulating BR activity. Transgenic plants overexpressing UGT73C5 show a typical BR-deficient phenotype.	brassinosteroid metabolic process, detoxification	chloroplast, integral component of membrane, intracellular membrane-bounded organelle
	At2g37550	R > C	A member of ARF GAP domain (AGD), A thaliana has 15 members, grouped into four classes.	vesicle-mediated transport	Golgi apparatus, cytosol, nucleus

spec3 sp1 ppi1	At3g11530	M > I	Vacuolar protein sorting 55 (VPS55) family protein	late endosome to vacuole transport via multivesicular body sorting pathway, negative regulation of protein localization to cell surface	endosome, vacuole
	At3g17510	A > T	Encodes a CBL-interacting protein kinase. Specifically interacts with ECT1 and ECT2	intracellular signal transduction, protein phosphorylation, response to abscisic acid, response to osmotic stress, response to salt stress	cytoplasm, nucleus, plasma membrane
	At3g19570	Q > *	Encodes SCO3 (snowy cotyledon3), a member of a largely uncharacterized protein family unique to the plant kingdom. The sco3-1 mutation alters chloroplast morphology and development, reduces chlorophyll accumulation, impairs thylakoid formation and photosynthesis in seedlings, and results in photoinhibition under extreme CO(2) concentrations in mature leaves. SCO3 is targeted to the periphery of peroxisomes	centrosome cycle, chloroplast organization, spindle assembly	centrosome, cytoplasm, nuclear microtubule, nucleus, peroxisome
	At3g21450	D > N	Protein kinase superfamily protein	phosphorylation, protein phosphorylation	nucleus, plasma membrane
	At3g21175	3' UTR	member of a novel family of plant-specific GATA-type transcription factors	N/A	nucleus
	At3g25840	R > K	Protein kinase superfamily protein	protein phosphorylation	nucleus

Table 5.4 A brief description on the *spec* mutant gene candidates. The information in the columns entitled Description, Biological Process and Cellular Component was obtained from TAIR (https://www.arabidopsis.org) and is presented here without editing.

Chapter 6

GENERAL DISCUSSION

CHAPTER 6: GENERAL DISCUSSION

The first aim of this thesis was to characterize the roles of the *Arabidopsis* SP1 homologues, SPL1 and SPL2. In particular, it was to investigate and assess their potential involvement in the regulation of chloroplast protein import and plastid biogenesis. Regulation of a single set of substrates by multiple E3 ligases would not be without precedent. For example, such a scenario has been well documented in animal systems with regard to the ubiquitination and degradation of the transcription factor p53 (Lee & Gu, 2010). Additionally, with regard to signalling of the plant hormone abscisic acid (ABA), at least two different classes of E3 ligases are reported to regulate the degradation of the ABA INSENSITIVE5 (ABI5) transcription factor (Kelley & Estelle, 2012).

A second aim of this thesis was to discover novel components involved in the SP1-mediated regulation of chloroplast biogenesis and/or new factors that regulate plastid protein import, in order to expand our knowledge concerning the molecular mechanisms underpinning chloroplast and plastid biogenesis.

To achieve these aims, I began by obtaining *spl1* T-DNA mutant lines and confirmed their knockout status by RT-PCR-based methods (Figure 3.2). There were no suitable T-DNA lines available for *SPL2,* and so *SPL2* knockdown (KD) lines were generated in advance of this study by Dr Qihua Ling, using an artificial microRNA-based method. The efficiency of the construct was tested by qRT-PCR and the extent of the knockdown of native *SPL2* gene expression was confirmed to be ~20% relative to the wild type (Figure 3.2). Crossing of the *spl1* mutations and of the *SPL2-KD* transgene into the chlorotic *ppi1* mutant background was then performed, to generate *spl1 ppi1* and *SPL2-KD ppi1* double mutants. The effects on the *ppi1*

phenotype in the double mutants were observed and analysed, and chloroplast development was investigated by TEM. Where relevant, further characterization of the mutants was conducted, such as investigating their ability to import proteins into the chloroplast, and testing for the specificity of the suppression phenotype by crossing with other protein import mutants. Abundance of chloroplast protein import machinery components was also assessed, and the protein-protein interactions were confirmed by immunoprecipitation. An extensive analysis of the evolutionary relationships between the homologues was conducted to achieve a greater understanding of the homology relationships across different species of the Viridiplantae. The ubiquitination activity of the homologues was investigated using a similar approach to SP1 (Ling *et al.*, 2012). Lastly, the effects of the *spl1* mutations and the *SPL2-KD* transgene on plastid developmental transitions was investigated by analyzing leaf senescence.

The secondary aim of this thesis involved the use of a chemical mutagen, EMS to conducting a new forward-genetic screen. In this screen, mutant plants displaying a similar or larger size with a greener phenotype than *sp1 ppi1* were isolated, and their chlorophyll levels were quantified. To expedite identification of the causative gene in each case, the mutant loci were mapped using NGS followed by computational analysis.

In the following sections, I will summarize the main findings presented in the thesis, and provide an appraisal of these findings, as well as suggesting potential avenues of future research.

6.1 Summary of key findings

In this project, both SPL1 and SPL2 were characterized. While SPL1 was shown to have little or no effect on TOC protein degradation, SPL2 on the other hand displayed partial redundancy in function with SP1 as a regulator of the chloroplast protein import machinery. Interestingly, SPL1 appeared to have an effect on the activity of SP1, which led to the hypothesis that SPL1 might be a regulator of SP1.

In accordance with the observation that SPL2 has an effect on TOC protein abundance, the knockdown of *SPL2* led to a recovery in protein import into chloroplasts in the *ppi1* background. Although the suppression phenotype mediated by *SPL2-KD* was weak, and it was unable to suppress the other *toc* mutations tested, direct interaction was observed between the SPL2 protein and the three main TOC proteins, via IP using SPL2-YFP as bait.

The inability of *spl1* mutants to mediate suppression of *ppi1* was consistent with the results of an *in vitro* self-ubiquitination experiment, where unlike SP1 and SPL2, SPL1 did not show any signs of ubiquitination activity *in vitro*.

SP1 was previously reported to have an influence on the rate of plastid transitions during different developmental stages. Therefore, the involvement of the two homologues in the senesce of leaves was tested, and it was found that SPL2 behaves similarly to SP1. The depletion of SPL2 delayed the onset of senescence, while the overexpression of *SPL2* led to an acceleration in the transition process. In contrast, mutation of *SPL1* led unexpectedly to an acceleration of senescence, phenocopying *SP1* overexpression. Moreover, it was found that the presence of SPL1 could negatively influence the activity of SP1, both in an *in vitro* ubiquitination experiment, and in double-overexpression genetic studies.

The screen for enhancers of *sp1* identified three new *spec* mutants. Mapping of the *spec* mutants led to the identification of several interesting gene candidates in each case. The gene candidates for the *spec1* mutation were: an enzyme of the purine biosynthesis pathway; a component annotated as having links to vacuolar biogenesis; and an uncharacterized putatively chloroplast-localized with possible nucleic acid-binding activity. The gene candidates for the *spec2* mutation included: a previously uncharacterized protein with a domain of unknown function; an enzyme putatively involved in brassinosteroid homeostasis; and a protein of the ARF GAP family, termed AGD7 (ARF-GAP DOMAIN 7). Lastly, the gene candidates for the *spec3* mutation included: three different types of protein kinase; a GATA-type transcription factor; a vacuolar protein sorting-associated protein; and a peroxisome-localized protein that was previously characterized to influence chloroplast biogenesis.

6.2 Appraisal of the new regulators of the chloroplast protein import machinery

6.2.1 SPL1

The intriguing discovery of the effect of *spl1* mutation during leaf senescence (Figure 4.6) demonstrated for the first time SPL1's involvement in plastid biogenesis. This was an unexpected discovery as under normal growth conditions, loss of *SPL1* expression did not cause either enhancement or suppression of the chlorotic *ppi1* phenotype (Figure 3.3, Ling *et al.*, 2012). The suggested role of SPL1 in senescence was supported by gene expression data retrieved using Genevestigator (Figure 4.7). However, the absence of SPL1 ubiquitination activity (Figure 4.4) indicated a lack of evidence to conclude a direct effect of SPL1 on the TOC machinery (unlike SP1). Moreover, a question was raised as to how the absence of SPL1 promotes the destabilization of the TOC machinery in order for the plastid transition to take place. One possibility is through negatively influencing the E3 ligase activity of SP1, as suggested by the Genevestigator data (Figure 4.7); the level of *SP1* expression increases during plant senescence, while at the same time the expression of *SPL1* decreases. The data from *in vitro* ubiquitination assays (Figure 4.8) and from plants engineered to overexpress both *SP1* and *SPL1* (Figure 4.9) pointed strongly to a scenario in which SPL1 negatively influences the action of SP1 on the TOC complex. Moreover, plants overexpressing *SPL1* in the *ppi1* mutant background displayed a greener and slightly larger phenotype than *ppi1*, which is reminiscent of the *sp1 ppi1* phenotype (Ling *et al.*, 2012), further suggesting an SPL1 inhibitory effect on the SP1 E3 ligase. However, exactly how SPL1 affects the performance of SP1 needs further investigation.

In considering these findings, it is appropriate to consider again the evolutionary analysis of SP1 and SPL1 (Figure 4.1). Although there are no reports in plants on pseudogenic proteins that are not fully functional but which can interact with and alter the parental protein function, various studies in cancer biology have suggested their existence (Poliseno, 2012; Xiao-Jie *et al.*, 2015). The placement of the SPL1 clade in Figure 4.1 is supported by the percentage ultrafast bootstrap values, and is consistent with the experimental data discussed earlier. However, several questions remain. Why is SPL1 limited to the Brassicaceae, and what happened to the SPL1 in the other families of the land plants? If SPL1 is required to regulate the activity of SP1 during senescence, how is SP1 activity controlled in non-Brassicaceae? Moreover, what are the signalling events that control the abundance of SPL1 during certain developmental stages in the Brassicaceae? This is a similar sequence of events happening with the identity of the TIC complex (Nakai, 2015); although SPL1 is not a prerequisite for *Arabidopsis thaliana* to function. The Brassicaceae are especially diverse, and their speciation rate was reported to be the highest for any land plant group (Hohmann *et al.*, 2015). Could the existence of SPL1 in the Brassicaceae therefore be attributed as a result of duplication error during early speciation event, and over time presumably has evolved to become a pseudogenic protein? The observation for multiple SPL1 paralogues in the Brassicaceae family (Figure 4.1) could support the duplication error hypothesis. A detailed evolutionary analysis of this genus might unravel the mystery. It might also be worth studying the differences of SP1 function in other families of living plants during plant senescence to compare the observations made with SPL1 containing species.

6.2.2 SPL2

A second chloroplast-localized E3 ligase that similarly ubiquitinates the TOC complexes to promote their degradation was identified and characterized in Chapter 3. In the phylogeny, SPL2 forms a discrete monophyletic group that is distinct from the SP1 clade (Figure 4.1). Like SP1, the activity of SPL2 relies on the presence of the conserved C3HC4 zinc finger motif (Figure 4.4). The involvement of SPL2 in leaf senescence mimics that of SP1 (Figure 4.6, Ling *et al.*, 2012); knockdown of *SPL2* delayed the transition process, whereas overexpression of SPL2 accelerated it. As noted earlier in this chapter, multiple E3 ligases can act on the same substrate to enable proper regulation (Lee & Gu, 2010; Kelley & Estelle, 2012). Although it is not known whether SPL2 and SP1 function independently or cooperate together in a single pathway, recent studies in substrate ubiquitination have demonstrated that multiple E3 ligases can act successively on a substrate, instead of the classical view involving a linear sequence as follows: E1 > E2 > a single E3 ligase (Hwang *et al.*, 2010; Scott *et al.*, 2016). The possibility of an association between SPL2 and SP1 should therefore be investigated.

The data suggest an interesting possibility regarding TOC machinery regulation by SPL2 as part of the abiotic stress response. Previously, SP1 has been shown to have an influence on the process of protein import (specifically the TOC complex) during unfavourable growth conditions, thereby promoting stress tolerance (Ling & Jarvis, 2015b). A different study proposed a requirement for the *NERF* (*SPL2*) 3' UTR in promoting drought tolerance in *Arabidopsis* (Gao *et al.*, 2015); however, in that study the possible involvement of NERF/SPL2 E3 ligase activity in drought stress was largely overlooked. In other systems, multiple RING E3 ligases have been documented as controlling various signalling events during abiotic stress

responses (Cho *et al.*, 2017). Therefore, the differential targeting of TOC complexes by two different E3 ligases, possibly with slightly different activities, e.g., in response to different abiotic stress signalling pathways, might allow for greater control of the regulation of the process.

6.3 Conclusion and perspectives

The unprecedented role of SPL1 as a negative regulator has expanded our knowledge of the regulatory mechanisms governing SP1 function in plastid biogenesis. In addition, a second direct regulator of the chloroplast protein import machinery, SPL2, was uncovered in this thesis. The discovery of such complex regulation by multiple E3 ligases (and their regulators) will enable us to better understand the overall process. The data also pave the way for further exploration of the implications of these components in processes such as the aforementioned abiotic stress tolerance. This, along with the homologues' potential or known involvement in plastid transitions, could lead to applications in agricultural settings. For example, SPL2 might prove useful during the process of fruit ripening in certain crops when chloroplasts transform into chromoplasts. The SP1 and SPL1 proteins (and possibly SPL2 as well) might also be relevant in biotic stress tolerance, as has been previously suggested by a study on programmed cell death (Basnayake *et al.*, 2011).

Although we now have a much greater understanding of the roles of SP1, SPL1 and SPL2 in plastid biogenesis, additional work is still required in order to understand the molecular mechanisms fully. Moreover, the identification of *sp1* enhancer mutants has unlocked the possibility of studying this novel area of chloroplast protein import regulation in greater detail and in novel directions. The results of such work will have implications for further understanding the functions of SP1, and their use in crop improvement technologies.

BIBLIOGRAPHY

Albrecht V, Simková K, Carrie C, Delannoy E, Giraud E, Whelan J, Small ID, Apel K, Badger MR, Pogson BJ (2010) The cytoskeleton and the peroxisomal-targeted snowy cotyledon3 protein are required for chloroplast development in Arabidopsis. *Plant Cell* **22**: 3423–38.

Arcus V (2002) OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* **12**: 794–801.

Aronsson H, Jarvis RP (2011) Rapid isolation of arabidopsis chloroplasts and their use for in vitro protein import assays. *Meth. Mol. Biol.* **774**: 281–305.

Baldwin A, Wardle A, Patel R, Dudley P, Park SK, Twell D, Inoue K, Jarvis P (2005) A molecular-genetic study of the Arabidopsis Toc75 gene family. *Plant Physiol.* **138**: 715–33.

Basnayake BMVS, Li D, Zhang H, Li G, Virk N, Song F (2011) Arabidopsis DAL1 and DAL2, two RING finger proteins homologous to Drosophila DIAP1, are involved in regulation of programmed cell death. *Plant Cell Rep.* **30**: 37–48.

Bauer J, Chen K, Hiltbunner A, Wehrli E, Eugster M, Schnell D, Kessler F (2000) The major protein import receptor of plastids is essential for chloroplast biogenesis. *Nature* **403**: 203–207.

Bédard J, Kubis S, Bimanadham S, Jarvis P (2007) Functional similarity between the chloroplast translocon component, Tic40, and the human co-chaperone, Hsp70-interacting protein (Hip). *J. Biol. Chem.* **282**: 21404–14.

Bédard J, Trösch R, Wu F, Ling Q, Flores-Pérez Ú, Töpel M, Nawaz F, Jarvis P (2017) Suppressors of the chloroplast protein import mutant tic40 reveal a genetic link between protein import and thylakoid biogenesis. *Plant Cell* **29**:

1726–1747.

Bruce BD (2000) Chloroplast transit peptides: Structure, function and evolution. *Trends Cell Biol.* **10**: 440–447.

Brzovic PS, Keeffe JR, Nishikawa H, Miyamoto K, Fox D, Fukuda M, Ohta T, Klevit R, Klevit R (2003) Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 5646–51.

Chen L-J, Li H (2017) Stable megadalton TOC-TIC supercomplexes as major mediators of protein import into chloroplasts. *Plant J.* **92**: 178–188.

Chen Y-L, Chen L-J, Li H-M (2016) Polypeptide transport-associated domains of the Toc75 channel protein are located in the intermembrane space of chloroplasts. *Plant Physiol.* **172**: 235–43.

Cheng H-C, Qi RZ, Paudel H, Zhu H-J (2011) Regulation and function of protein kinases and phosphatases. *Enzyme Res.* **2011**: 794089.

Cho SK, Ryu MY, Kim JH, Hong JS, Oh TR, Kim WT, Yang SW (2017) RING E3 ligases: key regulatory elements are involved in abiotic stress responses in plants. *BMB Rep.* **50**: 393–400.

Chou M-L, Chu C-C, Chen L-J, Akita M, Li H (2006) Stimulation of transit-peptide release and ATP hydrolysis by a cochaperone during protein import into chloroplasts. *J. Cell Biol.* **175**: 893–900.

Clough SJ, Bent AF (1998) Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. *Plant J.* **16**: 735–743.

Collingridge PW, Kelly S (2012) MergeAlign: improving multiple sequence alignment performance by dynamic reconstruction of consensus multiple sequence alignments. *BMC Bioinformatics* **13**: 117.

Constan D, Patel R, Keegstra K, Jarvis P (2004) An outer envelope membrane component of the plastid protein import apparatus plays an essential role in *Arabidopsis*. *Plant J.* **38**: 93–106.

David Y, Ziv T, Admon A, Navon A (2010) The E2 ubiquitin-conjugating enzymes direct polyubiquitination to preferred lysines. *J. Biol. Chem.* **285**: 8595–604.

Day PM, Potter D, Inoue K (2014) Evolution and targeting of Omp85 homologs in the chloroplast outer envelope membrane. *Front. Plant Sci.* **5**: 535.

Deshaies RJ, Joazeiro CAP (2009) RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**: 399–434.

Drescher A, Ruf S, Calsa T, Carrer H, Bock R (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* **22**: 97–104.

Evans JR, Caemmerer S von (2011) Enhancing photosynthesis. *Plant Physiol.* **155**: 19.

Flores-Pérez Ú, Jarvis P (2013) Molecular chaperone involvement in chloroplast protein import. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**: 332–340.

Flores-Pérez Ú, Jarvis P (2017) Isolation and suborganellar fractionation of arabidopsis chloroplasts. *Meth. Mol. Biol.* **1511**: 45–60.

Flynn RL, Zou L (2010) Oligonucleotide/oligosaccharide-binding (OB) fold proteins: a growing family of genome guardians. *Crit. Rev. Biochem. Mol. Biol.* **45**: 266–275.

Gao W, Liu W, Zhao M, Li W-X (2015) NERF encodes a RING E3 ligase important for drought resistance and enhances the expression of its antisense gene NFYA5 in Arabidopsis. *Nucleic Acids Res.* **43**: 607–17.

Ginalski K, Kinch L, Rychlewski L, Grishin N V. (2004) BOF: a novel family of

bacterial OB-fold proteins. *FEBS Lett.* **567**: 297–301.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178-86.

Gutensohn M, Schulz B, Nicolay P, Flugge U-I (2000) Functional analysis of the two Arabidopsis homologues of Toc34, a component of the chloroplast protein import apparatus. *Plant J.* **23**: 771–783.

Heins L, Mehrle A, Hemmler R, Wagner R, Küchler M, Hörmann F, Sveshnikov D, Soll J (2002) The preprotein conducting channel at the inner envelope membrane of plastids. *EMBO J.* **21**: 2616–25.

Hofgen R, Willmitzer L (1988) Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Res.* **16**: 9877–9877.

Hohmann N, Wolf EM, Lysak MA, Koch MA (2015) A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* **27**: 2770–2784.

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P (2008) Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. *Adv. Bioinformatics* **2008**: 1–5.

Huang W, Ling Q, Bédard J, Lilley K, Jarvis P (2011) In vivo analyses of the roles of essential Omp85-related proteins in the chloroplast outer envelope membrane. *Plant Physiol.* **157**: 147–59.

Hung W-F (2004) Characterization of arabidopsis glutamine phosphoribosyl pyrophosphate amidotransferase-deficient mutants. *Plant Physiol.* **135**: 1314–

1323.

Hussain H, Chong NF-M (2016) Combined overlap extension PCR method for improved site directed mutagenesis. *Biomed Res. Int.* **2016**: 8041532.

Hust B, Gutensohn M (2006) Deletion of core components of the plastid protein import machinery causes differential arrest of embryo development in Arabidopsis thaliana. *Plant Biol.* **8**: 18–30.

Hwang C-S, Shemorry A, Auerbach D, Varshavsky A (2010) The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. *Nat. Cell Biol.* **12**: 1177–85.

Hyman S, Jarvis RP (2011) Studying arabidopsis chloroplast structural organisation using transmission electron microscopy. *Methods Mol. Biol.* **774**: 113–132.

Inaba T, Li M, Alvarez-Huerta M, Kessler F, Schnell DJ (2003) atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. *J. Biol. Chem.* **278**: 38617–27.

Inoue H, Li M, Schnell DJ (2013) An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 3173–8.

Inoue H, Rounds C, Schnell DJ (2010) The molecular basis for distinct pathways for protein import into Arabidopsis chloroplasts. *Plant Cell* **22**: 1947–60.

Ivanova Y, Smith MD, Chen K, Schnell DJ (2004) Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids. *Mol. Biol. Cell* **15**: 3379–92.

James GV, Patel V, Nordström KJ, Klasen JR, Salomé PA, Weigel D, Schneeberger K (2013) User guide for mapping-by-sequencing in Arabidopsis. *Genome Biol.* **14**: R61.

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map-based cloning in the post-genome era. *Plant Physiol.* **129**: 440–50.

Jarvis P (2008) Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytol.* **179**: 257–258.

Jarvis P, Chen LJ, Li H, Peto CA, Fankhauser C, Chory J (1998) An Arabidopsis mutant defective in the plastid general protein import apparatus. *Science* **282**: 100–3.

Jarvis P, López-Juez E (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.* **14**: 787–802.

Karimi M, Depicker A, Hilson P (2007) Recombinational cloning with plant gateway vectors. *Plant Physiol.* **145**: 1144–1154.

Karimi M, Inzé D, Depicker A (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.* **7**: 193–195.

Kasmati AR, Töpel M, Khan NZ, Patel R, Ling Q, Karim S, Aronsson H, Jarvis P (2013) Evolutionary, molecular and genetic analyses of Tic22 homologues in Arabidopsis thaliana chloroplasts. *PLoS One* **8**: e63863.

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**: 511–518.

Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**: 729–48.

Kelley DR, Estelle M (2012) Ubiquitin-mediated control of plant hormone signaling. *Plant Physiol.* **160**: 47–55.

Kessler F (2012) Chloroplast Delivery by UPS. *Science* **338**: 622 LP-623.

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T,

Nakai M (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **339**: 571–4.

Kikuchi S, Hirohashi T, Nakai M (2006) Characterization of the preprotein translocon at the outer envelope membrane of chloroplasts by blue native PAGE. *Plant Cell Physiol.* **47**: 363–371.

Kikuchi S, Oishi M, Hirabayashi Y, Lee DW, Hwang I, Nakai M (2009) A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane. *Plant Cell* **21**: 1781–97.

Kouranov A, Chen X, Fuks B, Schnell DJ (1998) Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. *J. Cell Biol.* **143**: 991–1002.

Kovacheva S, Bédard J, Patel R, Dudley P, Twell D, Ríos G, Koncz C, Jarvis P (2004) In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. *Plant J.* **41**: 412–428.

Kubis S, Baldwin A, Patel R, Razzaq A, Dupree P, Lilley K, Kurth J, Leister D, Jarvis P (2003) The Arabidopsis ppi1 mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. *Plant Cell* **15**: 1859–71.

Kubis S, Patel R, Combe J, Bédard J, Kovacheva S, Lilley K, Biehl A, Leister D, Ríos G, Koncz C, Jarvis P (2004) Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. *Plant Cell* **16**: 2059–77.

Lee JT, Gu W (2010) The multiple levels of regulation by p53 ubiquitination. *Cell Death Differ.* **17**: 86–92.

Lee KH, Kim SJ, Lee YJ, Jin JB, Hwang I (2003) The M domain of atToc159 plays an essential role in the import of proteins into chloroplasts and chloroplast biogenesis. *J. Biol. Chem.* **278**: 36794–805.

Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**: W242–W245.

Li H, Chiu C-C (2010) Protein transport into chloroplasts. *Annu. Rev. Plant Biol.* **61**: 157–180.

Ling Q, Huang W, Baldwin A, Jarvis P (2012) Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science* **338**: 655–659.

Ling Q, Huang W, Jarvis P (2011) Use of a SPAD-502 meter to measure leaf chlorophyll concentration in Arabidopsis thaliana. *Photosynth. Res.* **107**: 209–214.

Ling Q, Jarvis P (2015a) Functions of plastid protein import and the ubiquitin–proteasome system in plastid development. *Biochim. Biophys. Acta - Bioenerg.* **1847**: 939–948.

Ling Q, Jarvis P (2015b) Regulation of chloroplast protein import by the ubiquitin E3 ligase SP1 is important for stress tolerance in plants. *Curr. Biol.* **25**: 2527–34.

Ling Q, Li N, Jarvis P (2017) Chloroplast ubiquitin E3 ligase SP1: does it really function in peroxisomes? *Plant Physiol.* **175**: 586–588.

Ling Q, Trösch R, Jarvis P (2013) The ins and outs of chloroplast protein transport. In: *Plastid Development in Leaves During Growth and Senescence, Advances in Photosynthesis and Respiration*. Springer, Dordrecht, pp. 239–280.

Lopez-Juez E, Pyke KA (2005) Plastids unleashed: their development and their integration in plant development. *Int. J. Dev. Biol.* **49**: 557–77.

Lorick KL, Jensen JP, Fang S, Ong AM, Hatakeyama S, Weissman AM (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 11364–9.

Maple J, Møller SG (2007) Mutagenesis in Arabidopsis. Humana Press, pp. 197–206.

Min MK, Kim SJ, Miao Y, Shin J, Jiang L, Hwang I (2007) Overexpression of arabidopsis AGD7 causes relocation of golgi-localized proteins to the endoplasmic reticulum and inhibits protein trafficking in plant cells. *Plant Physiol.* **143**: 1601–1614.

Nakai M (2015) The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochim. Biophys. Acta*. **1847**: 957-967.

Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: Experimental applications in Arabidopsis thaliana genetics. *Plant J.* **14**: 387–392.

Nelson N, Ben-Shem A (2004) The complex architecture of oxygenic photosynthesis. *Nat. Rev. Mol. Cell Biol.* **12**: 971-982.

Neuhaus HE, Emes MJ (2000) Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**: 111–140.

Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* **53**: 674–690.

Ottoline Leyser HM, Furner IJ (1992) Characterisation of three shoot apical meristem mutants of Arabidopsis thaliana **116**: 397–403.

Paila YD, Richardson LGL, Schnell DJ (2015) New insights into the mechanism of chloroplast protein import and its integration with protein quality control,

organelle biogenesis and development. *J. Mol. Biol.* **427**: 1038–1060.

Pan R, Hu J (2017) Sequence and biochemical analysis of Arabidopsis SP1 protein, a regulator of organelle biogenesis. *Commun. Integr. Biol.* **10**: e1338991.

Pan R, Satkovich J, Chen C, Hu J (2018) The E3 ubiquitin ligase SP1-like 1 plays a positive role in peroxisome biogenesis in Arabidopsis. *Plant J.* **94**: 836–846.

Pan R, Satkovich J, Hu J (2016) E3 ubiquitin ligase SP1 regulates peroxisome biogenesis in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **113**: E7307–E7316.

Patey CA, Shaw G (1973) Purification and properties of an enzyme duet, phosphoribosylaminoimidazole carboxylase and phosphoribosylaminoimidazolesuccinocarboxamide synthetase, involved in the biosynthesis of purine nucleotides de novo. *Biochem. J.* **135**: 543–5.

Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP (1997) The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* **11**: 3194–3205.

Perry SE, Keegstra K (1994) Envelope membrane proteins that interact with chloroplastic precursor proteins. *Plant Cell* **6**: 93–105.

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**: e45.

Poliseno L (2012) Pseudogenes: newly discovered players in human cancer. *Sci. Signal.* **5**: re5.

Porra RJ, Thompson WA, Kriedemann PE (1989) Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *BBA - Bioenerg.* **975**: 384–394.

Reyes-Prieto A, Weber APM, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* **41**: 147–168.

Richardson LGL, Small EL, Inoue H, Schnell DJ (2018) Molecular topology of the transit peptide during chloroplast protein import. *Plant Cell* **30**: 1789-1806.

Richardson LGL, Singhal R, Schnell DJ (2017) The integration of chloroplast protein targeting with plant developmental and stress responses. *BMC Biol.* **15**: 118.

Rowe B, Mccairns E, Madsen G, Sauer D, Elliott H (1978) De novo purine synthesis in avian liver . Co-purification of the enzymes and properties of the pathway . P B Rowe , E McCairns , G Madsen , D Sauer • When this article is cited De Novo Purine Synthesis in Avian Liver. *J. Biol. Chem.* **253**: 7711–7721.

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**: 725–738.

Saibo NJM, Lourenço T, Oliveira MM (2009) Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. *Ann. Bot.* **103**: 609–23.

Sakamoto W, Miyagishima S-Y, Jarvis P (2008) Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. *Arab. B.* **6**: e0110.

Schelbert S, Aubry S, Burla B, Agne B, Kessler F, Krupinska K, Hörtensteiner S (2009) Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during leaf senescence in Arabidopsis. *Plant Cell* **21**: 767–85.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source

platform for biological-image analysis. *Nat. Methods* **9**: 676–682.

Schleiff E, Soll J, Küchler M, Kühlbrandt W, Harrer R (2003) Characterization of the translocon of the outer envelope of chloroplasts. *J. Cell Biol.* **160**: 541 LP-551.

Schneeberger K (2014) Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* **15**: 662–676.

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D, Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* **6**: 550–551.

Schnell DJ, Blobel G, Keegstra K, Kessler F, Ko K, Soll J (1997) A consensus nomenclature for the protein-import components of the chloroplast envelope. *Trends Cell Biol.* **7**: 303–304.

Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell* **18**: 1121–33.

Scott DC, Rhee DY, Duda DM, Alpi AF, Harper JW, Schulman BAH, Edu JWH) (2016) Two distinct types of E3 ligases work in unison to regulate substrate ubiquitylation. *Cell* **166**: 1198–1214.

Shi L-X, Theg SM (2013) The chloroplast protein import system: From algae to trees. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**: 314–331.

Smith MD, Rounds CM, Wang F, Chen K, Afitlhile M, Schnell DJ (2004) atToc159 is a selective transit peptide receptor for the import of nucleus-encoded chloroplast proteins. *J. Cell Biol.* **165**: 323–34.

Stanga JP, Boonsirichai K, Sedbrook JC, Otegui MS, Masson PH (2009) A role

for the TOC complex in Arabidopsis root gravitropism. *Plant Physiol.* **149**: 1896–905.

Stewart MD, Ritterhoff T, Klevit RE, Brzovic PS (2016) E2 enzymes: more than just middle men. *Cell Res.* **26**: 423–40.

Stone SL, Hauksdóttir H, Troy A, Herschleb J, Kraft E, Callis J (2005) Functional analysis of the RING-type ubiquitin ligase family of Arabidopsis. *Plant Physiol.* **137**: 13–30.

Strittmatter P, Soll J, Bölter B (2010) The chloroplast protein import machinery: a review. *Meth. Mol. Biol.* **619**: 307-321.

Su P-H, Li H-M (2008) Arabidopsis stromal 70-kD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds. *Plant Physiol.* **146**: 1231–41.

Sun H, Schneeberger K (2015) SHOREmap v3.0: Fast and accurate identification of causal mutations from forward genetic screens. *Meth. Mol. Biol.* **1284**: 381–95.

Sveshnikova N, Soll J, Schleiff E (2000) Toc34 is a preprotein receptor regulated by GTP and phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **97**: 4973–8.

Teng Y-S, Su Y, Chen L-J, Lee YJ, Hwang I, Li H (2006) Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *Plant Cell* **18**: 2247–57.

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**: 123–135.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**:

W232–W235.

Tsai J-Y, Chu C-C, Yeh Y-H, Chen L-J, Li H-M, Hsiao C-D (2013) Structural characterizations of the chloroplast translocon protein Tic110. *Plant J.* **75**: 847–57.

Tutar Y (2012) Pseudogenes. *Comp. Funct. Genomics* **2012**: 424526.

Vierstra RD (2009) The ubiquitin–26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **10**: 385–397.

Vries J de, Sousa FL, Bölter B, Soll J, Gould SB (2015) YCF1: A green TIC? *Plant Cell* **27**: 1827–33.

Whatley JM (1978) A suggested cycle of plastid developmental interrelationships. *New Phytol.* **80**: 489–502.

Woodson JD, Joens MS, Sinson AB, Gilkerson J, Salomé PA, Weigel D, Fitzpatrick JA, Chory J (2015) Ubiquitin facilitates a quality-control pathway that removes damaged chloroplasts. *Science* **350**: 450–4.

Wu F-H, Shen S-C, Lee L-Y, Lee S-H, Chan M-T, Lin C-S (2009) Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods* **5**: 16.

Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X (2015) Pseudogene in cancer: real functions and promising signature. *J. Med. Genet.* **52**: 17–24.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A,

Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P (2018) Ensembl 2018. *Nucleic Acids Res.* **46**: D754–D761.

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40.

Zhong R, Thompson J, Ottesen E, Lamppa GK (2010) A forward genetic screen to explore chloroplast protein import in vivo identifies Moco sulfurase, pivotal for ABA and IAA biosynthesis and purine turnover. *Plant J.* **63**: 44–59.

Zufferey M, Montandon C, Douet V, Demarsy E, Agne B, Baginsky S, Kessler F (2017) The novel chloroplast outer membrane kinase KOC1 is a required component of the plastid protein import machinery. *J. Biol. Chem.* **292**: 6952–6964.