Transcriptional regulation of the SUMO system in Arabidopsis thaliana

GARRIDO-ALTAMIRANO, EMMA

TRANSCRIPTIONAL REGULATION OF THE SUMO SYSTEM IN *ARABIDOPSIS THALIANA*

EMMA GARRIDO ALTAMIRANO

THESIS SUBMITTED FOR THE DEGREE OF MASTERS BY RESEARCH IN BIOLOGICAL SCIENCES

DEPARTAMENT OF BIOSCIENCES – DURHAM UNIVERSITY

2018

SUPERVISOR: PROF. ARI SADANANDOM

ABSTRACT

Transcriptional regulation of the SUMO system in *Arabidopsis thaliana*
Emma Garrido Altamirano

Posttranslational modifications provide an important mechanism for plants to adapt to changes in their environment. The Small Ubiquitin-like Modifier (SUMO) has been implicated in the plant response to a number of environmental stresses, such as temperature, drought and salinity stress.
Bio-informatic analysis showed that the deSUMOylation system in crop cereals is disproportionately elaborate when compared to that of the non-cultivated *Brachypodium dystachon.*

RT-qPCR was used to study the changes in transcription of deSUMOylating enzymes, SUMO paralog and SUMOylating enzymes in response to various causes of SUMO conjugate accumulation: SUMO overexpression, loss of the ULP SUMO proteases OTS1 and OTS2 and abiotic stress treatment.
In order to assess the impact of SUMO overexpression on transcription in the SUMO system, two transgenic lines were produced which overexpress SUMO1 to different levels. These lines were characterised phenotypically and were then used to confirm previous data concerning the implication of SUMO in abscisic acid signalling.

Very few overarching patterns were found in the transcriptional response of ULP SUMO protease family members, implying they have individual regulation patterns. Additionally, this study provides evidence that the commonly used deconjugase/maturase dichotomy model requires review.

In contrast, the genes which encode the SUMO paralogs SUMO1/2/3/5 show a clear downregulation pattern to a set level in response to both SUMO overexpression and abiotic stress. These downregulation responses interact with each other and occur in a dose-dependent manner. This study also provides preliminary evidence that *SUMO4* and *SUMO6*, which are generally deemed to be pseudogenes, are in fact transcribed at low levels.

Transcriptional profiling of the genes encoding SUMOylation enzymes yielded a variety of patterns. Remarkably, NaCl treatment led to the inversion of transcriptional dominance between the E3 ligases SIZ1 and HPY2. Finally, this study also uncovers a possible link between SUMOylation and deSUMOylation processes.

STATEMENT OF COPYRIGHT

I can confirm that I did not knowingly plagiarise any of the material in this document.

CONTENTS

ACKNOWLEDGEMENTS

The acknowledgements for this thesis are split in seven ways:

To my supervisor, Prof. Ari Sadanandom, thank you for giving me the opportunity to complete my Master's degree in the lab. The learning curve was steep, but the results are plenty.

To the Department of Biosciences at Durham University, thank you for funding this work through the Grevillea Trust studentship and allowing me extend my Durham adventure.

To all the members of lab 6/7: thank you for your help and sense of humour. Special thanks go out to Dr. Beatriz Orosa, Dr. Moumita Srivastava and soon-to-be-Dr. Linda Millyard for showing me the ropes; to Dr. Beatriz Orosa, Dr. David Lloyd and Bekkie Morrell for their support, and to Dr. Cunjin Zhang for being the life and soul of the lab 6/7 party.

To my BFF, the qPCR machine, thank you for only giving up on me once.

To all those with whom I shared my breaks, whether it be in the Biosciences atrium or in the kitchens of all the houses I have lived in, thank you for the laughter, the opportunities to vent frustrations and the fascinating conversations on a wide range of unexpected topics.

To Jeanna Spencer, Dr. Kate Bruce and the community at St. John's College MCR: thank you for your extensive and enduring advice, support and tea and biscuit provision.

To family, and especially to my parents: thank you for the unwavering support in making all this possible. Thanks you for putting up with my wacky biorhythm and my endless confused ramblings about science, and for sticking with me until the very end.

Choose Corrour.

To all the disabled students who have faced ignorance, incompetence or discrimination in the course of their studies but have refused to remain silent about it:

I dedicate this to you.

LIST OF ABBREVIATIONS

ABA	abscisic acid
ABI5	ABSICISC ACID INSENSITIVE 5
APS	ammonium persulphate
ASP1	ARABIDOPSIS SUMO PROTEASE 1
At	*Arabidopsis thaliana*
ATP	adenosine triphospate
CaMV	Cauliflower mosaic virus
CAT3Ct	CATALASE 3 C-terminal domain
CBP20	NUCLEAR CAP-BINDING PROTEIN 20 kDa subunit
cDNA	copy deoxyribonucleic acid
Col-0 WT	Columbia-0 wild type
COP1	CONSTITUTIVE PHOTOMORPHOGENIC 1
Ct	cycle threshold
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DOG1	DELAY OF GERMINATION 1
DREB2	DEHYDRATION RESPONSIVE ELEMENT BINDING PROTEIN 2
DSUL	DI-SUMO-LIKE
DTT	dithiotreitol
ECL	enhanced chemoluminescence
EDTA	ethylenediaminetetraacetic acid
ELS1	ESD4-LIKE SUMO PROTEASE 1
ESD4	EARLY IN SHORT DAYS 4
EST	expressed sequence tag
FLC	FLOWERING LOCUS C
G1/2-phase	growth phase
GA	gibberellin
GID1	GA INSENSITIVE DWARF 1
GST	glutathione-S-transferase
GUS	β-glucuronidase
HAB1	HYPERSENSITIVE TO ABA 1
HPY2	HIGH PLOIDY 2
Hs	*Homo sapiens*
ICE1	INDUCER OF CBP EXPRESSION 1
ICS1	ISOCHORISMATE SYNTHASE 1
kDA	kilodaltons
MMS21	METHYL METHANE SULFONATE SENSITIVITY 21
MOM1	MORPHEUS MOLECULE 1
M-phase	mitotic phase
MS	Murashige & Skoog
NEDD8	NEURAL PRECURSOR CELL EXPRESSED DEVELOPMENTALLY DOWN-REGULATED PROTEIN 8

NEM	N-Ethylmaleimide
NIA1/2	NITRATE REDUCTASE 1/2
NLS	nuclear localisation signal
NPR1	NONEXPRESSER OF PR GENES 1
Os	*Oryza sativa*
OTS1/2	OVERLY TOLERANT TO SALT 1/2
PAGE	polyacrylamide gel electrophoresis
PCNA	PROLIFERATING CELL NUCLEAR ANTIGEN
PCR	polymerase chain reaction
Pi	inorganic phosphate
PIAL1/2	PROTEIN INHIBITOR OF ACTIVATED STAT-LIKE 1/2
PP2C	PROTEIN PHOSPHATASE 2C
PTM	post translational modification
PVDF	polyvinylidene difluoride
PYL	PYR1-LIKE PROTEIN
PYR1	PYRABACTIN RESISTANCE 1
qPCR	quantitative polymerase chain reaction
RanGAP	RAN GTPASE-ACTIVATING PROTEIN
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
SA	salicylic acid
SAE1/2	SUMO ACTIVATING ENZYME 1/2
Sc	*Saccharomyces cerevisiae*
SCE1	SUMO CONJUGATING ENZYME 1
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SENP	SENTRIN-SPECIFIC PROTEASE
sid2	SALICYLIC ACID INDUCTION DEFICIENT 2
SIM	SUMO interaction motif
SIZ1	SAR AND MIZ 1
SLX5/8	SYNTHETIC LETHAL OF UNKNOWN FUNCTION PROTEIN 5/8
SLY1	SLEEPY 1
SMC5/6	STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6
SMT3	SUPPRESSOR OF MIF TWO 3
SNC	SUPPRESSOR OF NPR1-1 CONSTITUTIVE 1
SnRK6.2	SNF1-RELATED PROTEIN KINASE 6.2
SOX	SUMO overexpression
SPF1/2	SUMO PROTEASE RELATED TO FERTILITY 1/2
S-phase	synthesis phase
SP-ring	Siz/PIAS RING
SUL-γC	SUMO-LIKE-ГC
SUMO	SMALL UBIQUITIN-LIKE MODIFIER
TAE	tris acetic acid EDTA
TAF7	TATA-BOX-BINDING PROTEIN-ASSOCIATED FACTOR 7
TBST	tris buffered saline tween

TCP	TEOSINTE BRANCHED1 - CYCLOIDEA - PROLIFERATING CELL NUCLEAR ANTIGEN FACTORT 1
TEMED	tetramethylethylenediamine
UBC9	UBIQUITIN CARRIER PROTEIN 9
ULP	UBIQUITIN-LIKE PROTEASE
UTR	untranslated region
UV	ultra violet
Y2H	yeast-2-hybrid

1 INTRODUCTION

In recent years, the posttranslational modification (PTM) of proteins has emerged has a key process that integrates plant growth and its response to a changing environment. The effects of PTMs emanate from the addition or removal of a modifier alters the surface characteristics of a PTM target protein, thus influencing the target's ability to interact with other proteins. This can lead to changes in protein abundance, function or localisation (Vierstra & Callis 1999; Hunter 2007). PTMs are widespread and allow for the activation or deactivation of stress sensors and downstream transcription factors that control the expression of hundreds of genes, providing the plant with a fast and flexible way to adapt to changes in its environment (Casey et al. 2017; Spoel 2018).

Posttranslational modifiers can be inorganic, in the cases such as phosphorylation or carboxylation, or small organic groups in cases such as methylation or acetylation. These last two modifications play an important role in the epigenetic regulation of gene expression through the modification of histones (Jenuwein & Allis 2001). Additionally, a number of small peptides also act as modifiers. Ubiquitin is the best studied of these. It is known to be involved in protein degradation via the 26 S proteasome, but also in a diverse range of other function such as iron homeostasis, DNA repair and auxin signalling (Bachmair et al. 2001; Walsh & Sadanandom 2014).

In *Arabidopsis*, ubiquitin is translated in an immature form, either as polyubiquitin or as a fusion with a ribosomal or ubiquitin-like protein. It is then proteolytically processed into its mature, conjugatable form by ubiquitin-specific proteases (Bachmair et al. 2001).Subsequently, ubiquitin is conjugated to its substrates through a mechanism known as the E1-E2-E3 cascade. In a first step, an E1 activating enzyme uses ATP to form a thioester bond with ubiquitin. The ubiquitin moiety is then transferred to form a thioester bond with an E2 conjugating enzyme. Finally, an E3 ligase facilitates the transfer of ubiquitin to its substrate. Depending on the type of E3 ligase involved, it can either act as dual docking site for the E2 conjugating enzyme and the substrate, enabling the transfer of the ubiquitin moiety between the two, or it can itself form a thioester bond with ubiquitin and transfer the posttranslational modifier to its substrate (Weissman 2001).

Arabidopsis thaliana encodes two ubiquitin E1 activating enzymes (Hatfield et al. 1997), some 36 E2 conjugating enzymes (Chen & Hellmann 2013) and up to 1500 E3 ligases (Hua & Vierstra 2011), allowing for an extraordinary degree of specificity. In addition to monomeric attachment to substrates, ubiquitin is also present in polymeric chains, often indicating the targeting of a substrate to the 26 S proteasome (Vierstra 2003). However, modification of a substrate with ubiquitin is a reversible process: deubiquitylating enzymes can detach ubiquitin monomers from their substrates and hydrolyse the bonds between individual ubiquitin subunits in a polyubiquitin chain. *Arabidopsis* is thought to encode around 50 of these deubiquitylating enzymes (Isono & Nagel 2014).

However, ubiquitin is not the only peptide posttranslational modifier. A whole family of ubiquitin-like modifiers exist, incorporating the likes of the SMALL UBIQUITIN-LIKE MODIFIER (SUMO) and NEURAL PRECURSOR CELL EXPRESSED DEVELOPMENTALLY DOWN-REGULATED PROTEIN 8 (NEDD8) (Hochstrasser 2009). The ubiquitin-like modifiers exhibit structures, conjugation and deconjugation mechanisms similar to those of ubiquitin (Miura & Hasegawa 2010). This study focusses on the small ubiquitin-like modifier (SUMO). SUMO is a small (approx. 15 kDa) peptide tag which exhibits ubiquitin-like fold, despite sharing less than 20% sequence similarity with ubiquitin (Vierstra & Callis 1999) Crystal structures of human ubiquitin and SUMO can be seen in figures 1A and 1B respectively.

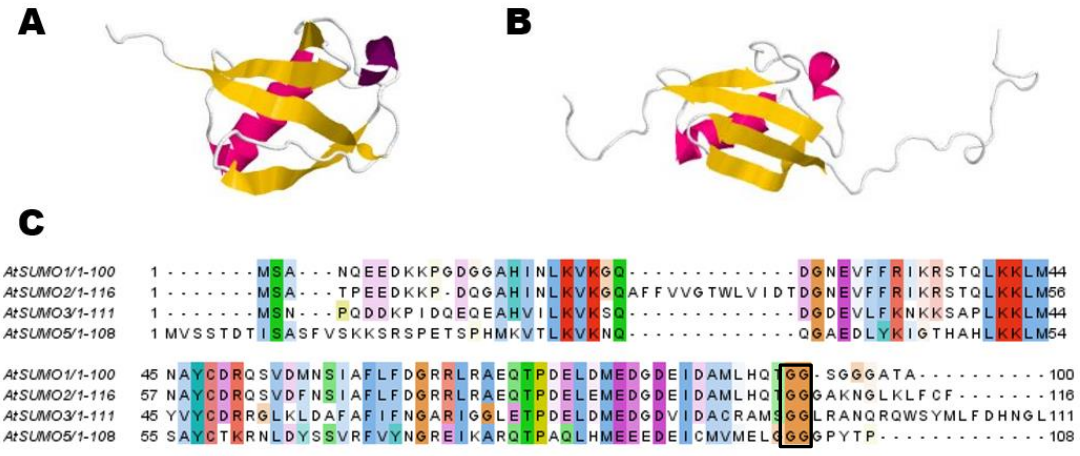

Figure 1: Structural analysis of *Arabidopsis* SUMO proteins

A and B: Human ubiquitin and SUM exhibit a similar fold. *Structures of HsUbiquitin (1UBQ) and HsSUMO1 (1A5R) were taken from the protein data bank and visualised using Jsmol.*

B: Sequence alignment of the four SUMO proteins expressed in Arabidopsis thaliana. *The highly conserved di-glycine motif after which pre-SUMO is cleaved is indicated in black. Sequences were retrieved from the TAIR database, aligned using ClustalX and visualised in Jalview. Jalview's ClustalX-based colour scheme, which colour-codes amino acids by residue type, was applied. The intensity of the colouring reflects the degree of amino acid conservation.*

Similarly to ubiquitin, it is translated in an immature form: SUMO is translated as pre-SUMO, which exhibits an extended C-terminal tail. This C-terminal tail is then partially cleaved by a UBIQUITIN-LIKE PROTEASE (ULP) to expose the highly conserved di-glycine motif in its flexible tail (Hay 2007) which can be conjugated to a lysine residue present in the target protein (Figure 1B).

While there is only one ubiquitin isoform present (Novatchkova et al. 2004) in *Arabidopsis*, it encodes eight SUMO homologs of which four (SUMO1, SUMO2, SUMO3 and SUMO5) are known to be expressed *in vivo* (Kurepa et al. 2003; Saracco et al. 2007), each in a characteristic expression pattern (van den Burg et al. 2010). An alignment of the 4 *Arabidopsis* SUMOs known to be expressed is shown in figure 1C. SUMO1 and SUMO2 share 62% identity and have higher expression levels that SUMO3 and SUMO5. They are considered to act at least in part redundantly as the main plant SUMOs (Saracco et al. 2007). The literature concerning the *Arabidopsis* SUMO isoforms is discussed more extensively in the introduction to chapter 6.

Additionally, *Arabidopsis* encodes seven putative ULP SUMO proteases of which six have been characterised and confirmed as *bona fide* SUMO proteases. Figure 2 shows the structure of the *Saccharomyces cerevisiae* ULP1 in complex with the SUMO homolog SUPPRESSOR OF MIF TWO 3 (SMT3). (Kenji Miura et al. 2007). The scope of the literature concerning the *Arabidopsis* ULP SUMO proteases and their relation to the *Saccharomyces cerevisiae* ULP SUMO protease currently available is limited. The available data is set out in detail in the introduction to chapter 5.

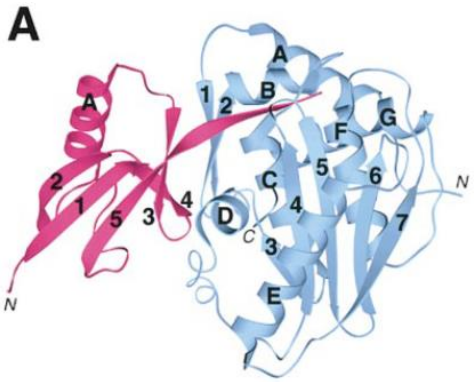

Figure 2: Side (A) view of the yeast SUMO homolog SMT3 (red) in complex with the yeast SUMO protease ULP1 (blue). *The C-terminal flexible tail where the cleavage and conjugation site of SUMO homologs is located, protrudes into ULP1. Figure adapted from* (Mossessova et al. 2000).

In analogy with the ubiquitin system, SUMO conjugation is mediated by an E1 – E2 – E3 cascade (Saitoh et al. 1997). However, the SUMO conjugation cascade encompasses a much smaller group of proteins than the plethora of enzymes which play a role in the ubiquitin conjugation system. The conjugation cycle involves only the heterodimeric E1 SUMO ACTVATING ENZYME (SAE1), the E2 SUMO CONJUGATING ENZYME (SAE1) and the two E3

ligases SAR AND MIZ1 (SIZ1) and HIGH PLOIDY2 (HPY2). Interestingly, SCE1 has also been shown to act as a ligase *in vitro* (Novatchkova et al. 2004; Miura & Hasegawa 2010) and the balance between E2 and E3-mediated SUMOylation *in vivo* is unknown.

Poly-SUMO chains can then be constructed by the E4 ligases PROTEIN INHIBITOR OF ACTIVATED STAT LIKE 1/2 (PIAL1 & PIAL2) (Tomanov et al. 2014) Currently, very little is known about the function of poly-SUMO chains in plants, but PIAL1/2 are thought to be involved in stress tolerance and transposition (Nukarinen et al. 2017). The SUMOylation enzymes, in particular the E3 ligases SIZ1 and HPY2 are the most extensively characterised section of the *Arabidopsis* SUMO system. A review of the current state of knowledge on the *Arabidopsis* SUMOylation enzymes can be found in the introduction to chapter 7.

As with ubiquitination, SUMOylation is reversible, meaning there is also a deconjugation part to the SUMO cycle. Interestingly, deconjugation is accomplished by the same ULP SUMO proteases that act in SUMO maturation (Geiss-Friedlander & Melchior 2007). The SUMO cycle is illustrated in figure 2. Together with the multiplicity of SUMO paralogs, the variety of ULP SUMO protease provide potential for diversity within the SUMO system (Mukhopadhyay & Dasso 2007).

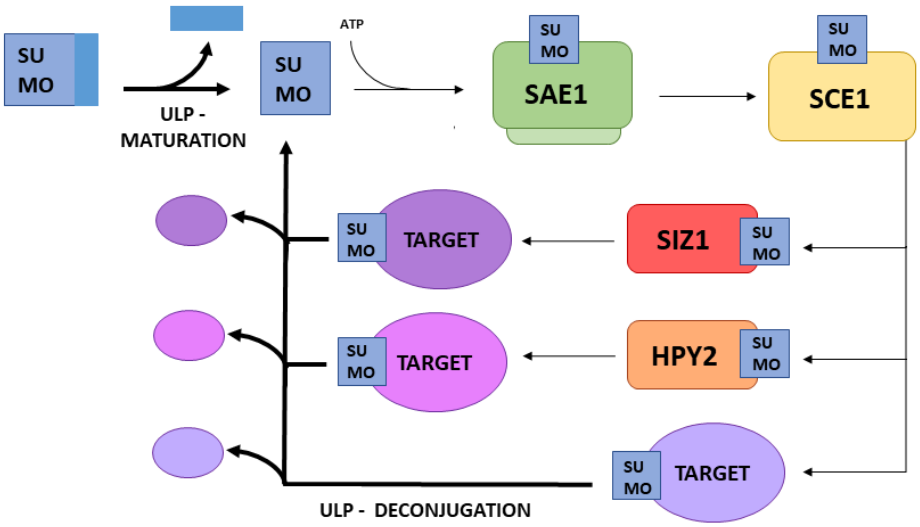

***Figure 3**: The SUMO cycle. SUMO is produced as immature pre-SUMO. ULP SUMO proteases cleave pre-SUMO to produce mature SUMO. The mature SUMO is then activated by the E1 SAE and transferred to the E2 SCE. SCE can directly SUMOylate target proteins or it can pass SUMO on to the E3s HPY2 and SIZ1, which in turn SUMOylate their target proteins. When SUMOylation of a target protein is no longer required, ULP SUMO proteases can remove the tag in a process called deconjugation. SAE: SUMO activating enzyme; SCE: SUMO conjugating enzyme; SIZ1; SAP and Miz1; HPY2: High Ploidy2.*

While the process of SUMOylation may seem simple, it is essential: *sum1sum2, sae1b* and *sce1* mutant are embryonically lethal (Saracco et al. 2007). SUMOylation of target proteins generally takes place at the SUMO modification consensus motif, ΨKXE/D where Ψ is a large hydrophobic amino acid and X can be any amino acid (Rodriguez et al. 2001). Other SUMOylation motifs have also been identified, such as ones characterised by phosphorylatable, negatively charged or hydrophobic amino acids (Tomanov 2014). However, covalent conjugation is not necessary for SUMO-target interactions: some proteins, notably the gibberellin receptor GA INSENSITIVE DWARF 1 (GID1) (Conti et al. 2014), possess SUMO INTERACTION MOTIF (SIM) domains which allow them to interact with SUMO noncovalently (Perry et al. 2008; Gareau & Lima 2010).

When a plant perceives stress, SUMO conjugates accumulate (Kurepa et al. 2003; Castro et al. 2012). However, when the situation has passed, plants must deSUMOylate proteins in order to return to a non-stressed metabolism. As mentioned above both these processes involve SUMO proteases: the first to produce mature SUMO from pre-SUMO, the second to deconjugate SUMO conjugates. The process of stress-induced modification of proteins (with SUMO as well as with other posttranslational modifiers) involves a wide range of targets.

Many key transcriptional regulators including DEHYDRATION RESPONSIVE ELEMENT BINDING PROTEIN 2 (DREB2), INDUCER OF CBP EXPRESSION 1 (ICE1) (controlling cold, heat, salt and drought stress) and ABSCISIC ACID INSENSITIVE 5 (ABI5) (regulator of stress hormone abscisic acid) have been shown to undergo posttranslational modification in order to be effective in promoting plant stress adaptation (Miura & Hasegawa 2010). PTMs have also been shown to be influential in countering biotic stresses (Casey et al. 2017).

PTMs are also relevant to agronomically important alleles. For example, the DELLA proteins, which are responsible for dwarfed, high-yielding varieties of the “green revolution” (Peng et al. 1999) undergo multiple PTMs: they have been shown to undergo phosphorylation (Itoh et al. 2005), ubiquitination (Dill 2004), and SUMOylation (Conti et al. 2014).

These multiple PTMs can take place either at the same or a different site, each modification leading to a different outcome for the target (Hunter 2007). For example, in yeast, SUMOylation, monoubiquitination or polyubiquitination of the same lysine in the PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) protein lead to different pathways of DNA replication or repair (Gill 2004).

However, the mechanism through which these difference PTMs influence protein fate is likely to be more complex than simple competition. Ulrich reviewed a number of mechanisms through which phosphorylation, ubiquitination and SUMOylation may interact, such as loss of interaction surface, sequestration or nuclear translocation (Ulrich 2005).

More recently, it has been discovered that the different protein modifiers can also act together. A class of SUMO-targeted ubiquitin ligases, which require the target to be SUMOylated before it can be ubiquitinated, has been identified, challenging the notion that SUMO and ubiquitin necessarily act competitively and/or antagonistically (Perry et al. 2008). Subsequently, it was shown that hybrid SUMO-ubiquitin chains can be specifically recognised as such (Guzzo & Matunis 2013), opening up a whole new angle on PTM cross-talk. While the study of interactions between posttranslational modifiers will undoubtedly lead to a better understanding of the intricacies of plants' ability to adapt to their environment at the molecular level, the focus of this study is limited to SUMO and the SUMO system.

Over the past two decades, it has become increasingly clear that SUMO has an important role to play as a post-translational modifier in both plant growth and development and in resistance to biotic and abiotic stresses. This study focussed on its role within the stress response in *Arabidopsis thaliana*.

This study explores the importance of SUMO conjugation and deconjugation in plants. Firstly, bio-informatic tools are used to investigate the evolutionary development of the ULP SUMO protease family in crop and non-crop plants and assess whether cultivation level influences the elaboration level of the deSUMOylations system.
Secondly, the influence of SUMO overexpression was further characterised. SUMO overexpression was previously been reported to impact abscisic acid (ABA) signalling (Lois et al. 2003). Expanding on this knowledge, the influence of both low and high levels of SUMO overexpression is assessed phenotypically.
Finally, RT-qPCR is used to transcriptionally profile the tree branches of the SUMO system: the deSUMOylation enzymes, SUMO paralogs and deSUMOylation enzymes. This study assesses responses of these genes to a number of changes which are known to cause the accumulation of SUMO conjugates: SUMO overexpression, loss of OVERLY TOLERANT TO SALT 1/2 (OTS1/2) SUMO protease function and exposure to abiotic stress.

2 MATERIALS AND METHODS

2.1 BIO-INFORMATIC ANALYSIS

NCBI-blast (p-blast, DELTA-blast & PSI-blast) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to retrieve protein sequences (Altschul et al. 1997), using the AtOTS2 catalytic domain as a query for the Brassica crops and the AtOTS2 and BdOTS2 catalytic domains for the cereal crops. As the rice genome is well annotated, the putative function search tool from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) (Kawahara et al. 2013) was also consulted to find putative members of the ULP1 family. Alignments were made using ClustalX (https://www.ebi.ac.uk/Tools/msa/clustalo/) and visualised in Jalview (http://www.jalview.org/). Alignments were colour-coded using the ClustalX-based colour scheme (http://www.jalview.org/help/html/colourSchemes/clustal.html). For conservation-based intensity colourings, the conservation colour increment value was set to 10. Bootstrap neighbour-joining trees were produced using ClustalX and visualised using Figtree (http://tree.bio.ed.ac.uk/software/figtree/). SUMO and SIM site predictions were carried out with GPS-SUMO (http://sumosp.biocuckoo.org/online.php). Available microarray data for *SUM4* and *SUM6* was accessed and visualised using Genevestigator (https://genevestigator.com/gv/index.jsp) (Hruz et al. 2008) and eFP Gene Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) (Winter et al. 2007).

2.2 PLANT CULTURE

2.2.1 Medium preparation

0.5 Murashige & Skooge (MS) medium was prepared using 2.2 g MS basal salt mixture (Duchefa, https://www.duchefa-biochemie.com/) and 7.8 g phytoagar (Melford, https://www.melford.co.uk/) per litre. For abscisic acid treatment plates, abscisic acid (Sigma, https://www.sigmaaldrich.com/) dissolved to 100 mM in DMSO (VWR, https://www.vwr.com/) was added to a final concentration of 1µM after autoclaving. For NaCl treatment plates, NaCl (Sigma, https://www.sigmaaldrich.com/) dissolved to 5 M in Milli-Q water was added to final concentrations of 50 mM and 100 mM after autoclaving.

2.2.2 Seed sterilisation

Seeds were gas-sterilised in an air-tight container. Seeds were placed in open Eppendorf tubes and left overnight with an Erlenmeyer flask containing 97 ml NaClO (VWR, https://www.vwr.com/) and 3 ml 37% HCl (VWR, https://www.vwr.com/). Excess gas was then removed by airing the open tubes in a laminar flow hood for 20 minutes.

2.2.3 Growth conditions

Seeds were spread on 0.5 MS plates before being transferred to a 4°C room for 72 hours. Seedlings were then grown at 22°C in short-day conditions (8 h light/16 h dark) with a light intensity of 100 $\mu mol \cdot m^{-2} \cdot s^{-1}$.

2.2.3.1 Seedlings

Seedlings were transferred to treatment plates on the 4th day after transfer to the growth cabinet. For root length and seedling qPCR assays, plants were spaced at 24 seedlings per plate.

2.2.3.2 Mature plants

Mature plants were grown on 0.5 MS up to 21 days and then transferred to soil in soil. Plants were subsequently grown for 14 days in short day conditions before tissue harvest.

2.2.4 Tissue storage

After harvesting, tissues were immediately frozen in liquid nitrogen. Processed samples were stored at -80 °C.

2.3 PHENOTYPIC ASSAYS

2.3.1 Germination assay

Following stratification treatment for 3 days at 4°C, seeds were transferred to 0.5 MS plates in a 1 x 1 cm grid. Germination was determined as the emergence of the root tip and was assessed every 24 hours starting 30 hours after transfer. Groups of 39 seeds were assessed together in order to calculate germination rate.

2.3.2 Root length assay

Seedlings were grown vertically on 0.5 MS plates for 4 days, then transferred to control or treatment plates until they were 10 days old. Increase in root length from time of transfer was measured with ImageJ (https://imagej.nih.gov/ij/index.html).

2.3.3 Fresh weight assay

Seedlings were grown vertically on 0.5 MS plates for 4 days, then transferred to control or treatment plates until they were 10 days old. 10 to 12 seedlings were weighed together to calculate fresh weight.

2.4 RNA EXTRACTION

For each extraction, 24 10 day-old seedlings were frozen in liquid nitrogen and ground to a fine powder using a pestle and mortar. Following the Direct-zol™ RNA Miniprep kit protocol (Zymo Research, https://www.zymoresearch.eu/), 1 ml TRIzol (Zymo Research, https://www.zymoresearch.eu/) was added and the tissue was ground with the TRIzol for 30 s. The mix was transferred to a 1.5 ml Eppendorf tube, to which 250 µl chloroform (VWR, https://www.vwr.com/) was added. The tubes were vortexed for 15 s at maximum speed before centrifugation at 12,000 g for 15 mins at 4°C.

After centrifugation, the supernatant was removed carefully and transferred to a clean 1.5 ml Eppendorf tube. An equal volume of 100% ethanol was added, and the contents transferred to Direct-Zol™ RNA miniprep (Zymo Research, https://www.zymoresearch.eu/) columns. The RNA was washed according to the Direct-Zol™ miniprep kit protocol, including treatment with DNAse I, and eluted into 50 µl DNAse-RNAse-free water. RNA concentration was measured using a NanoDrop 1000 spectrophotometer (ThermoFisher, https://www.thermofisher.com/). RNA was stored at -80°C.

2.5 cDNA SYNTHESIS

cDNA synthesis was carried out using a modified protocol based on the Superscript II reverese transcription protocol (Invitrogen, https://www.thermofisher.com/). The manipulations for this reaction were carried out on ice. 1 µg RNA, diluted in 10 µl sterile distilled water was used per synthesis reaction. 1µl of 10 mM oligo dT (VWR, https://www.vwr.com/) was added and the mixture returned to the ice to allow annealing. It was then incubated at 65 °C for 5 min. Subsequently, the following reagents were added:

Reagent	Volume (µl)
5X strand buffer (Invitrogen, https://www.thermofisher.com/)	4
0.1 mM DTT (Invitrogen, https://www.thermofisher.com/)	2
10 mM each dNTP mix (VWR, https://www.vwr.com/)	1
RNAse OUT (Invitrogen, https://www.thermofisher.com/)	1

The mixtures were spun down in a microcentrifuge and incubated at 42 °C for 1 min. Thereafter, 1 µl (200 U) Superscript II (Invitrogen, https://www.thermofisher.com/) was added. The mixture was incubated at 42 °C for 50 min, then at 72 °C for 15 min. Finally, 30 µl of sterile distilled water was added. DNA concentration was measured using a NanoDrop 1000 spectrophotometer (ThermoFisher, https://www.thermofisher.com/). cDNA was stored at -20 °C.

2.6 PCR

2.6.1 Primer design

All primers were ordered from Integrated DNA Technologies (https://eu.idtdna.com/pages/home) and used at 10 µM concentration. The sequences of all primers used in this study can be found in supplementary table 1.

2.6.1.1 qPCR primers

Primers were designed with Primer3 (http://bioinfo.ut.ee/primer3/) and NCBI primer-blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers were tested by performing a PCR reaction followed by gel electrophoresis. In the case of a single visible band, they were further tested using the StepOnePlus (Applied Biosystems, https://www.thermofisher.com/) qPCR standard curve programme with 1/1, ½ and ¼ dilutions of stock cDNA. Only primers with an R^2 value above or equal to 0.95 were retained.

2.6.2 mRNA amplification for sequencing

Standard PCR was performed using MyTaq™ Red Mix (Bioline, https://www.bioline.com/). The reaction was performed in a 50 µl volume using the following reagents:

PCR mix	Volume (µl)
MyTaq™ Red Mix (Bioline, https://www.bioline.com/)	25
Sterile distilled water	17.5
Forward primer (Integrated DNA Technologies https://eu.idtdna.com/pages/home)	2.5
Reverse primer (Integrated DNA Technologies https://eu.idtdna.com/pages/home)	2.5
cDNA stock	2.5

Primers were used at a 10µM stock concentration. Mixtures were spun down in a microcentrifuge. PCR reactions were run in a Veriti ™ Thermal cycler (Applied Biosystems, https://www.thermofisher.com/) on the following programme:

94 °C	3 min	
94 °C	30 s	Repeat for
55 °C	90 s	35
72 °C	30 s	cycles
72 °C	7 min	

The fragment was visualised on a 0.8% agarose (Severn Biotech, http://www.severnbiotech.com/) gel with 1µl ethidium bromide (VWR, https://www.vwr.com/) per 100 ml volume. The band was excised from the gel and purified.

2.6.3 qPCR

RT-qPCR was performed using a modified protocol based on the universal SYBR Green Quantitative PCR protocol (Sigma, https://www.sigmaaldrich.com/technical-documents/protocols/biology/sybr-green-qpcr.html#protocol).

For each 10 µl reaction, the following mixes were prepared:

SYBR mix	**Volume (µl)**	**cDNA mix**	**Volume (µl)**
SYBR Green (Agilent, https://www.agilent.com/)	5	cDNA stock	0.5
Forward primer (Integrated DNA Technologies https://eu.idtdna.com/pages/home)	0.5	DEPC water	3.5
Reverse primer (Integrated DNA Technologies https://eu.idtdna.com/pages/home)	0.5		
ROX (Agilent, https://www.agilent.com/)	0.15		

Primers were used at a 10 µM stock concentration. 6.2 µl of SYPBR mix and 3.8 µl of cDNA mix were then combined in a 96-well FAST plate (ThermoFisher, https://www.thermofisher.com/) The reactions were run in a StepOnePlus qPCR machine (ThermoFisher, https://www.thermofisher.com/) on the following programme:

95 °C	3 min	
94 °C	10 s	Repeat for
60 °C	30 s	40 cycles

The cycle threshold (Ct) readings were then normalised to two housekeeping genes. Relative transcript abundance was calculated using the $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen 2001).

2.7 GEL ELECTROPHORESIS

Gels were prepared using 0.8-2% agarose (Severn Biotech, http://www.severnbiotech.com/) in 1X TAE buffer. The mixture was heated in a microwave until the agarose was fully dissolved, the cooled under running water before addition of 1 µl ethidium bromide (VWR, https://www.vwr.com/) per 100 ml volume. The gels were left until solidified.

The gels were submerged in gel electrophoresis tanks filled with 1X TAE buffer and wells were loaded with 3 µl of ladder or 9 µl of PCR sample. Depending on the expected size of the fragment, 1 kb or 50 bp hyperladder (Bioline, https://www.bioline.com/) was used. Gels were run at ~100 V and visualised using a UV transilluminator.

2.8 GEL EXTRACTION

Using a UV transilluminator, the relevant band was excised and transferred to a pre-weighed Eppendorf tube. The tube was re-weighed to determine the weight of the gel fragment. The extraction was performed as instructed by the Zymoclean™ Gel DNA Recovery kit (Zymo Research, https://www.zymoresearch.eu/). The final elution volume was 50 µl.

2.9 PROTEIN EXTRACTION

Plant tissue was frozen in liquid nitrogen and ground to a fine powder. The tissue was weighed and 1 µl/mg of protein extraction buffer was added. The mixture was centrifuged for 8 minutes at 10,000 RPM. The supernatant was isolated and re-centrifuged under the same conditions. The supernatant was re-isolated and an equal volume of 2x Laemelli buffer was added. The mixture was then incubated at 98 °C for 5 minutes. The protein concentration was measured using a Direct Detect Spectrometer (Merck Milipore, www.merckmillipore.com/).

2.10 SDS-PAGE

Protein samples were separated by electrophoresis on a 10% SDS gel. Resolving gels were prepared as follows:

Reagent	Volume
Sterile H_2O	3.5 ml
30 % acrylamide solution (Sigma, https://www.sigmaaldrich.com/)	3.4 ml
1.5 M Tris PH 8.8 (Melford, https://www.melford.co.uk/)	2.6 ml
10 % SDS (Melford, https://www.melford.co.uk/)	100 µl
10 % APS (ThermoFisher, https://www.thermofisher.com/)	100 µl
TEMED (ThermoFisher, https://www.thermofisher.com/)	8 µl

After pouring using 7 ml of the above mixture, gels were left to polymerise under a layer of 100 % isopropanol (VWR, https://www.vwr.com/).
The stacking gel was then prepared as follows:

Reagent	Volume
Sterile H_2O	3.4 ml
30 % acrylamide solution (Sigma https://www.sigmaaldrich.com/)	830 µl
1 M Tris-HCl PH 6.8	630 µl
10 % SDS (Melford, https://www.melford.co.uk/)	50 µl
10 % APS (ThermoFisher, https://www.thermofisher.com/)	50 µl
TEMED (ThermoFisher, https://www.thermofisher.com/)	5 µl

3 ml of the above mixture was used to pour the stacking gel. The gel was left to polymerise with a 15 well comb insert. Equal amounts of protein were loaded onto the gel together with 20 µl 4x SDS loading buffer, together with 5µl PAGE ruler protein ladder (ThermoFisher, https://www.thermofisher.com/).

The gel was run in 1x running buffer at 60 V for 3 hours or until fully separated. Proteins were transferred to PVDF membrane (ThermoFisher, https://www.thermofisher.com/) in 1x transfer buffer. The transfer was run overnight at 30 V and 4°C.

2.11 Western Blotting

After transfer, membranes were blocked in TBST with 5% skimmed milk powder for 1 hour. They were then incubated with primary (rabbit α-strep) antibody for 2 hours at 1:10000 dilution in TBST. They were subsequently washed 5 times for 5 minutes in TBST before the secondary antibody (α -rabbit) was added at 1:20000 dilution and the membrane was incubated for 1 hour. 5 more TBST washes were then performed. Blots were then exposed on photographic film using ECL solution (ThermoFisher, https://www.thermofisher.com/). Loading controls were visualised using Ponceau staining.

2.12 Buffers

2.12.1 TAE buffer

Reagent	Concentration
Tris PH 8 (Melford, https://www.melford.co.uk/)	40 mM
Glacial acetic acid (VWR, https://www.vwr.com/)	20 mM
EDTA (Sigma, https://www.sigmaaldrich.com/)	1 mM

2.12.2 **Protein extraction buffer**

Reagent	Concentration
NaCl (Sigma, https://www.sigmaaldrich.com/)	150 mM
Tris-HCl pH 8	50 mM
Igepal (Sigma, https://www.sigmaaldrich.com/)	1%
Sodium deoxycholate (Sigma, https://www.sigmaaldrich.com/)	0.5%
SDS (Melford, https://www.melford.co.uk/)	0.1%
EDTA (Sigma, https://www.sigmaaldrich.com/)	1 mM
NEM (Alfa Aesar, https://www.alfa.com/)	50 mM

Additionally, a proteinase inhibitor tablet (La Roche, https://www.roche.com/) was added to the solution.

2.12.3 2x Laemelli buffer

Reagent	Concentration
Tris-HCl pH 6.8	125 mM
SDS (Melford, https://www.melford.co.uk/)	4 %
Glycerol (VWR, https://www.vwr.com/)	20 %
β-mercapto ethanol (Sigma, https://www.sigmaaldrich.com/)	10 %
Bromophenol blue	0.005 %

2.12.4 **4x SDS loading buffer**

Reagent	Concentration
SDS (Melford, https://www.melford.co.uk/)	8 %
β-mercapto ethanol (Sigma, https://www.sigmaaldrich.com/)	20 %
Glycerol (VWR, https://www.vwr.com/)	40%
Tris-HCl PH 6.8	200 mM
Bromophenol blue	0.008%
EDTA (Sigma, https://www.sigmaaldrich.com/)	50 mM

2.12.5 **Running buffer**

Reagent	Concentration
Tris (Melford, https://www.melford.co.uk/)	25 mM
Glycine (VWR, https://www.vwr.com/)	190 mM
SDS (Melford, https://www.melford.co.uk/)	0.1 %

2.12.6 **Transfer buffer**

Reagent	Concentration
Tris (Melford, https://www.melford.co.uk/)	25 mM
Glycine (VWR, https://www.vwr.com/)	190 mM
Methanol (VWR, https://www.vwr.com/)	20 %

2.12.7 **TBST**

Reagent	Concentration
Tris (VWR, https://www.vwr.com/)	20 mM
NaCl (Sigma, https://www.sigmaaldrich.com/)	150 mM

The solution pH was adjusted to 7.6 before adding 0.1 % Tween 20.

2.13 DATA ANALYSIS

2.13.1 qPCR data analysis

Cycle threshold (Ct) data generated using the RT-qPCR protocol described above were normalised to Ct values obtained for *Arabidopsis* housekeeping genes *UBIQUITIN CARRIER PROTEIN 9* (*UBC9)* and N*UCLEAR CAP-BINDING PROTEIN 20 kDa SUBUNIT (CBP20)*.

Relative transcript abundance was calculated using the $2^{-\Delta\Delta CT}$ (Livak & Schmittgen 2001). Error due to differences in amplification efficiency were minimised through careful selection of primers (see 2.6.1). The combined effect of genotype and stress treatment was assessed by comparing the relevant ΔCt values to those obtained for the corresponding gene in Col-0 WT plants grown in control MS medium conditions. The two effects were also assessed separately. In the case of genotype effect, the ΔΔCt values were calculated relative to ΔCt values obtained in Col-0 WT plants exposed to the relevant stress conditions. For the assessment of the effect of abiotic stress treatment, , the ΔΔCt values were calculated relative to ΔCt values obtained in plants of the relevant genotype grown in control MS medium conditions.

2.13.2 Statistical analysis

In sample groups too small for their distribution to be assessed, the Mann-Whitney U-test was used. This was the case for fresh weight and RT-qPCR assays. The latter were performed using three biological and two technical repeats. The larger sample groups generated in the root length and germination time assay were assessed using Student's t-test. All error bars show standard error and significance was assessed for each sample relative to the relevant control sample.

3 BIO-INFORMATIC ANALYSIS OF ULP NUMBERS IN CROPS AND THEIR NON-CULTIVATED RELATIVES

3.1 INTRODUCTION

As mentioned in the introduction, the *Arabidopsis* SUMOylation machinery is substantially less extensive than its ubiquitination machinery. Specifically, only three ligation mechanisms have currently been demonstrated *in vitro*: either directly by the E2 SCE1 or through the E3s SIZ1 or HPY2 (Colby et al. 2006; Ishida et al. 2009; Miura et al. 2005). Meanwhile, the ULP family of SUMO proteases is generally considered to have 7 members (Novatchkova et al. 2004), of which 6 have been confirmed as bona fide SUMO proteases (Reeves 2002; Hermkes et al. 2011; Conti et al. 2008; Kong et al. 2017; Liu et al. 2017).

Additionally, a number of new ULP candidates were identified though bio-informatics approaches (Kurepa et al. 2003; Lois 2010). However, few of these have large numbers of Expressed Sequence Tags (ESTs) and many were later hypothesised to be the result of a transposition event and therefore not to encode functional cellular ULPs (Hoen et al. 2006).

Currently, the most likely ULP candidate outside the consensus group of 7 is At3g48480, sometimes termed ULP1e (Castro 2013). It has an EST number comparable the low-expressed ULPs *ELS1* and *ULP1b*. Together with the transposon-ULPS, it forms a separate clade to the Ulp1s and Ulp2s. It was therefore renamed ULP3 in this study. It is phylogenetically close to the human SENTRIN-SPECIFIC PROTEASES 6 and 7 (HsSENPs 6/7) (Lois 2010), which are thought to be poly-SUMO deconjugases (Lima & Reverter 2008). In order to establish whether homologues are present in *Arabidopsis* relatives, it was included in the analysis.

Previous data mining using bioinformatics revealed an increased number of ULP SUMO proteases in the crop plants *Oryza sativa* and *Zea mays* when compared to *Arabidopsis* (Yates et al. 2016). Multiple factors may lie at the root of this increase, including a difference between monocot and dicot plants or a difference in cultivation status between the highly cultivated rice and maize crops and the uncultivated *Arabidopsis*. A number of the processes known to involve SUMOylation relate to characteristics commonly selected for in plant

breeding. In order to establish whether the cultivation level (non-cultivated 'weed' versus crop) has any influence on the size of the ULP family, bioinformatic tools were used to investigate the number of ULP SUMO proteases in crops and their non-crop relatives. Phylogenetic trees were then constructed in order to ascertain whether the phylogenetic groups observed in *Arabidopsis* are conserved in other plants or whether new clades emerge.

3.2 RESULTS

3.2.1 ULP SUMO proteases in the Brassica family

As the *Arabidopsis thaliana* proteome is well characterised, I started out by comparing it to Brassica crops *Brassica rapa* and *Brassica oleracae,* whose cultivars make up most of the vegetable cabbages, and to the oilseed rape *Brassica napus*. The number of sequences retrieved for each species can be found in table 1.

Table 1: ULP SUMO protease sequences retrieved in Arabidopsis thaliana and crop Brassicas. *The number of putative ULPs is conserved in Brassica rapa, while one sequence fewer was retrieved in Brassica oleracea. Meanwhile, two more were retrieved in Brassica napus.*

Organism	Number of ULP sequences found
Arabidopsis thaliana	**8**
Brassica rapa	8
Brassica napus	10
Brassica oleracea	7

The sequences found show distinct homology within the catalytic domain (Figure 4).

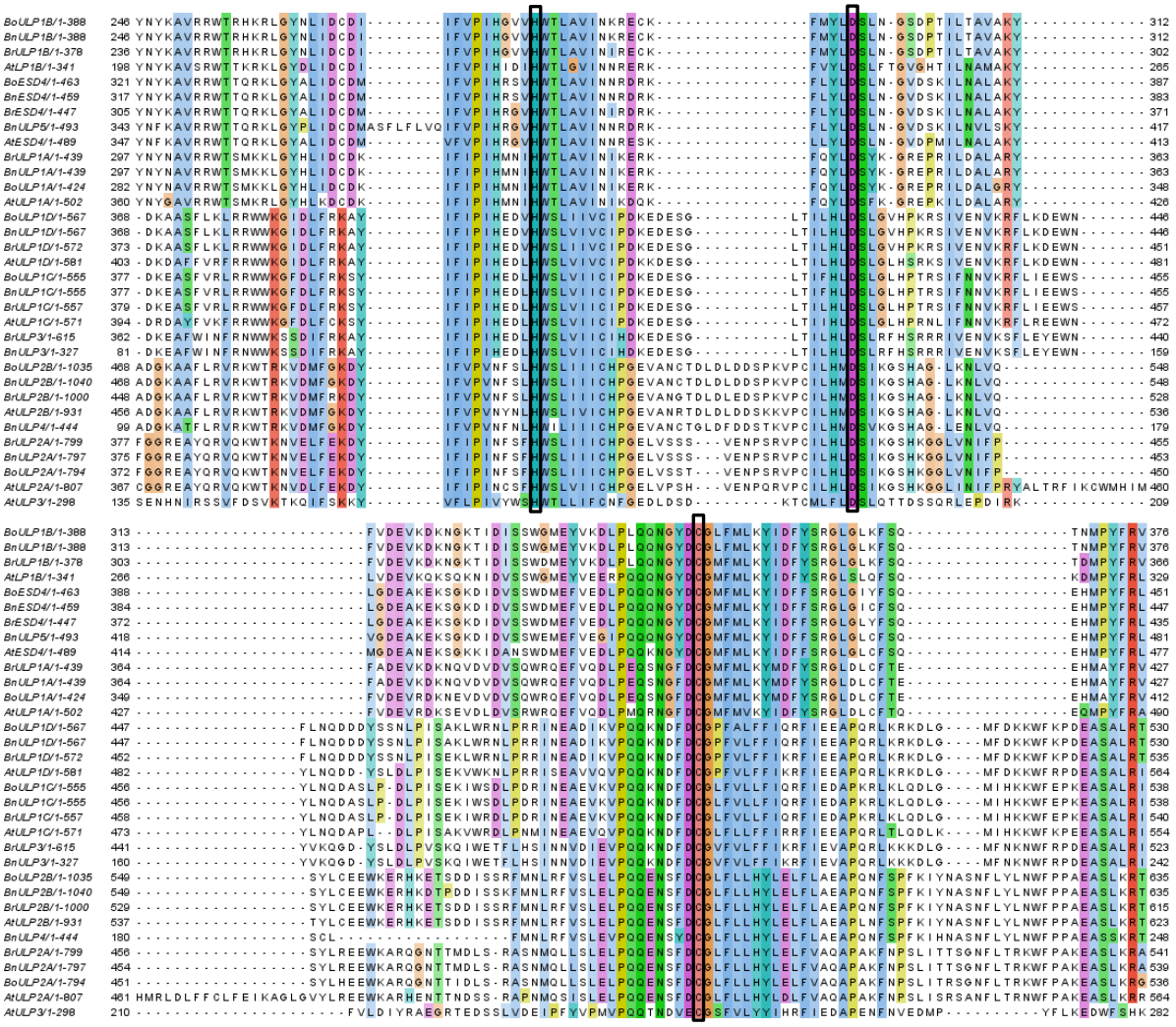

Figure 4: Alignment of Brassica ULP sequences. *The H-D-C catalytic triad characteristic of the ULP SUMO proteases is marked in black. The areas surrounding the key amino acid residues show strong conservation across species. Sequences were obtained using the NCBI-BLAST tool, aligned using ClustalX and visualised in Jalview. Jalview's ClustalX-based colour scheme, which colour-codes amino acids by residue type, was applied. The intensity of the colouring reflects the degree of amino acid conservation.*

Phylogenetically, the Brassica ULPs sort into 3 branches: ESD4-ULP1A-ULP1B (ESD4 group), ULP1C-ULP1D (OTS group) and ULP2A-ULP2B group (ULP2 group) (Figure 5). Accession codes for the proteins used can be found in supplementary table 3.

Only *B. napus* showed a slight increase in the number of ULPs. As this crop was bred for the properties of its seed, I decided to further investigate our hypothesis by focussing on cereal crops, all of which have undergone selection for seed-related traits.

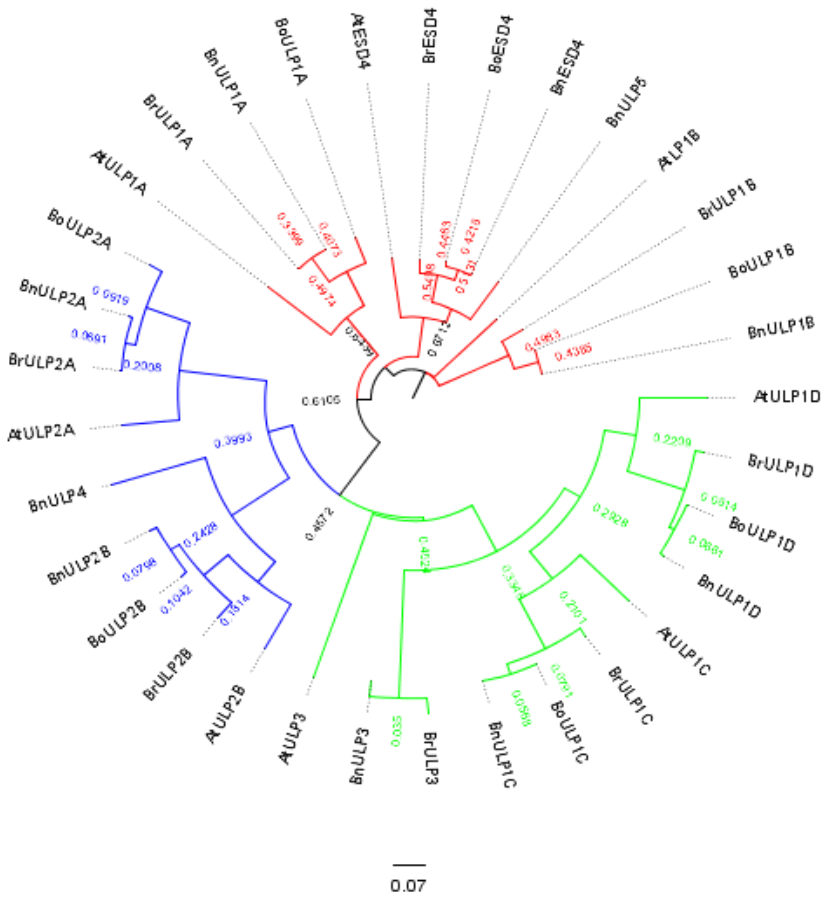

Figure 5: Phylogenetic tree of Arabidopsis thaliana and crop Brassicas putative ULP sequences. *The sequences cluster into three groups. The closely conserved ESD4 group incorporates ESD4, ULP1A and ULP1B homologs and is coloured in red. The OTS group incorporates OTS1 and OTS2 homologs and is coloured in green. The ULP2 group incorporates ULP2a and ULP2b homologs*

3.2.2 ULP SUMO proteases in the cereal family

While the differences in ULP number in the Brassica family are only subtle, those in the cereal family are clearer. Crop plants *Oryza sativa* and *Zea mays* encode more than double the number of ULP sequences found in *B. dystachon* (Table 2).

Table 2: ULP SUMO protease sequences retrieved in Brachypodium dystachon and crop cereals. *A number of crops encode more putative ULPs than B. dystachon, with Oryza sativa and Zea mays encoding more than double. Due to the underdevelopment of proteome data, very few sequences were recovered from Hordeum vulgare and Triticum aestivum.*

Organism	Number of ULP sequences retrieved
Brachypodium dystachon	**10**
Oryza sativa	22
Zea mays	21
Sorghum bicolor	13
Hordeum vulgare	3
Triticum aestivum	7

It is tempting to hypothesize that the reduced increase in ULP sequences found in *Sorghum bicolor* is due to a lack of resolution of the sorghum proteome compared to the highly resolved rice and maize proteomes. This is supported by the fact that only 3 and 7 ULP protein sequences were found in barley and wheat respective, both of whose proteome data is still underdeveloped.

Phylogenetic analysis of the sequences recovered shows partial conservation of the OTS and ESD4 groups (Figure 6), but a number of novel groups also emerge, as was previously suggested (Yates et al. 2016). Meanwhile, the ULP2 group is less clearly defined in cereals. Accession codes for the proteins used can be found in supplementary table 3.

Overall, an increase in the number of ULP family members was observed in crops bred for their seed. Within the cereal family, there is an increase in ULP numbers in cultivated crops in comparison with to the non-crop grass *Brachypodium dystachon*. The following section discusses a number of possible causes of the disproportionately intricate deSUMOylation system in cultivated cereals.

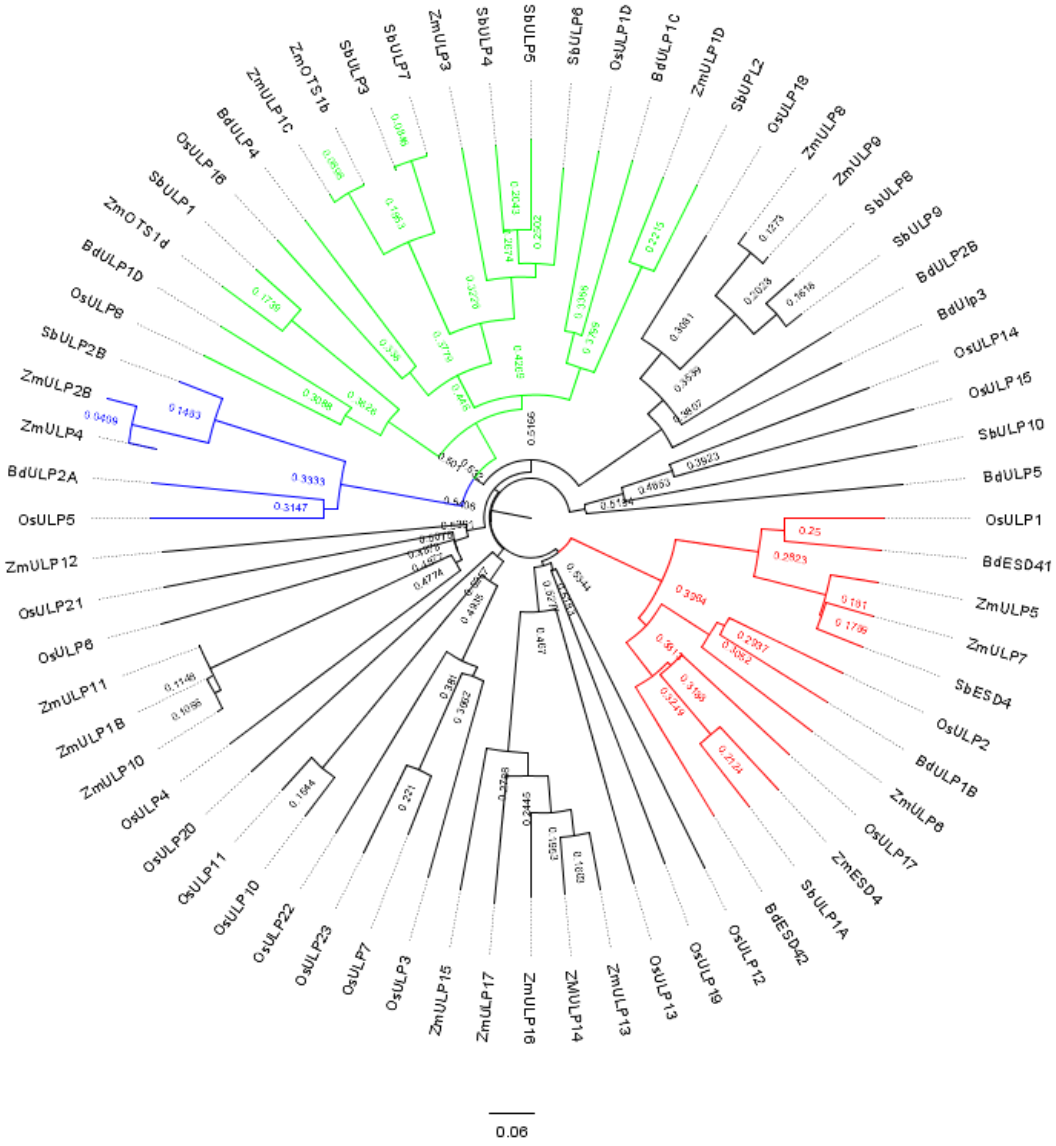

Figure 6: Phylogenetic tree of Brachypodium dystachon and crop cereals putative ULP sequences. *The grouping of sequences is less clear in cereals in comparison to Brassicas. This is due in part to the increase in number of sequences and the lack of characterisation of crop ULPs. However, the ESD4 group (coloured in red) and the OTS group (coloured in green) remain recognisable. The placement of BdULP2b outside the ULP2 group (coloured in blue) is likely due to the fact that only a partial sequence was recovered. However, new groups have seemingly emerged.*

3.3 DISCUSSION

3.3.1 The selection pressures applied during domestication and breeding target a number of processes which are known to involve SUMOylation.

The centuries-long process of domestication has driven the evolution of low-yielding ancestral plants into the high-yielding crops we know today. This phenotypic transformation is the result of a number of selection pressures exerted in order to make crops easier to grow and harvest and to improve the yield of the tissue destined for consumption. The main characteristics selected for in cereals have been reviewed both archaeologically and biologically by Fuller and Allaby. They include reducing shattering so as to enable effective harvesting of the seed head, and reducing the size and number of seed dispersal appendages to avoid unwanted dispersal by wind or animals. Also important are the control of germination time so as to produce seed that does not germinate in dry storage conditions but germinates promptly and strongly when sown in order to outcompete weeds; increasing seed size and therefore yield, and adapting to changes in environment as the crop spread to areas by developing ways of coping with new (a)biotic stresses without compromising greatly on yield (Fuller & Allaby 2009). More fundamentally, the plant must reliably produce viable seed, produce enough vegetative growth to sustain the development of many and/or large seeds. A number of these aspects have already been linked to (de)SUMOylation, notably fecundity, germination time, seed size and stress tolerance.

3.3.1.1 Fertility

SUMOylation is involved in various aspects of fecundity, from the ability to form viable embryos to correct flower development and preventing early flowering.

SUMOylation is known to be essential for embryogenesis: homozygous *Arabidopsis* mutants for either both the canonical SUMOs (SUMO1 and SUMO2), the E1 activating subunit SAE2 or the E2 conjugating enzyme SCE1 are non-viable, aborting in early embryogenesis (Saracco et al. 2007). Additionally, mutants lacking the two currently known E3 ligases SIZ1 and HPY2 were also non-viable (Ishida et al. 2012) . This phenotype coincides with high levels of SUMO1 and/or SUMO2 expression in various parts of developing flowers, seeds and embryo in wild-type plants as measured by GUS-staining (van den Burg et al. 2010).

Less dramatic but still highly impactful are the single *siz1* and *hpy2* mutants, which are both strongly dwarfed. *hpy2* mutants often do not survive bolting and a considerable proportion of the seeds they do generate are aborted (Ishida et al. 2012). siz1 mutants do generally bolt and

produce seed. However, they also show and increased number of aborted seeds (Ling et al. 2012)

Not only defective SUMOylation, but also defective deSUMOylation lie at the cause of the dwarfed phenotype. *esd4* mutants are severely dwarfed and early flowering with a reduced number of flowers and amount of pollen (Murtas et al. 2003; Reeves 2002; Villajuana-Bonequi et al. 2014). *ots1/ots2* mutants share the first two of these characteristics and exhibit reduced seed numbers in all but the ideal growth conditions (Conti et al. 2008; Campanaro et al. 2016). Fertility is also severely affected in *spf1* mutants, with less than half of seeds developing normally and in the *spf1/2* double mutant, in which only 15% of seeds complete full normal development (Liu et al. 2017).

One of the causes of the lack of optimal seed development in SUMO mutants could be malformations in floral organ development. For example, *ots1ots2* double mutants exhibit a reduction in stamen elongation which, when rescued by crossing in a *della* mutation, also restores the otherwise reduced seed formation rate back to wild type levels (Campanaro et al. 2016).

Conversely, some *spf1/2* double mutant flowers exhibit increased style length, causing a physical pollination barrier. Interestingly, this phenotype is only present in approximately one third of flowers. These double mutants also produced a smaller proportion of viable pollen grains and showed slower pollen tube growth (Liu et al. 2017).

Additionally, *siz1* mutants also show disrupted guidance of the pollen tube, again reducing the chances of successful fertilisation (Ling et al. 2012). As protein and/or transcript levels of almost all characterised ULP SUMO proteases (the northern blot performed in *esd4* mutants is inconclusive) are known to be elevated in developing flowers (Hermkes et al. 2011; Kong et al. 2017; Castro et al. 2016; Murtas et al. 2003), the ULP mutants may be harbouring undiscovered floral phenotypes. For example, *esd4* mutants exhibit deformed and irregularly placed siliques (Reeves 2002) reminiscent of *sum1-1 amiR-SUM2 SUM1/2* knockdown plants (van den Burg et al. 2010), which may be caused by defective flower or seed formation earlier in development.

Complementarily, high levels of both SUMO conjugates and SUMOylation elements SUMO1, SAE1 and SCE1 were also observed in flowers (Saracco et al. 2007), emphasising the critical importance of (de)SUMOylation in floral development.

3.3.1.2 Seed size

The end target of cereal breeding of yield is to maximise the amount of seed endosperm produced. Transcriptome analysis of a range of maize tissues showed transcriptional upregulation of the two canonical SUMOs, the maize SAE, a number of SCEs and E3 ligases and all but 2 of the 8 investigated ULPs in the endosperm (Augustine et al. 2016). Interestingly, the same study found an increase in the transcript levels of SUMO-v, a non-conjugatable SUMO analogue presumably acting through SIM-domains in endospermal tissue. The enriched endospermal SUMO system could provide an explanation for the increase in ULP number observed in *B. napus*, a crop bred for its seed, in comparison to other Brassica cultivars.

The difference in the extent of ULP family expansion may be due to the different pathways of endosperm formation and fate observed in monocot grasses and dicot Brassicas (Olsen 2004). However, very few analyses of the endospermal SUMO system have been published. Furthermore, the data available is limited to transcriptome analysis. Further research into both the transcriptional and (post)translational level of the endospermal SUMO system in a wider variety of plant species is needed to solidify this hypothesis.

3.3.1.3 Germination time

For farmers to be able to differentiate clearly between crop and weed seedlings and for the crop seedling to outcompete their weed competitors, crop seeds must have a narrow germination time window and germinate quickly. Both the lack of SUMOylation in the *siz1* mutant and the lack of deSUMOylation in the *ots1ots2* double mutant exhibits a late germination phenotype (S. Il Kim et al. 2016; Castro et al. 2016) while SUMO overexpression leads to abscisic acid hyposensitivity (Lois et al. 2003). Additionally, the data presented in chapter 4 show that overexpression of SUMO1 to high levels also leads to a delay in germination. The disparate causes of the delayed germination phenotype suggest that SUMOylation must be tightly regulated for the germination process to occur correctly.

3.3.1.4 Stress tolerance

In the field, crops are grown close together and must be able to withstand the associated abiotic stresses. Moreover, they may have lost defensive toxins in the breeding process, increasing their reliance on other defensive mechanisms to respond to the biotic stresses they encounter in the field. SUMO has long been established as a strong player in the plant stress system.

A variety of abiotic stresses are known to cause the accumulation of SUMO conjugates, including oxidative, salt, osmotic and temperature stress. These conjugates disappear as the plant is given time to recover (Catala et al. 2007; Conti et al. 2008; Kurepa et al. 2003). Meanwhile, *esd4*, *ots1/2,* and *spf1* mutants exhibit higher levels of SUMO conjugation in non-stressed conditions (Hermkes et al. 2011; Conti et al. 2008; Liu et al. 2017) and the ability to recover from stress induced SUMO conjugate accumulation was shown to be compromised *spf1* mutants (Kong et al. 2017). Phenotypically, *ots1/2* mutants are more sensitive to salt stress. Interestingly, while their root growth was more affected by osmotic stress, they did not show increased water loss (Castro et al. 2016; Conti et al. 2008). The salinity tolerance of the *esd4* mutant has not yet been studied, but as it is known to be hypersensitive to abscisic acid (Miura & Hasegawa 2010), it may also exhibit a salinity phenotype. The mutant phenotypes stand in contrast with the OTS1 overexpressor phenotype, was shown in both rice and *Arabidopsis* to be more salt-tolerant and accumulate fewer SUMO conjugates when exposed to stress (Conti et al. 2008; Srivastava, Zhang, Yates, et al. 2016).

SUMO also plays a role in the biotic stress system. Both the *ots1/2* and the *siz1* mutant exhibit increased levels of salicylic acid (Bailey et al. 2015; Jin et al. 2008). Salicylic acid is involved in the plant defence against biotrophic pathogens though the mechanism of programmed cell death. While this may confer resistance to *Pseudomonas syringae* to the *ots1/2* (Bailey et al. 2015), it is not necessarily a desirable characteristic. Constitutive hyperaccumulation of salicylic acid pushes the jasmonic acid-salicylic acid antagonism in one direction, leaving the plant unable to adapt to defending against necrotrophic pathogens. Indeed, plants with partially inhibited SUMOylation such as the *siz1* mutant showed higher susceptibility to the necrotrophic pathogens *Botrytis cinerea* and *Plectosphaerella cucumerina,* while SUMO1 overexpression led to increased resistance against these pathogens (Castaño-Miquel et al. 2017).

In addition to SUMO, most of these characteristics share a connection to gibberellin (GA) signalling. The GA-abscisic acid equilibrium mediates germination, the aleurone being the key seed layer in the perception of this equilibrium (Jacobsen & Beach 1985; Xie et al. 2006) The DELLA proteins, which are degraded in response to GA, are known SUMOylation targets and play a role in flower development (Campanaro et al. 2016), growth repression in abiotic stress situations (Conti et al. 2008), shade avoidance and in the equilibrium between growth and defence against biotic stresses mediated by GA and jasmonic acid (Pieterse et al. 2014). All this information leads us to hypothesize that the key to uncovering the importance of SUMOylation in crops may lie in the identification of SUMO targets in GA-related proteins.

3.3.2 A number of cereal-specific SUMO system components provide a basis for a more complex SUMO system in cereals

Characterisation of the SUMO system in cereals has led to the discovery of a number of components not present in dicots. Firstly, a peptide similar to di-SUMO was identified and found to be expressed at low levels in the maize female gametophyte (Srilunchang et al. 2010). This same peptide was later discovered to also be present in the male gametophyte (Augustine et al. 2016).

Secondly, cereal family have been found to contain a new subclass of SCE proteins. Active site modelling revealed an increased proportion of negatively charged amino acids around the active site, suggesting they may exhibit altered specificity (Augustine et al. 2016). However, neither the number of class I nor class II SCEs is consistently more elevated in cereal crops than in *B. dystachon*. (Table 3)

Table 3: Number of crop SCE proteins found by Augustine et al. (Augustine et al. 2016) *Cereals have a higher number of (putative) SCE proteins than Arabidopsis thaliana. Specifically, class II SCEs are only found in cereals. However, crop cereals do not consistently encode more SCEs than Brachypodium dystachon.*

Organism	**Class I**	**Class II**
Arabidopsis thaliana	1	0
Brachypodium dystachon	2	1
Oryza sativa	2	1
Zea mays	4	3
Sorghum bicolor	2	3

3.3.3 DeSUMOylation as mechanism for specificity in the SUMO system

Even when the increased number of SCE enzymes in cereals are taken into account, the ULP SUMO proteases outnumber the SUMOylating enzymes (Tables 1, 2 and 3). This raises an interesting question with regard to specificity within the SUMO system. Specificity may be imparted by deSUMOylation rather than by SUMOylation, especially as the ULP SUMO proteases are unlikely to be the only class of SUMO proteases in plants; Two other classes of SUMO proteases have previously been identified in mammals (Hickey et al. 2012).

As the post-translational modification process is conserved in eukaryotes, they may also be present in plants. Further investigation of these classes in plants could turn the ubiquitin-based model of specificity in posttranslational peptide tags upside down.

In order to investigate whether the ULP SUMO proteases are indeed a viable candidate for specificity within the SUMO system, this study profiles the transcriptional changes in the *ULP* genes (and later the genes encoding both the SUMO paralogs and SUMOylation enzymes) in response to different parameters, including abiotic stress and constitutive SUMO overexpression. The aim of this profiling was to establish whether the genes showed individually specific regulation patterns in response to these factors, which may be indicative of a role as a specificity mechanism.

As mentioned above, SUMOylation is strongly linked to the plant stress response, including the response to abiotic stress (Novatchkova et al. 2004). As plant exposure to stress generally leads to the accumulation of SUMO conjugates, I was keen to investigate whether the changes in gene transcription associated with the accumulation of SUMO conjugates though constitutive SUMO overexpression mimicked those in plants subjected to abiotic stress. Two lines overexpressing SUMO1 in the Col-0 background were therefore generated profiled phenotypically.

4 PHENOTYPIC CHARACTERISATION OF SUMO OVEREXPRESSOR LINES

4.1 INTRODUCTION

A large amount of the characterisation of SUMO system components has taken place through mutant analysis. However, as *Arabidopsis* SUMO1 and SUMO2 mostly act redundantly and the *sum1sum2* mutation is lethal (Saracco et al. 2007), mutant genotypes could not be used to assess the impact of SUMO levels on transcription of the ULP SUMO proteases and other SUMO system components.

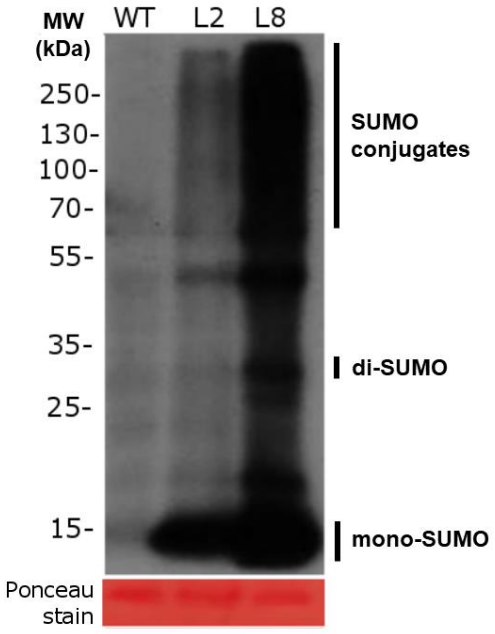

Figure 7: Expression profile of overexpressor lines used in this report. *A rabbit-α-STREP-probed Western Blot shows the characteristic SUMO profile at a lower level in L2 and a high level in L8.*

This study therefore made use of two independent SUMO overexpression (SOX) lines, L2 (low OX) and L8 (high OX) previously generated in the lab. Both these lines were generated by Dr. Moumita Srivastava using a using a CaMV35::Strep-SUMO construct expressing mature SUMO1.

Western Blotting was used to compare the SUMO levels in the overexpressor lines to Col-0 WT plants, all of which were grown on unsupplemented MS medium. In an experiment performed by Rebecca Morrell, Both L2 and L8 showed large amounts of the α-STREP-probed tagged peptide of approximately 15 kDa, the expected size of mono-SUMO (Figure 7). Both also showed a high-molecular weight smear reminiscent of that observed when stressed wild-type plants are probed for SUMO conjugates (Kurepa et al. 2003), with the smear being substantially more intense in the L8 sample, indicating conjugation of tagged SUMO1 to SUMOylation targets.

SUMO overexpressing *Arabidopsis* lines were previously studied by Lois et al. (Lois et al. 2003), who reported a higher tolerance to high levels (10 µM) of the phytohormone abscisic acid (ABA) in plants overexpressing SUMO1 and SUMO2. The limited scope of available data raised

the question of whether the SUMO-mediated attenuation of ABA-mediated growth regulation held true at lower ABA concentrations. The following section addresses this question and assesses the effect of SUMO overexpression on seed germination, as this process is known to involve ABA signalling.

4.2 RESULTS

4.2.1 SUMO overexpression increases growth in response to ABA

The SUMO-mediated reduction in ABA-mediated root growth inhibition was previously reported by Lois et al. (Lois et al. 2003). However, extremely high concentrations of ABA were used in the study (10 and 50 µM ABA). In order to establish whether this effect is conserved at lower ABA levels, the experiment was repeated at 1µM ABA. Supplementation of MS agar medium with 1 µM ABA was previously shown not to affect root length, lying in between the low ABA concentrations which enhance root growth and the high ABA concentrations which inhibit it (Ghassemian et al. 2000).

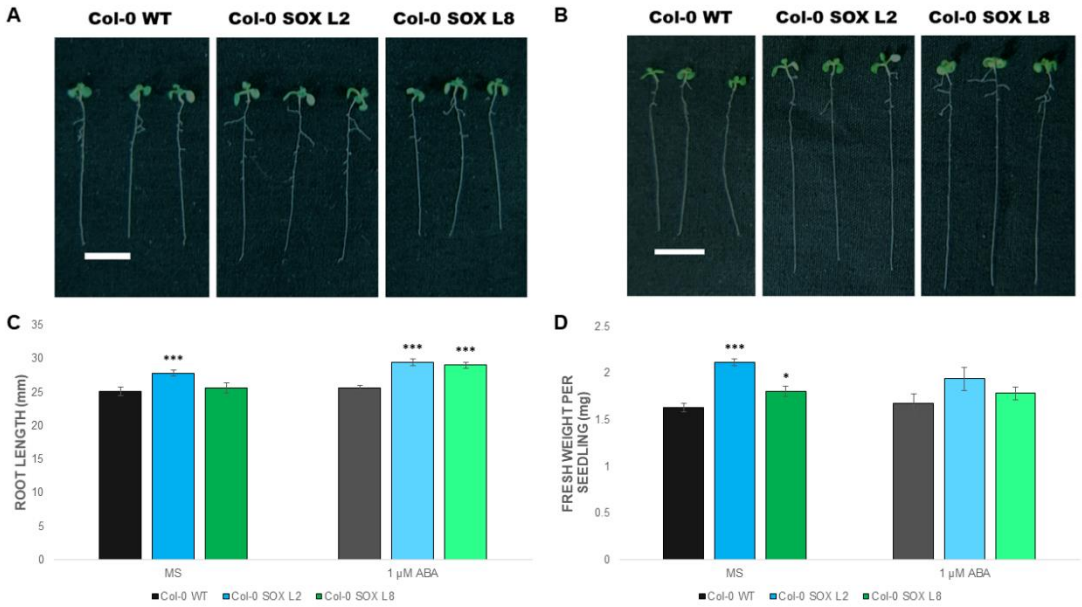

Figure 8: Root length is affected by ABA and SUMO expression. *A: Phenotypic appearance of 10 day-old seedlings grown on MS medium. B. Phenotypic appearance of 10 day-old seedlings grown on MS medium supplemented with 1 µM ABA. C: Increase in root length in 10-day-old seedlings after 6 days of treatment. D: Biomass at 10 days following 6 days of exposure to treatment. Error bars represent standard error. Significance was assessed using Student's t-test for root length assay and Mann-Whitney U-test for fresh weight assay. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor.*

In non-ABA-supplemented MS medium, Col-0 SOX L8 show no difference in root length with respect to Col-0 WT plants. However, Col-0 SOX L2 seedlings did exhibit longer roots (Figure 8).

After 1µM ABA treatment, Col-0 WT seedlings showed no change in root length when compared to Col-0 WT plants grown in control conditions. Meanwhile, roots of SUMO overexpressing plants were significantly longer than those of Col-0 WT plants. Although the level of SUMO overexpression differs between Col-0 SOX L2 and L8, there was no significant difference in root length between them. This implies a greater increase in root length in response to 1 µM ABA treatment in Col-0 SOX L8 plants than in Col-0 SOX L2.

In order to establish whether this phenotype affected the growth of other plant tissues, a fresh weight assay was performed. In MS conditions, both Col-0 SOX L2 and L8 showed increased biomass with respect to Col-0 WT plants. However, these differences were not significant in 1 µM ABA conditions. While there was little difference in the means between conditions, ABA treatment led to a greater variance between results.

4.2.2 SUMO conjugate accumulation delays germination

A link between SUMO and germination has been previously reported. Interestingly, both a lack of SUMOylation through SIZ1 and lack of deSUMOylation through OTS1/2 (Miura et al. 2009; Castro et al. 2016) cause delayed germination. I therefore assessed the germination rate of the Col-0 SOX lines (Figure 9). As the *ots1ots2* double mutant also shows an increase in SUMO conjugates but its germination phenotype has not yet been quantified, it was also included in the experiment. Col-0 SOX L2, which accumulates large amounts mono-SUMO but has only moderately increased SUMO conjugate levels, does not exhibit a significant delay in germination. However, Col-0 SOX L8 does show a delayed germination phenotype, with a germination rate of approximately 60% at the first time point, when over 80% of Col-0 WT seeds had germinated. At the second time point, the delay in germination observed in Col-0 SOX L8 was only significant to $p<0.1$.

In general, germination was not as delayed in Col-0 SOX L8 as in the *ots1/2* double mutant, indicating the process involved maybe dependent on deSUMOylation by OTS1 or OTS2 specifically.

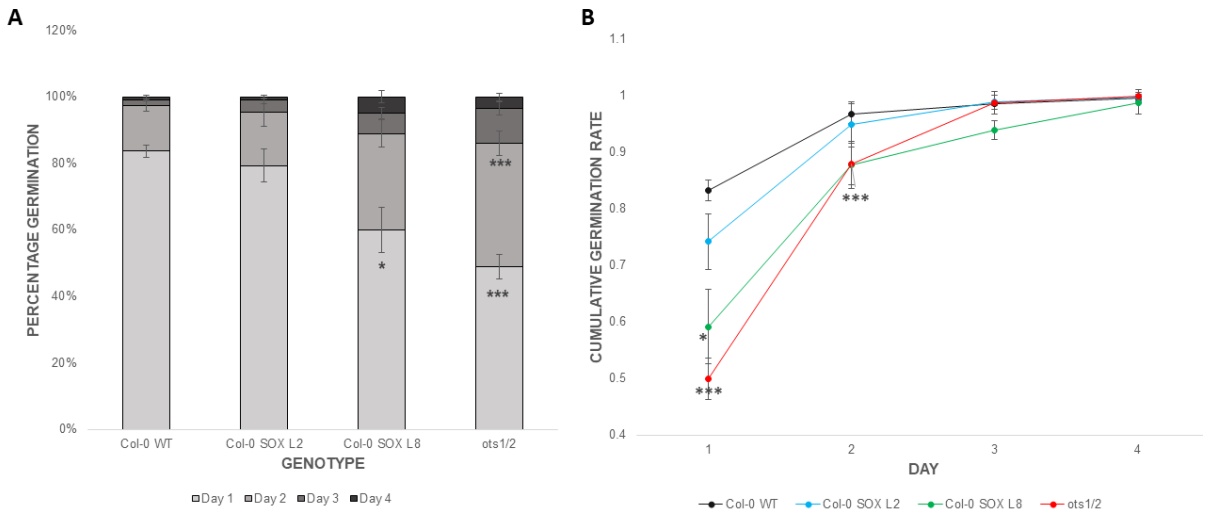

Figure 9: High SUMO conjugate levels delay germination. *Daily germination chart (A) and germination curve (B) of germination on MS medium. Col-0 SOX L2 plants show no significant delay in germination, while Col-0 SOX L8 plants do. However, germination is not retarded to the level observed in ots1/2 double mutant. Error bars represent standard error. Significance was assessed using Student's t-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

4.3 Discussion

At first glance, the two phenotypic response exhibited by the Col-0 SOX lines may seem contradictory: the increased root length would suggest a decreased ABA response while the delayed germinated would indicate an increased ABA response. However, more careful analysis potentially provides a more plausible response.

4.3.1 SUMO1 overexpression may cause abscisic acid insensitivity

In the case of root length, ABA treatment is known to increase root length in concentrations below 1 µM and decrease it in concentrations above 1 µM in a dose-dependent manner (Ghassemian et al. 2000). In our experiments, SOX lines showed an increase in root length, which would imply a decreased ability to sense external ABA. This hypothesis is in accordance with results obtained in high ABA concentrations (10 and 50 µM) (Lois et al. 2003) where plants overexpressing SUMO showed a smaller decrease in root length when compared to Col-0 plants grown in the same conditions.

Little is currently known about the interaction between SUMO and the ABA signalling pathway. As a result, the molecular mechanism behind this ABA hyposensitivity remains a topic of speculation. ABA perception in plant has been shown to occur though complexes of PYRABACTIN RESISTANCE 1 (PYR1) and the PYR1-LIKE PROTEIN (PYL) superfamily and the PROTEIN PHOSPHATASE 2C (PP2C) phosphatases (Yu et al. 2016; Raghavendra et al. 2010).

Members of clade A of the PP2CA family in particular have been identified as regulators of the ABA response (Schweighofer et al. 2004). Using the SUMO and SIM site prediction software GPS-SUMO (Zhao et al. 2014) to screen the sequences of PYR1 and the PYLs for SIM sites and/or SUMO sites, searches revealed at least one of these was predicted to present in PYR1 and all PYLs except PYL9 when stringency was set to the medium setting, sites were predicted in PYR1, PYL3-6, PYL8 and PYL10-12 when the most stringent setting was used. Meanwhile, all clade A PP2Cs have predicted SIMs and/or SUMO sites that are visible at the highest stringency setting.

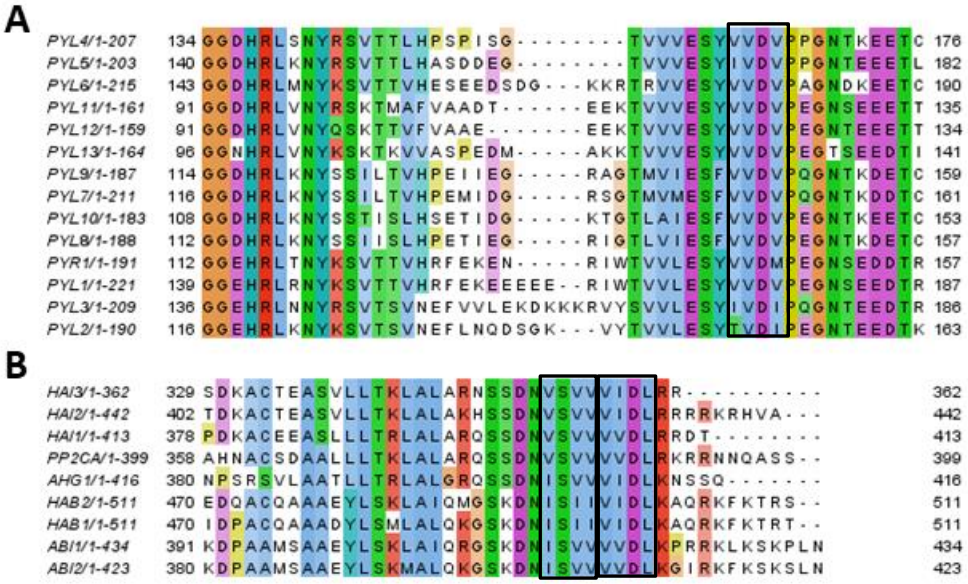

Figure 10: The PYR1/PYL and PP2CA families both exhibit potential SIM sites*. An alignment of PYL (A) and clade A PP2C (B) sequences shows a high degree of conservation of the putative SIM sites, marked in black. Sequences were recovered from the TAIR database, aligned using ClustalX and visualised in Jalview. Jalview's ClustalX-based colour scheme, which colour-codes amino acids by residue type, was applied. The intensity of the colouring reflects the degree of amino acid conservation.*

When the prediction data was compared to alignments of the PYR/PYL and the clade A PP2Cs, both families were found to exhibit a conserved hydrophobic patch at their respective C-termini. These take the shape of consensus SIM motifs (Merrill et al. 2010), with the PP2C proteins showing two adjacent SIM motifs with inverse orientations (Figure 10). It is tempting to speculate that if either of both of these proteins interact with SUMO, it may impact their ability to act as an ABA receptor. In the case of the PP2C HYPERSENSITIVE TO ABA 1 (HAB1), R505, an amino acid important for interaction with its target SNF1-RELATED PROTEIN KINASE 6.2 (SnRK6.2) (Soon et al. 2012), is directly adjacent to the second putative SIM site.

The larger increase in root length after 1µM ABA treatment observed in Col-0 SOX L8 when compared to Col-0 SOX L2 could be the result of a greater inhibition of ABA perception. This hypothesis could be tested by repeating the experiments at lower and higher ABA concentrations. However, a better profiling of root length responses to low (0-0.1 µM) ABA concentrations would be needed to establish the concentration at which point the maximum root length increase occurs. Below this concentration, Col-0 SOX L8 plants should exhibit a smaller root length increase than Col-0 SOX L2. At high (10-50 µM) ABA concentrations, Col-0 SOX L8 plants should show less root growth inhibition than Col-0 SOX L2 plants.

The increased SUMO levels in Col-0 SOX L2 and L8 could also play a role downstream of ABA perception: for example, the ABA-related transcription factor ABI5 is known to be SUMOylated (Miura et al. 2009). *abi5* mutants complemented with a non-SUMOylated form of ABI5 showed a stronger decrease in root length grown on both unsupplemented MS medium and in response to ABA than Col-0 WT. This opens up the possibility that increased levels of SUMOylated ABI5 could lead to an increase in root length.

However, any effect of SUMO overexpression is likely to be a balancing act between different growth mediators. For instance, SUMOylation of the DELLA proteins represses growth rather than increasing it (Conti et al. 2014). This may explain why Col-0 SOX L8 plants do not show the same increase in growth observed in Col-0 SOX L2 plants when grown in MS medium. Another possible explanation relies on the delayed germination phenotype observed in Col-0 SOX L8: as all seedlings were transferred for treatment and assessed on the same day, Col-0 SOX L8 seedlings are *de facto* younger and may therefore show a delay in development.

4.3.2 Increased ABA concentrations in plants accumulating SUMO conjugates may lead to delayed germination.

The process from pollinated flower to germinating seed is an intricate one, involving extensive remodelling of the cell types present in the seed throughout its development. After pollination, the seed and its embryo develop through the morphogenesis and maturation phases. If the prevailing conditions at the time when the seed is fully formed are not optimal for germination, the seed can undergo desiccation and become dormant (Locascio et al. 2014). When environmental conditions become more favourable to growth, water is reabsorbed in a multi-step process termed imbibition. This rehydration is the first step towards the breaking of dormancy and the start of germination. Germination itself is a complex process regulated by balance between the phytohormones GA and ABA, but also by light, temperature and nitrogenous compounds (Nambara et al. 2010).

SUMOylation is also known to play a part in the breaking of dormancy. While dormant seeds accumulate SUMO conjugates, these disappear upon stratification, when *Arabidopsis* seeds are exposed to moisture and undergo cold treatment in order to stimulate germination. Additionally, mass spectrometry analysis found a total of 53 putative SUMOylation targets present in seeds. However only 3 targets were present in both dormant and non-dormant seeds, suggesting that the breaking of dormancy entails an extensive reprogramming through changes in SUMOylation (Castaño Miquel 2015).

During seed development, ABA is both imported from maternal tissues and synthesised endogenously, the latter of which induces dormancy (Karssen et al. 1983). After during and after imbibition, ABA levels drop sharply. Indeed, imbibed seeds must maintain *de novo* ABA biosynthesis if they are to remain dormant. This extended dormancy might ensue if seeds encounter stresses such as ABA treatment, temperature stress, salinity or osmotic stress (Joosen et al. 2012).

Seeds which accumulate lower amounts of ABA have been shown to exhibit precocious germination, while those overaccumulating ABA show enhanced dormancy (Nambara et al. 2010). As Col-0 SOX lines were previously hypothesised to produce ABA insensitive seedlings, it is likely that the delay in germination in Col-0 SOX L8 plants is due to an increase in ABA levels rather than to an increased sensitivity to ABA. Indeed, rice OsOTS1 knockdown lines, which exhibit delayed germination, were shown to accumulate increased ABA levels (Srivastava et al. 2017).

Additionally, ABA accumulation is known to increase drought tolerance. The observation that Col-0 SOX L8 plants are more drought tolerant than Col-0 WT and Col-0 SOX L2 plants, though slightly less drought tolerant than *ots1/2* plants (Garrido & Orosa, personal communication) would perfectly reflect the pattern of germination observed in these genotypes. In the following chapter this hypothesis is expanded upon using transcriptional data gathered to identify transcriptional differences which could explain the striking difference in germination rate between Col-0 SOX L2 and L8.

5 TRANSCRIPTIONAL ANALYSIS OF *ARABIDOPSIS* ULP SUMO PROTEASES

5.1 INTRODUCTION

Very little is currently known about the *Arabidopsis* ULP SUMO proteases, especially in relation to their targets. Initially, the *Arabidopsis* system for deSUMOylation by ULPs was modelled on the yeast system. This consists of two proteins, ULP1 and ULP2 (Li & Hochstrasser 1999; Li & Hochstrasser 2000), with distinct cellular localisations (Schwienhorst et al. 2000; Panse et al. 2002; Sydorskyy et al. 2010) and putative functions and only one SUMO isoform, SMT3 (Johnson et al. 1997).

The *ulp1* mutation is lethal, once again showing the importance of SUMOylation *in vivo* (Li & Hochstrasser 1999). The current model for yeast SUMOylation is that tasks of maturation and deconjugation are predominantly split between the proteases: ULP1 acting as a maturase and ULP2 as a deconjugase (Melchior et al. 2003). This presumption is currently also common in *Arabidopsis* (Hermkes et al. 2011; Chosed et al. 2006), but a combination of the data gathered in this study and data collated from more recent literature suggest this may not be the case.

Twelve candidate *Arabidopsis* SUMO proteases were named systematically based on their homology with yeast ULP1 or ULP2 (Kurepa et al. 2003). This list was later refined to the seven most likely candidates - those which exhibited the highest to homology to known SUMO proteases (Novatchkova et al. 2004). These candidates, grouped by putative yeast homologue, are shown in table 4. The accession numbers of these 7 canonical SUMO proteases can be found in supplementary table 2.

Table 4: Comparative ULP phylogeny. *While only two ULP SUMO proteases have been identified in Saccharomyces cerevisiae, Arabidopsis thaliana encodes seven canonical SUMO proteases. The Arabidopsis ULP SUMO proteases were named systematically based on their phylogenetic relatedness to their yeast homologues.*

Yeast SUMO protease	***Arabidopsis* ULP SUMO protease**
ULP1	ESD4
	ULP1A/ELS1
	ULP1C/OTS2
	ULP1D/OTS1
ULP2	ULP2A/SPF2
	ULP2B/ASP1/SPF1

Of the putative *Arabidopsis* SUMO proteases, six have already been characterised to some extent: EARLY IN SHORT DAYS 4 (ESD4), ESD4-LIKE SUMO PROTEASE 1/UBIQUITIN-LIKE PROTEASE 1A (ELS1/ULP1A), OVERLY TOLERANT TO SALT 2/UBIQUITIN-LIKE PROTEASE 1C (OTS2/ULP1C), OVERLY TOLERANT TO SALT 1/UBIQUITIN-LIKE PROTEASE 1D (OTS1/ULP1D), SUMO PROTEASE RELATED TO FERTILITY 2/UBIQUITIN-LIKE PROTEASE 2A (SPF2/ULP2A) and SUMO PROTEASE RELATED TO FERTILITY 1/ ARABIDOPSIS SUMO PROTEASE 1/UBIQUITIN-LIKE PROTEASE 2B (SPF1/ASP1/ULP2B). ULP1B remains to be characterised. All cysteine proteases, the ULPs have a conserved catalytic domain featuring a highly conserved H-D-C catalytic triad. Their mutant phenotypes are only rescued though complementation when the active cysteine is intact. An alignment of the most conserved region of the 7 ULP sequences can be seen in figure 11.

Figure 11: Alignment of the catalytic regions of Arabidopsis ULP SUMO proteases. *The sequences show both areas of homology and variation. The intensity of the colouring reflects the degree of amino acid conservation. The H-D-C catalytic triad is strongly conserved across all sequences. Sequences of the seven putative ULP SUMO proteases were retrieved from the TAIR database, aligned using ClustalX and visualised in Jalview. Jalview's ClustalX-based colour scheme, which colour-codes amino acids by residue type, was applied.*

As described in chapter 3, the ULPs sort into three branches phylogenetically: the ESD4 branch (red), the OTS branch (green), and the ULP2 branch (blue) (Figure 12). Interestingly, the OTS branch is more closely related to the ULP2 branch than to its fellow ULP1s in the ESD4 group.

The Ulp1s (ESD4, ELS1, ULP1B, OTS1, OTS2) were proven to process SUMO1 and SUMO2 to mature SUMO1/2 to various extents *in vitro* (Chosed et al. 2006). The Ulp2s (SPF1, SPF2) were proven to mature (i.e. to process SUMO from its immature to its mature state by cleaving the protein to expose the C-terminal diglycine motif, see Figure 1C) SUMO1, but not SUMO2 or SUMO3 (Liu et al. 2017). As SUMO1 and SUMO2 are often considered to be interchangeable due to their high degree of homology, this differentiation is remarkable.

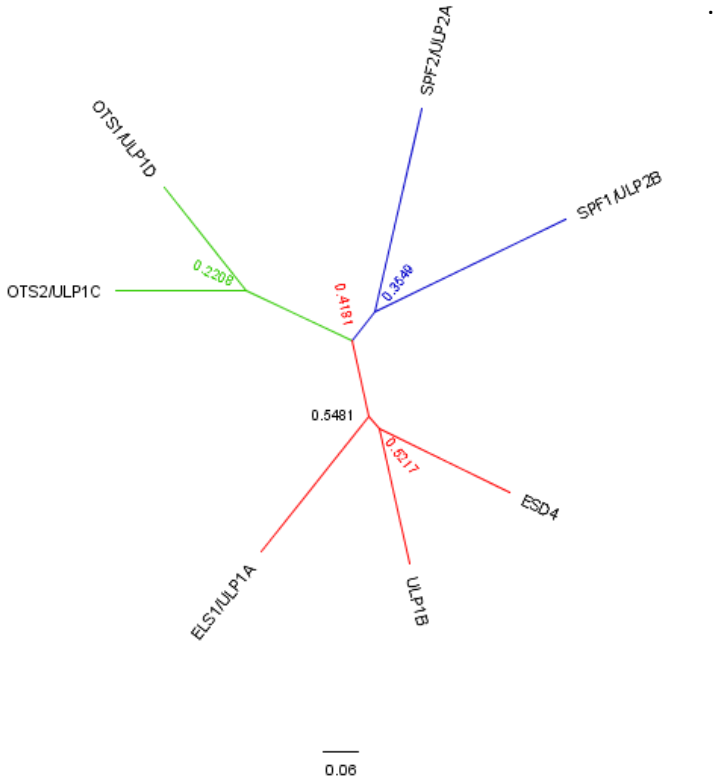

Figure 12: Phylogenetic tree of the Arabidopsis ULP SUMO proteases. *Three branches are visible: the ESD4 branch (red), the OTS branch (green) and the ULP2 branch (blue). The phylogenetic tree was generated based on an alignment made in ClustalX and was visualised in Figtree.*

Only ELS1 was able to process SUMO3-HA but not SUMO3-Gly-Gly-X_{18} to mature SUMO3 (Chosed et al. 2006), highlighting the limitations of *in vitro* maturation assays. As expected, maturase activity was abolished when the conserved SUMO di-glycine motif is mutated to di-alanine (Murtas et al. 2003; Hermkes et al. 2011; Budhiraja et al. 2009). Interestingly, SPF1 was observed to interact with SUMO3 even though it showed no SUMO3 maturase activity (Liu et al. 2017). No ULPs have yet been shown to mature SUMO5. However, as both SUMO3 and

SUMO5 have been shown to form conjugates, it is possible that they are matured by a different class of proteases. Indeed, HsSUMO4, which is not processed by the human ULP family of proteases due to the Q90P amino acid substitution (Owerbach et al. 2005), is matured by an as yet unidentified hydrolase in stressed cells (Wei et al. 2008). As SUMO3 and SUMO5 exhibit similar substitutions in this position (to methionine and leucine respectively), they may also be subject to maturation by non-ULP SUMO proteases.

In vitro deconjugation assays also possess certain limitations. OTS1, OTS2 and ESD4 were clearly shown to deconjugate SUMO1 and SUMO2 conjugates, but not SUMO3 conjugates of the yeast protein ScPCNA *in vitro* (Colby et al. 2006), However, when GST-SUMO1/2 fused to mammalian RAN GTPASE-ACTIVATING PROTEIN (RanGAP) was used as a substrate, only ELS1 and ESD4 produced clear deconjugation patterns, while OTS1/2 showed only a minor degree of deconjugation. None of the proteases were able to deconjugate GST-SUMO3/5-mRanGAP (Chosed et al. 2006).

The following section provides an overview of the 6 currently characterised *Arabidopsis* SUMO proteases through the knock-out mutants which have been isolated for each of them.

5.1.1 ESD4

EARLY IN SHORT DAYS 4 (ESD4) was the first ULP SUMO protease to be characterised in *Arabidopsis.* The *esd4* mutant exhibits reduced size and an early flowering phenotype under short day conditions, with siliques appearing deformed and at irregular positions (Reeves 2002). At a molecular level, they accumulate increased levels of SUMO conjugates and possess lower levels of free SUMO1 compared to *Landsberg erecta* wild-type plants (Murtas et al. 2003).

Overexpression of the mature and immature and immature forms of SUMO1, SUMO2 and SUMO3 enhanced the *esd4* mutant phenotype, leading to the hypothesis that *in vivo,* ESD4 mainly functions in the deconjugation of SUMO from its substrates rather than in SUMO maturation (Murtas et al. 2003). *ESD4* was later shown to complement the *ulp2* mutation in yeast (Hermkes et al. 2011). Ulp2 is thought to be the yeast SUMO deconjugase. Lack of Ulp2 function also leads to the accumulation of high molecular weight SUMO conjugates in yeast cells (Bylebyl et al. 2003).

Fluorescence assays showed that ESD4 localises to the nuclear periphery and the nuclear envelope (Xu et al. 2007; Murtas et al. 2003), roughly in accordance with the nucleoplasm localisation of ULP2 in yeast (Li & Hochstrasser 2000; Panse et al. 2002) (Figure 8 and 9).

esd4 mutants are also known to accumulate increased levels of the phytohormone salicylic acid (SA). Interestingly, their SUMO conjugate accumulation seems to correlate with the levels of free SA present in the plant rather than with total SA content. The *esd4* phenotype is partially alleviated by the additional mutation of the salicylic acid biosynthesis gene *ISOCHORISMATE SYNTHASE 1* (*ICS1)*. *esd4 sid2* mutants are larger and flower later than *esd4* mutants, and accumulate less SA. They also accumulate fewer SUMO conjugates, with levels falling back to wild-type or slightly above wild type depending on the background (Villajuana-Bonequi et al. 2014). This last observation implies that the increase in SUMO conjugates visible in *esd4* mutants may be caused in part by an increase in SUMOylation rather than by a decrease in deSUMOylation.

While the inactivation of ICS1 reduces the levels of SUMO conjugates in *esd4* mutants, it does not increase the levels of free SUMO1/2 (Villajuana-Bonequi et al. 2014) as would be expected in the case of increased deconjugation. However, the exact relationship between SA and (free) SUMO is still unknown as both a variety of SUMO-related mutants, including *esd4, ots1/2*, *siz1* and *sum1-amiR SUM2,* and SUMO overexpressor lines show hallmarks of an increased SA response (Villajuana-Bonequi et al. 2014; Bailey et al. 2015; van den Burg et al. 2010).

5.1.2 ELS1

Unlike *esd4* mutants, EDS4-LIKE SUMO PROTEASE (*els1)* mutants -also known as *ulp1a*- do not exhibit an eye-catching phenotype: their growth is only mildly restricted, some but not all plants have thinner stems and there is no significant difference in flowering time comparted to wild type plants. When *ELS1* expression was studied using GUS, it was mainly detected in in root and vascular tissues (Hermkes et al. 2011). RT-PCR showed *ELS1* RNA to be more prevalent in flower tissue than in rosette leaves, cauline leaves or siliques (Hermkes et al. 2011).

The observations that the loss of yeast *ulp1* (the hypothesised SUMO maturase) can be rescued through ELS1 expression and that *els1* mutants show only slightly increased accumulation of high molecular weight SUMO conjugates led to the hypothesis that ELS1 is more likely to be involved in SUMO maturation than in deconjugation (Hermkes et al. 2011).

However, ELS1 and ESD4 showed equal SUMO1 and SUMO2 cleavage proficiency (Chosed et al. 2006), detracting from this hypothesis.

At the cellular level, ELS1 localises to the cytoplasm, which is surprising as the yeast homolog ULP1 naturally locates to the nuclear envelope (Li & Hochstrasser 2003), specifically to nuclear pore complexes (Panse et al. 2002) (Fig 13 and 14). However, modified ULP1 targeted to the cytoplasm was still able to complement the otherwise lethal *ulp1* mutation, indicating that the protein is still functional in the cytoplasm. In contrast, relocalisation to the nucleolus did not rescue lethality (Panse et al. 2002).

In yeast, the *ulp2* mutation can repair the lethality of the *ulp1* mutation. *Ulp1 ulp2* double mutants grow in a wider variety of environments accumulate fewer conjugates than either of the single mutants (Li & Hochstrasser 2000). This degree of interaction is not observed in *Arabidopsis eds4 els1* mutants. While the *els1* mutation does increase the viability of one of the germ lines in the *esd4* background, the *esd4* phenotype is conserved and levels of SUMO conjugate accumulation are unaltered. This implies that the SUMO protease system in *Arabidopsis* is more complex than the yeast model.

5.1.3 OTS1 – OTS2

OVERLY TOLERANT TO SALT (OTS) -1 and -2, also known as ULP1d and ULP1c, respectively, were first studied for their role in salt tolerance. These proteins are highly similar and exhibit a considerable degree of functional redundancy (Conti et al. 2008). As a result, *ots1* and *ots2* single mutants generally do not show an obvious phenotype under normal growth conditions. The only single mutant phenotype currently reported is increased drought tolerance in *ots1* single mutants, while *ots2* mutants exhibit the same level of phenotypic drought tolerance as Col-0 controls (Castro 2013). This drought tolerance phenotype is also observed in rice OTS1-RNAi lines (Srivastava et al. 2017).

The *ots1 ots2* double mutant however, displays a range of phenotypical alterations including early flowering under short day conditions, late germination, an increased relative reduction in root and leaf growth during prolonged salt stress and osmotic stress, increased drought tolerance increased stomatal aperture in adult plants, reduced stamen elongation and consequent reduced fertility, spontaneous lesions and increased resistance to the plant pathogen *Pseudomonas syringae* (Conti et al. 2008; Campanaro et al. 2016; Castro et al. 2016; Bailey et al. 2015; Conti et al. 2009; Castro 2013). The pleiotropic phenotype of *ots1 ots2* mutants suggest that the OTS proteases are involved in multiple processes.

At a molecular level, mutants permanently show the protein signature of stressed Col-0 WT plant: increased accumulation of SUMO conjugates, DELLA proteins and of the plant hormone salicylic acid. (Conti et al. 2008; Castro et al. 2016; Conti et al. 2014). OTS1 protein is known to be degraded in the presence of salt and has been implicated in the SUMOylation of the DELLA proteins, thus influencing plant growth (Bailey et al. 2015; Campanaro et al. 2016).

Both proteases have a similar expression pattern as observed through GUS staining: OTS1 and OTS2 are present mostly in the root tissue from the early developmental stages, with high expression levels in both the root and shoot vasculature of seedlings. OTS1 expression is generally more widespread, while OTS2 expression is more local to the petioles. In mature plants, the proteases were predominantly present in developing flowers, with high levels of OTS2 expression in filaments. Both OTS proteins are also observed at wounding sites (Castro et al. 2016). Within cells, both OTS1 and OTS2 localise to the nucleus, but while OTS1 is found throughout the nucleoplasm, OTS2 accumulates in speckle-like bodies (Conti et al. 2008) (Fig 14).

5.1.4 SPF1-SPF2

SUMO PROTEASE RELATED TO FERTILITY SPF1 (also named ASP1 and ULP2b) and SPF2 (ULP2a) are the most recent ULPs to be characterised. Both ULP2s, they function partially redundantly and are thought to be involved in the regulation of embryo development. While *spf2* mutants exhibit no clear phenotype, *spf1* mutants have more elongated leaves, flower late, have short siliques and show abnormal development of almost half of their seeds. These seeds either failed to develop or were aborted at various stages of embryogenesis (Castro 2013; Kong et al. 2017; Liu et al. 2017). Around a third of *spf1* flowers also exhibit increased style length, which poses a physical fertility barrier. The infertility is asymmetric: it is more pronounced when *spf1* is used as the female parent in a backcross, further implicating SPF1 in the development of the female reproductive organs (Liu et al. 2017).

Meanwhile, *spf1spf2* double mutants exhibited higher chlorophyll and anthocyanin levels in leaves (Castro 2013), even shorter siliques than *spf1* mutants and abnormal development of more than half of seeds. Additionally, pollen grains of double mutants exhibited growth retardation even though pollen tubes developed normally in both single mutants (Liu et al. 2017).

At the molecular level, neither single nor double mutants show increased levels of SUMO conjugates in seedlings (Liu et al. 2017). However, *spf1* seedlings accumulated more SUMO

conjugates and less free SUMO, and recovered more slowly in comparison to wild-type seedlings when exposed to heat stress (Kong et al. 2017). When other tissues were studied, it was observed that *spf1* and *spf1spf2* mutants do accumulate an increased number of SUMO conjugates in mature leaves and inflorescences (Castro 2013; Liu et al. 2017). When probed with α-SUMO3 antibody, *spf1* and *spf1spf2* mutants also exhibited a slightly modified SUMOylation pattern (Castro 2013). This is especially interesting as SPF1 and SPF2 were shown not to cleave SUMO3, but were able to interact with it in a Y2H assay (Liu et al. 2017) and *SPF1* expression colocalises with *SUMO3* expression in hydathodes (Kong et al. 2017; van den Burg et al. 2010).

GUS assays showed *SPF1* expression is ubiquitous in 2- and 4 day-old seedlings. In older seedlings, it is present in newly developing leaves, the hydathodes of developed leaves and the tips of both the primary and lateral roots. It also localises to embryo sacs, inflorescences, anthers and developing seeds (Kong et al. 2017; Liu et al. 2017). Meanwhile, SPF2 was more widespread in leaves, with a slightly higher presence in the vasculature, and localised to inflorescences and maternal floral tissues. Tissue-specific PCR revealed that *spf1* transcription is highest in inflorescences and cauline leaves, with intermediate expression levels in stems and rosette leaves. SPF2 transcription was seen to be at its highest in stems, cauline leaves, rosette leaves and middle-length siliques. Interestingly, no expression of SPF2 was detected in root tissue (Liu et al. 2017). Within the cell, both SPF1 and SPF2 localise to the nucleoplasm (Liu et al. 2017).

Figures 13 and 14 illustrate the differences in subcellular localisation of SUMO proteases in *Arabidopsis* and yeast cells. As discussed in chapter 3, deSUMOylation by the ULP SUMO proteases is emerging as a candidate mechanism for specificity within the SUMO system. These disparate subcellular localisations provide further evidence for this hypothesis.

Previous experiments have shown that although most stress-induced SUMOylation is nuclear, a smaller amount of SUMOylation also occurs outside the nucleus (Saracco et al. 2007). As the only ULP to be expressed in the cytosol (Hermkes et al. 2011), ELS1 is likely to mediate the deSUMOylation of extranuclear SUMO conjugates. Meanwhile, the nuclear ULPs also show differential distribution throughout the nucleus, further indicating that they are likely to perform specific tasks.

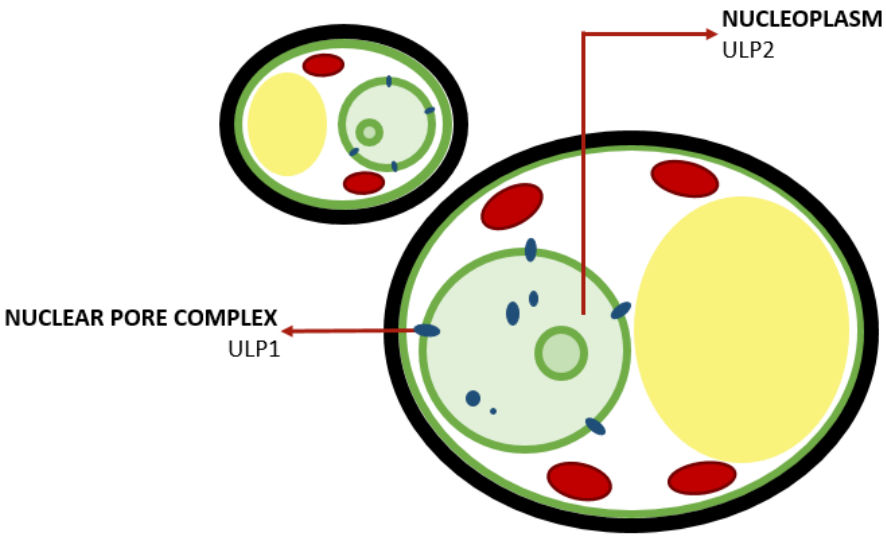

Figure 13: Illustration of the subcellular localisations of ULP SMT3 proteases in a yeast cell

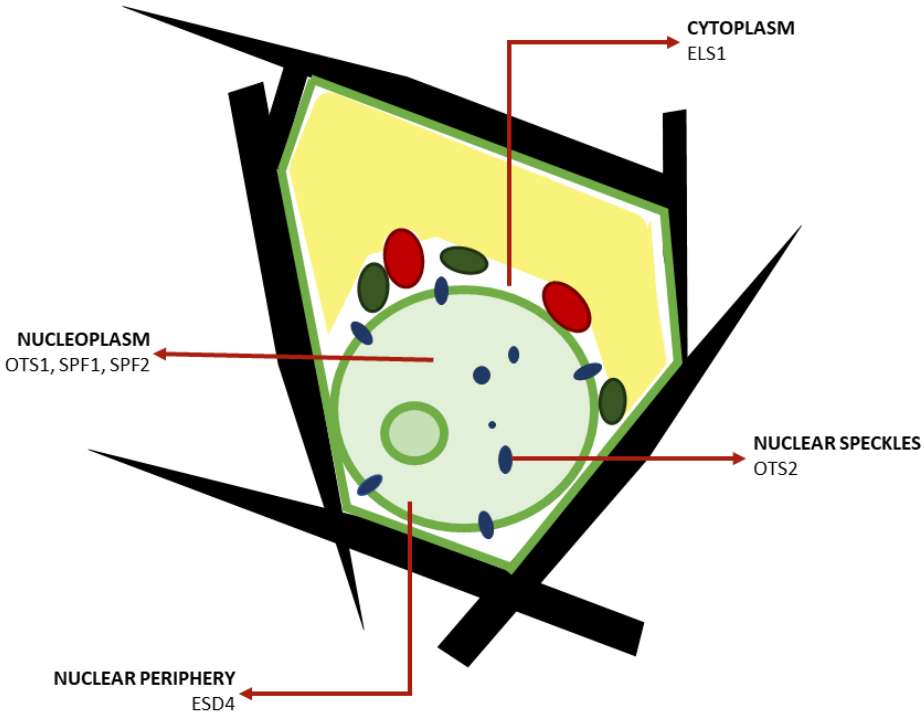

Figure 14: Illustration of the subcellular localisations of SUMO proteases in an Arabidopsis cell.

5.2 RESULTS

5.2.1 Differential transcription of ULP genes in mature plant tissues

RT-qPCR was used to compare transcript levels between individual SUMO proteases in a number of mature plant tissues, all of which were of the Col-0 accession. The results are shown in Figure 15. Figure 15A shows the results expressed relative to each other using $-\Delta C_T$ as a metric. Though this metric can express whether the abundance of a particular transcript is higher or lower in one tissue when compared to another or act as a gauge to assess which of two transcripts is more abundant in a given tissue, numerical differences should not be interpreted directly. Although primers were subjected to efficiency selection in order to minimise the error due to differing amplification efficiency, this factor should still be considered when comparing the abundance of different transcripts. Figure 15B shows the expressed as transcript abundance in mature plant tissues relative to those in 10 day-old seedlings.

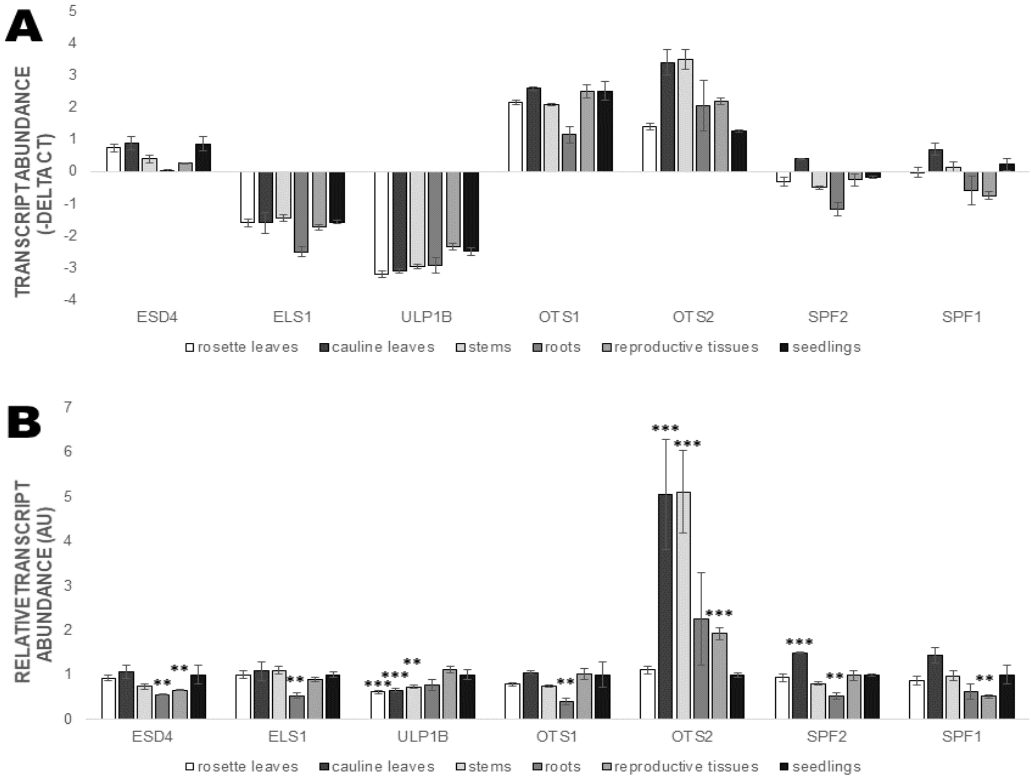

***Figure 15: Differential transcription of ULP genes in mature plant tissues**. Levels of ULP transcription vary greatly between ULP family members, but variation in transcript abundance is also present between different tissues. A: Abundance of ULP SUMO protease transcripts expressed using the -ΔCt metric. As this metric functions as a gauge rather than an accurate measure of absolute abundance, numerical differences should not be interpreted directly. B: Relative abundance of ULP SUMO protease transcripts in mature plant tissues in comparison to 10 day-old seedlings.*

*Ct values were normalised to UBC9 and CBP20 housekeeping genes. Error bars represent standard error based on three biological and two technical repeats for all tissues except the root tissue, where two biological replicates were used. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005.*

Overall, *OTS1* and *OTS2* transcripts are the most abundant, followed by *ESD4. SPF1* and *SPF2* mRNA was detected at intermediate levels, while both *ELS1* and *ULP1B* transcripts were much scarcer. For *ESD4*, the highest transcript levels are observed in leaves and seedlings, with the lowest levels observed in roots. When relative transcript abundance was calculated, only roots and reproductive tissues showed transcript levels which were significantly different to those detected in seedlings. In the case of *ELS1*, the lowest levels are again found in roots, no significant differences in transcript levels were detected in the other tested tissues.

ULP1B has the lowest transcript level of all 7 ULPs, independently of the tissue studied. Levels are slightly higher in seedlings and reproductive tissues (inflorescences, flowers and young siliques). Relative to the transcript levels detected in seedlings, those found in stems, rosette leaves and cauline leaves were significantly lower. Although the levels of transcript detected were low and the number of ESTs associated with *ULP1B* is comparable to a number of suspected pseudogenes in the family (Lois 2010), full-length *ULP1B* mRNA was recovered.

In the case of *OTS1,* the highest transcript levels were visible in cauline leaves, seedlings and reproductive tissues, with much lower levels in roots. This reduction in root *OTS1* transcript abundance was significant when compared to seedlings. For *OTS2*, the highest levels were observed in cauline leaves and stems and the lowest in rosette leaves and seedlings. Indeed, when compared to seedlings, transcript levels were significantly higher in both cauline leaves and stems, as well as in the reproductive tissues.

In both *SPF1* and *SPF2*, the highest transcript levels were found in cauline leaves. When transcript abundance was calculated relative to seedling levels, this increase was only significant for *SPF2.* The lowest transcript levels were found in roots for *SPF2* and in roots and reproductive tissues for *SPF1* both these decreases in transcript abundance were significant relative to transcript levels detected in seedlings.

It was very hard to compare the results of this experiment with previous localisation data from GUS assays. Firstly, because for most ULPs, GUS data is only available for seedlings and in some cases reproductive tissues. Secondly, the highest levels expression levels of *OTS1/2* and *SPF2* are found in the vasculature (Liu et al. 2017; Castro et al. 2016), but these high levels can be masked by the inclusion of the non-vascular organ tissue.

For example, GUS staining showed *ELS1* to be highly expressed in the roots of seedlings, mostly in developing lateral roots and the youngest section of the primary root (Hermkes et al. 2011). Meanwhile, the qPCR analysis performed here found root tissue to be the tissue showing the lowest levels of *ELS1* transcript. If the pattern observed in seedlings is conserved in mature plants, the low transcript level detected is likely to be due to the proportional shift in mature root/young root tissue in mature plants. However, even this pattern conservation need not be the true. In the case of *OTS1* and *OTS2*, true leaves in seedlings showed high levels of transcription, especially in the vasculature, while 5 week-old mature leaves showed much reduced transcription levels, with no apparent accumulation in the vasculature (Castro et al. 2016).

SPF1 and *SPF2* are the only ULPs for which the GUS assays were complemented with extensive tissue-specific PCR, albeit not qPCR. For both genes, the key result was mirrored in our analysis: high levels of *SPF1* transcript were detected in cauline leaves and low transcription of *SPF2* was detected in root tissue (Liu et al. 2017).

After uncovering differential transcription patterns in mature plants, I proceeded to establish whether expression of the ULP SUMO proteases is transcriptionally regulated by abiotic stress (abscisic acid, moderate and high salinity stress) or SUMO overexpression (using Col-0 SOX L2 and L8). RT-qPCR analysis of cDNA from 10 day-old seedlings was chosen as the experimental system. As Col-0 SOX L8 and *ots1/2* exhibit a certain degree of phenotypic overlap (see chapter 4), the *ots1/2* mutant was also included in the analysis.

If there is a considerable degree of functional equivalence between members of the ULP SUMO protease family, loss of OTS1/2 function may lead to differential transcription of other ULPs as part of a compensation mechanism. If however, they perform a specific function, this less likely to be the case. The accumulation of SUMO conjugates seen in *ots1/2* mutants as a result of defective deSUMOylation may also lead to altered transcription of the genes encoding SUMO isoforms or those encoding SUMOylation enzymes. This last hypothesis, as well as an analogous one concerning the possible transcriptional knock-on effects of the overaccumulation of SUMO conjugates in the Col-0 SOX lines, is tested in chapters 6 and 7.

5.2.2 Combined influence of SUMO overexpression and abiotic stress on ULP transcript abundance

In order to gain a general impression of SUM transcription levels in the different genotypes and stress conditions tested, all values were compared to the Ct values obtained for Col-0 WT plants grown in MS conditions. The resulting variations in transcript abundance are shown in figure 16. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005. However, variation between biological replicates was sometimes considerable and where highly significant data suggested a pattern, $p<0.1$ significance results are also described.

ESD4 transcription was lower in Col-0 SOX L8 plants in control, 1 µM ABA and 50 mM NaCl conditions, but not in the 100 mM NaCl condition. Additionally, transcript levels were decreased in Col-0 WT grown in the presence of 1 µM ABA, while the decrease observed after 50 mM NaCl treatment was only significant to $p<0.1$. Again, no significant differences were observed in plants subjected to 100 mM NaCl treatment. Surprisingly, no such downregulation was observed in Col-0 SOX L2 or *ots1/2* plants. These data seem to suggest that *ESD4* does undergo transcriptional regulation, although its trigger is unclear.

Conversely, *ELS1* transcription was only affected in plant exposed 100 mM NaCl, where it was increased in Col-0 SOX L2 and *ots 1/2*.

ULP1B showed a pattern more similar to *ESD4*, with no effects visible after 100 mM NaCl treatment, but a generalised downregulation in 1 µM ABA and 50 mM NaCl conditions, this time including *ots1/2* mutants. In 1µM ABA a decrease in transcript abundance was observed in all genotypes, with downregulation in *ots1/*2 mutants being significant to $p<0.1$ due to the large variation between samples. When plants were grown on medium supplemented with 50 mM NaCl, lower transcript levels were visible in Col-0 SOX plants, with significance for Col-0 SOX L8 at $p<0.1$, and in *ots1/2* mutants, while an increase in transcription was observed in Col-0 WT plants. However, variation between replicates was again very large. It remains tempting to speculate that moderate, but not severe stress causes a generalised downregulation of *ULP1B* transcript levels. Interestingly, more disparate effects were observed when plants were grown on unsupplemented MS medium: no significant differences were present in *ots1/2* mutants, while transcript levels were decreased in Col-0 SOX L2 and increased in Col-0 SOX L8.

The most prominent change in transcript abundance was a 15-fold decrease visible in *OTS1* when assessed in Col-0 SOX L8 in control conditions. This was reduced to a twofold decrease in plants treated with 50 mM NaCl, with no overall differences visible in 1 µM ABA or 100 mM

NaCl conditions. However, Col-0 SOX L2 did exhibit a two-fold reduction in transcription in the presence of 1 µM ABA. *OTS2* transcription was only affected in Col-0 WT plants grown in 100 mM NaCl.

SPF2 transcription was reduced in Col-0 SOX L8 plants grown in control conditions, but increased in the presence of 100 mM NaCl. In *ots1/2*, it was only affected by 1 µM ABA treatment. *SPF1* showed no overall variations in transcription.

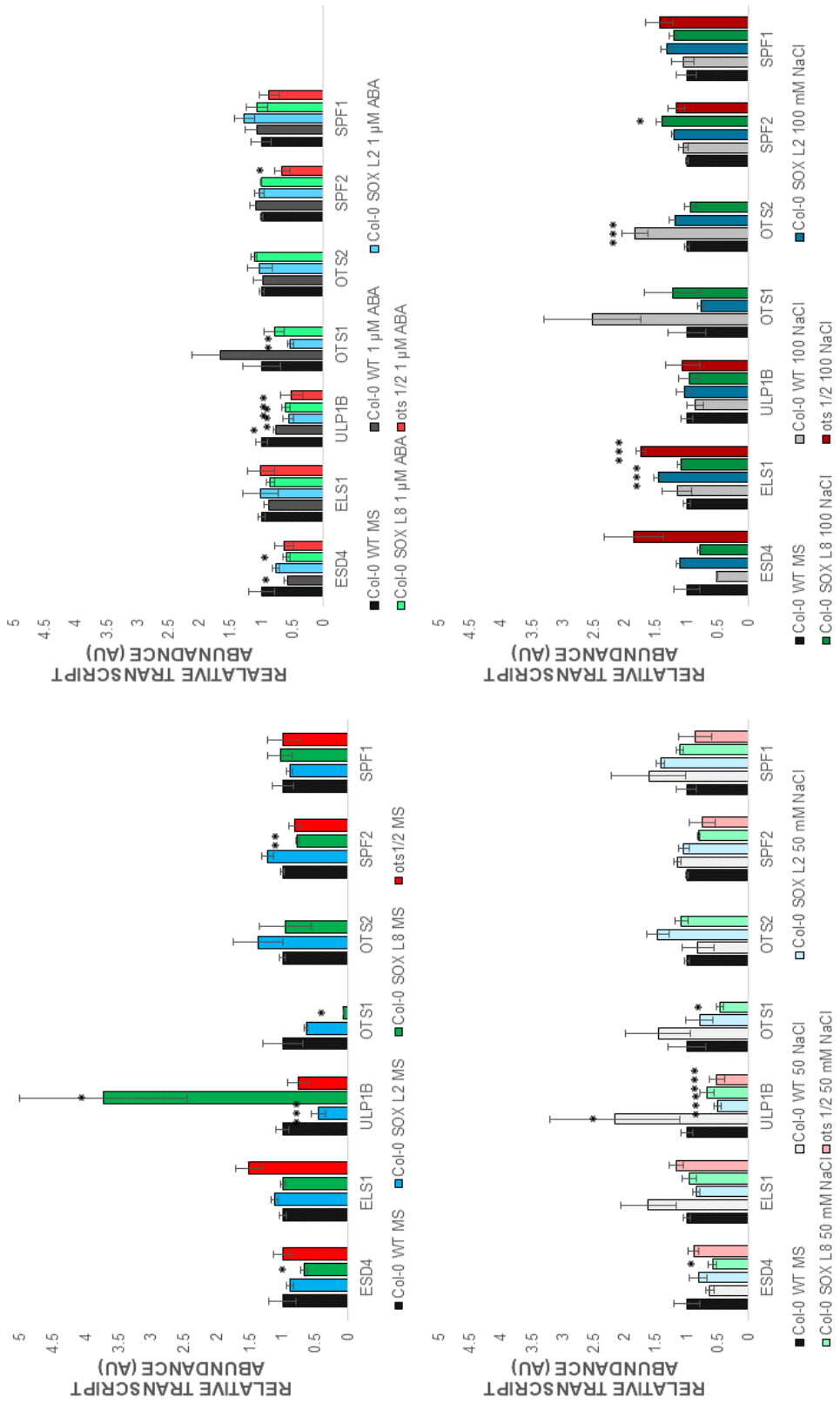

Figure 16: Combined influence of SUMO overexpression and abiotic stress on ULP transcription. *Differences in ULP transcript abundance in 10 day-old seedlings grown on A: unsupplemented MS medium B: MS medium supplemented with 1 µM ABA C: MS medium supplemented with 50 mM NaCl D: MS medium supplemented with 100 mM NaCl.*

Figure 16, continued: *All Ct values were compared to the Ct values obtained for Col-0 WT grown in MS conditions. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

5.2.3 Influence of SUMO overexpression on ULP transcript abundance

In order to assess the effect of SUMO levels on *ULP* transcription, Ct values for each growth condition were compared to values obtained for Col-0 WT plants grown in that condition. The resulting differences in transcript abundance are shown in figure 17. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005. However, variation between biological replicates was sometimes considerable and where highly significant data suggested a pattern, $p<0.1$ significance results are also described.

ESD4 transcription was lower in Col-0 SOX L8 than in Col-0 WT plants when grown in control conditions, but showed no significantly different mRNA levels in comparison to Col-0 WT plants grown in stress conditions. In *ots1/2* mutants, *ESD4* transcript levels were only affected in 100 mM NaCl conditions, where they were increased. Similarly, SUMO levels had no effect on *ELS1* transcription. Similarly, in *ots1/ots2* mutants *ELS1* transcript levels were only increased in the presence of 100 mM NaCl. Meanwhile, *ULP1B* transcript levels were lower than Col-0 WT in Col-0 SOX L2 plants grown in control and 50 mM NaCl conditions. Transcript levels were increased in Col-0 SOX L8 in comparison to Col-0 WT when grown on unsupplemented MS but not under any other condition.

OTS1 transcription was lower in Col-0 SOX L2 plants than in Col-0 WT plants in all stress conditions. Meanwhile, transcript levels were decreased in Col-0 SOX L8 in control and moderate stress conditions, but not in 100 mM NaCl. We hypothesise that the insignificance of this decrease, which just exceeded the $p<0.1$ significance levels, is due to a large variation between replicates. *OTS2* transcription was only affected in the presence of 100 mM NaCl, were it was decreased in both Col-0 SOX L2 and L8 when compared to Col-0 WT.

These data seem to suggest that high levels of SUMOylation lead to the downregulation of *OTS1.* In control conditions, this state is only achieved in Col-0 SOX L8 (Figure 17 A). However, when plants are exposed to stress conditions, this downregulation affects both Col-0 SOX L2 and L8. Remarkably, the sharp, more than 10-fold downregulation observed for Col-0 SOX L8 plants grown in control conditions is reduced in stress conditions. Here, transcript levels are 2- to 4-fold lower than those in Col-0 WT plants grown in the same conditions. Additionally, when plants were exposed to 100 mM NaCl, this downregulation also affects *OTS2*, the transcription of which is not affected in any other growth condition, indicating that the two closely related ULPs are not fully redundant.

SPF2 transcript levels were only affected in Col-0 SOX L8: they were decreased in control and 50 mM NaCl conditions, but increased after 100 mM NaCl treatment. *SPF1* transcript levels were unaffected by SUMO levels.

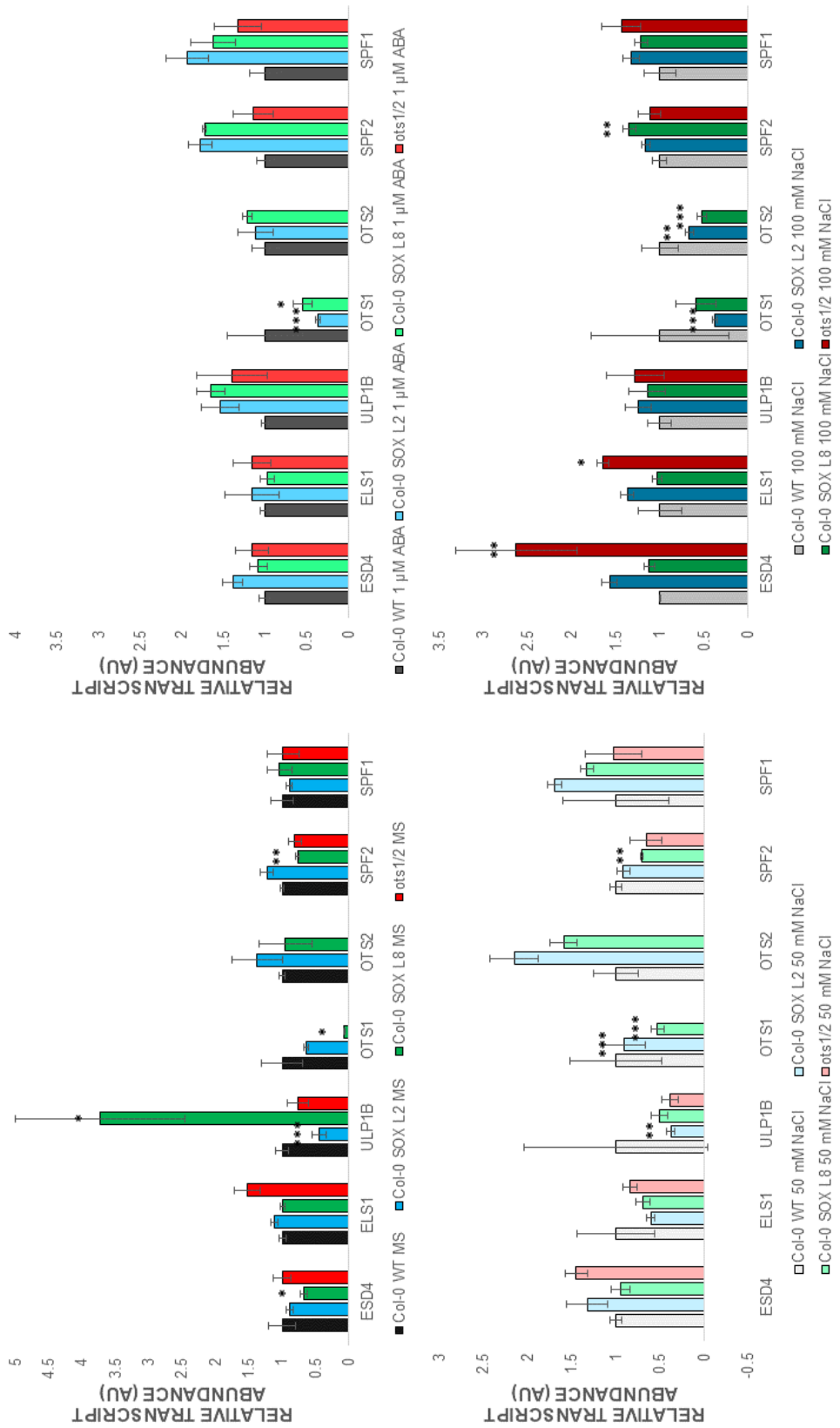

Figure 17: Influence of SUMO overexpression on ULP transcription. *Differences in ULP transcript abundance in 10 day-old seedlings grown on A: unsupplemented MS medium B: MS medium supplemented with 1 µM ABA C: MS medium supplemented with 50 mM NaCl D: MS medium supplemented with 100 mM NaCl.*

***Figure 17, continued**: Ct values for every condition were compared to the Ct values obtained for Col-0 WT grown in each respective conditions. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

5.2.4 Influence of abiotic stress on ULP transcript abundance

In order to assess the effect of abiotic stress on ULP transcription, Ct values for each growth condition were compared to values obtained for plants of that genotype grown on unsupplemented MS. The resulting changes in transcript abundance are shown in figure 18. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005. However, variation between biological replicates was sometimes considerable and where highly significant data suggested a pattern, $p<0.1$ significance results are also described.

Overall, the influence of abiotic stress on ULP transcription is limited in comparison to that of SUMO overexpression. However, some of the results described here do provide an interesting backdrop to the results discussed earlier.

In Col-0 WT plants, 1 µM ABA treatment downregulated *ESD4* and *ULP1B* transcription, while 50 mM NaCl treatment upregulated *ELS1*. 100 mM NaCl treatment increased *OTS2* transcript levels. However, both NaCl treatments showed an increase in *OTS1* transcript significant to $p<0.1$. This observation indicates that the downregulations described above in response to SUMO overexpression (Figure 17) may rather represent a lack of upregulation in response to abiotic stress, as could be the case for *OTS2* in 100 mM NaCl conditions.

In *ots1/2* mutants, the change in *SPF2* transcript level after 100 mM NaCl treatment was the only sample to show significant change in transcript abundance.

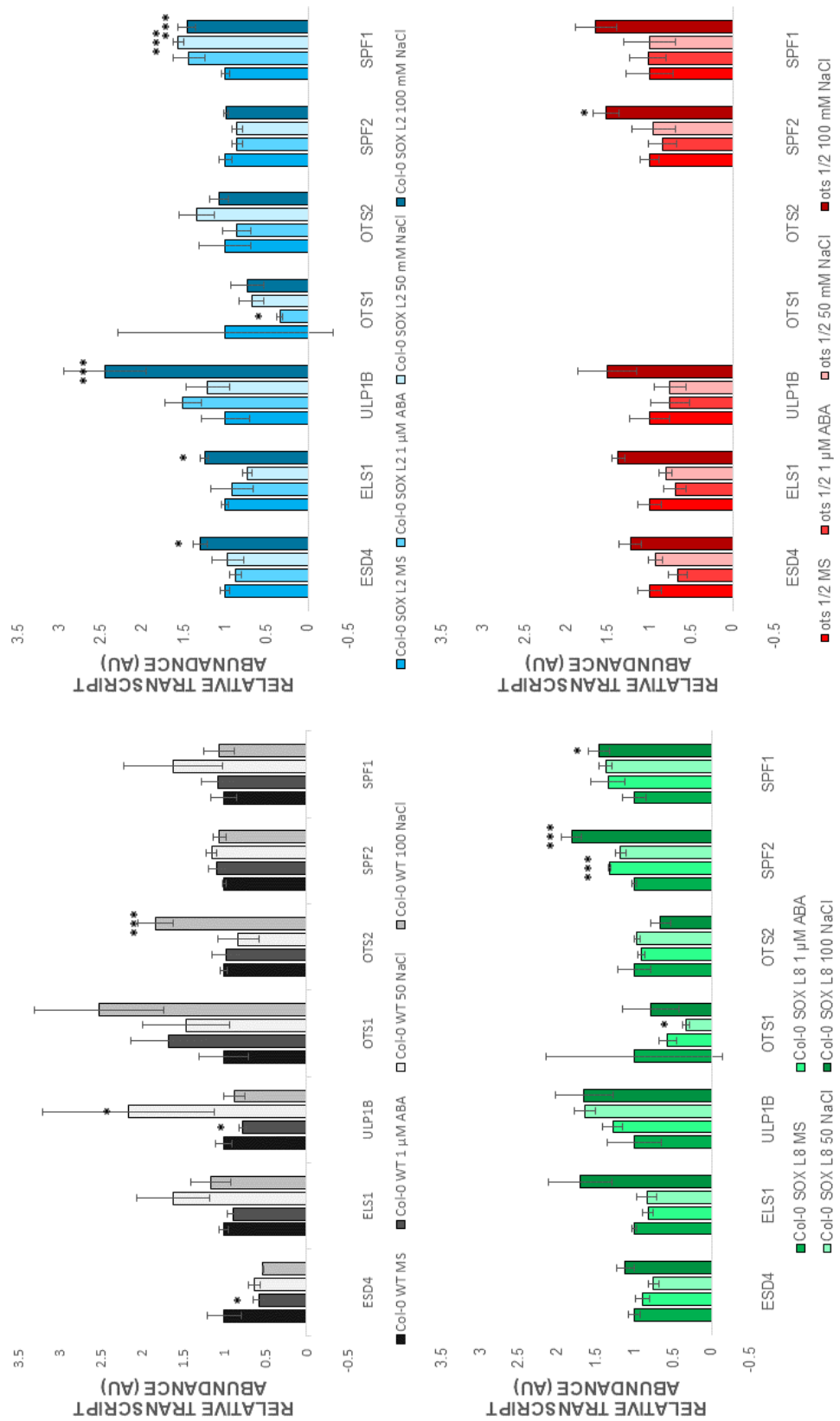

Figure 18: Influence of abiotic stress on ULP transcription. *Changes in ULP transcript abundance in 10 day-old seedlings of several genotypes after 6 days of treatment with 1 µM ABA, 50 mM NaCl and 100 mM NaCl. A: Col-0 WT, B: ots1/2 mutant, C: Col-0 SOX L2, D: Col-0 SOX L8.*

Figure 18, continued: *Ct values for each genotype were compared to the Ct values obtained for that genotype when grown on unsupplemented MS. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

In Col-0 SOX L2 plants, supplementation of the medium with 1 µM ABA only affected *OTS1* transcript levels, which it decreased. Meanwhile, 50 mM NaCl treatment only affected *SPF1* transcript abundance, which it increased. 100 mM NaCl treatment was more influential, upregulating *ESD4, ELS1, ULP1B* and *SPF1*.

In Col-0 SOX L8 plants, the presence of 1 µM ABA only affected *SPF2* transcription, which it upregulated. 50 mM NaCl treatment only affected *OTS1* transcript abundance, which was also upregulated. 100 mM NaCl treatment increased both *SPF1* and *SPF2* transcription levels.

Interestingly, the two reductions in *OTS1* transcription mentioned above (in 1µM ABA conditions for Col-0 SOX L2 and 50 mM NaCl for Col-0 SOX L2, respectively) are the only ones which remain visible when the overall effect of SUMO and abiotic stress on *ULP* transcript levels were assessed (see figure 16).

5.3 DISCUSSION

5.3.1 The members of the ULP gene family are expressed to varying levels.

While studies using the GUS reporter assay have previously shown that the ULP family members have individually specific expression patterns, data in the first part of this chapter showed that transcript levels of the ULP family members also vary greatly between individual genes. For example, there is an approximately 32-fold difference in transcript abundance between the highest levels detected (for *OTS1*) and lowest (for *ULP1B)* in Col-0 WT seedlings grown in control conditions. If a similar balance holds true at the protein level, this begs the question whether the Western blots of SUMO conjugates in seedlings commonly used to ascertain whether an *Arabidopsis* SUMO protease is likely to act as a maturase or a deconjugase provide a fair comparison.

Firstly, ELS1 and SPF2, both of which have been proposed to act as maturases due to a lack of visible SUMO conjugate profiles in their respective mutant seedlings have the two lowest transcript levels of all the confirmed ULPs. Secondly, the choice of 10-day old seedlings as a model tissue is an arbitrary one which disadvantages ULPS which are not highly expressed in seedlings, as it has previously been shown that SUMO conjugation patterns vary widely

between different plant tissues (Saracco et al. 2007). For example, 10 day-old *spf1/2* seedlings showed no increase in SUMO conjugates when compared to Col-0 WT seedlings. However, when a one month-old leaf extract was assayed, the mutant clearly accumulated more conjugates than the control (Castro 2013).

Thirdly, neither the *els1* nor the *spf2* mutant exhibited a clear phenotype. However, if either of them were the main SUMO maturase, I would expect to see a clear phenotype similar to that of SUMOylation impaired *siz1* and *hpy2* mutants. In-vitro maturation assays may therefore be a better way to discover which, if any, of the ULPs mainly take on the role of maturation. No experiment has yet assayed the 6 confirmed SUMO proteases side-by-side. In the most comprehensive comparison to date, the highest SUMO1 and SUMO2 conjugation proficiency was observed for both *ELS1* and *ESD4* (Chosed et al. 2006), while, as detailed above, *esd4* mutant seedlings accumulate SUMO conjugates and *els1* do not.

Taking all these arguments into account, it has become clear that the maturase-deconjugase dichotomy model based on the yeast system needs revising. Currently, it seems more likely that at least some ULP family members perform both functions. Alternatively, a different category of SUMO protease may assist in the maturation process.

5.3.2 Changes in OTS1 transcription provide a possible mechanism behind the Col-0 SOX L8 and *ots1/2* delayed germination phenotype

Chapter 4 described a significant delay in the germination of Col-0 SOX L8 seeds similar to that seen in *ots1/2*, while the germination of Col-0 SOX L2 seeds showed no difference to that Col-0 WT plants. In this chapter, a large decrease in *OTS1* transcription was observed in unsupplemented MS-grown Col-0 SOX L8 but not L2 plants. If this reduction also holds true at the protein level, it would provide an explanation for the way the Col-0 SOX germination rate mimics that of *ots1/2* mutants.

It is tempting to speculate that the reduction does indeed hold true at the protein level as previous data from rice indicates that various types of stress treatment (ABA, salinity, mannitol and desiccation) lead to the degradation of OsOTS1 protein (Srivastava et al. 2017). Additionally, salinity, cold and heat stress led to a reduction in *OsOTS1* transcription (Srivastava, Zhang & Sadanandom 2016). These data seem to suggest that in response to stress, OTS1 degradation and *OTS1* downregulation go hand in hand.

Additionally, it has previously been demonstrated in rice that depletion of OsOTS1 through RNAi increases ABA levels (Srivastava et al. 2017). *OsOTS1*-RNAi plants to indeed exhibit phenotypes which are generally associated with increased ABA: drought tolerance (Srivastava et al. 2017), a decrease of germination in severe stress conditions (Srivastava, Zhang & Sadanandom 2016) and delayed germination in non-stress conditions (Anjil Srivastava, personal communication).

In *Arabidopsis*, the relationship between SUMO, stress, OTS1 transcription and OTS1 protein levels has only been has been investigated in the context of salicylic acid (van den Burg et al. 2010; Bailey et al. 2015). In the light of the data presented above, I propose that future studies broaden the picture to include ABA. Indeed, a comparison of ABA levels in Col-0 WT, Col-0 SOX L2 and L8 and *ots1/2* could provide some of the evidence towards the confirmation of this hypothesis in *Arabidopsis.*

5.3.3 Transcriptional regulation of ULPs is a complex process

Generally speaking, there was no gene which was not affected transcriptionally by either one of the stress conditions or one of the genotypes. However, patterns were often hard to discern. For example, the combination of moderate stress and SUMO accumulation led to downregulation of *ULP1B* in Col-0 SOX lines and *ots1/2* mutants grown in the presence of 1 µM ABA and 50 mM NaCl (Figure 16, B-C). However, this effect could not be traced back to either a genotypic or environmental effect (Figure 17 B-C, Figure 18). The same is true for the downregulation of *ESD4* observed in Col-0 SOX L8 in 1 µM ABA and 50 mM NaCl conditions.

Even more complex is the transcriptional regulation of the *OTS* genes. When the influence of genotype alone was assessed, there was a clear downregulation of *OTS1* in Col-0 SOX lines grown in stress conditions when compared to Col-0 WT plants, which expanded to *OTS2* in severe salt stress. The downregulation of *OTS1* was also visible in Col-0 SOX L8 in control conditions (Figure 17). However, when Col-0 WT plant were exposed NaCl, both concentrations of which have been shown to increase both SUMOylation and free SUMO levels (Conti et al. 2008), *OTS1* showed an increase in transcription significant to $p<0.1$. Similarly, *OTS2* showed an increase in transcription in Col-0 WT plants grown in 100 mM NaCl (Figure 18 A). These opposing effects resulted in the loss of a net effect in the majority of cases when compared to the transcription levels found in Col-0 WT plants on unsupplemented MS medium: only the downregulations in Col-0 SOX L2 in 1 µM ABA and Col-0 SOX L8 in 50 mM NaCl were visible when the combined effect was assessed (Figure 16 B,C).

The cause of this lack of upregulation of *OTS1* in Col-0 SOX lines remains an open question. On the one hand, the differential factor between Col-0 WT and Col-0 SOX genotypes is SUMO overexpression. As the reduction in transcription was only observed in Col-0 SOX L8 plants in MS conditions, it is tempting to hypothesise that is requires either high levels of SUMO conjugates or extremely high levels of free SUMO (Figure 7). On the other hand, the increase in *OTS1* transcription in Col-0 plants occurs when these are grown in the presence of NaCl, which has previously been shown to increase both the accumulation of SUMO conjugates and of free SUMO.

Due to time constraints it was not possible to perform a Western blot analysis of the SUMOylation profiles of the various genotypes in all stress conditions. As Western blotting allows visualisation of both free SUMO and SUMO conjugate levels, it may be able to shed more light on this topic.

In addition to the transcriptional effects described above, it is likely that there are also (post)translational effects at play. While SA- and NaCl-mediated degradation of OTS1 have already been described (Conti et al. 2008; Bailey et al. 2015), the effects of stress on the stability of other ULPs has not yet been studied.

The ULPs may also themselves be SUMOylation targets. Indeed ESD4 has been confirmed as a SUMOylation target by mass spectrometry (Miller et al. 2010). Possible SUMO and SIM site predictions using GPS-SUMO (Zhao et al. 2014), showed at least one possibility for each mode of interaction in all ULPs, even at the highest stringency settings. In fact, all ULP1s exhibited a potential C-terminal SIM motif, while all family members except ESD4 and ULP1B exhibit a potential SIM motif in the centre of the protein sequence. Whether these predicted sequences encode *bona fide* SIM sites and, if so, how SUMOylation could affect ULP function are important questions which remains to be addressed.

6 TRANSCRIPTIONAL ANALYSIS OF THE *ARABIDOPSIS* SUMO PARALOGS

6.1 INTRODUCTION

After investigating the effect of stress and SUMO overexpression on ULP transcription, my attention turned to the substrates of these enzymes, the SUMO paralogues. As mentioned in the general introduction, expression of either of the dominant *SUM* genes, *SUM1* or *SUM2,* is essential for early plant development as double mutants were non-viable (Saracco et al. 2007). When a viable *sum1-amiR SUM2* SUMO knockdown line was generated, it displayed a pleiotropic phenotype which included dwarfism, early flowering, early leaf senescence and erratic silique development (van den Burg et al. 2010).

In contrast with most other eukaryotic model organisms which encode only one SUMO isoform (Flotho & Melchior 2013), the *Arabidopsis* genome contains nine *SUM* genes. Of these, four are known to be expressed: *SUM1, SUM2, SUM3* and *SUM5* (Saracco et al. 2007). *SUM9* encodes only a partial transcript and is widely accepted to be a pseudogene. The cases of *SUM4*, *SUM6*, *SUM7* and *SUM8* are more interesting: all encode a full-length transcript, but as they have no associated ESTs (Kurepa et al. 2003), they are generally deemed to have pseudogenised (Hammoudi et al. 2016), especially as SUMO4, SUMO6 and SUMO7 only retain half of the highly conserved diglycine motif necessary for conjugation (Saitoh et al. 1997) (Figure 19). Indeed, the low levels of *SUM4* transcription first detected in a microarray experiment could not be confirmed using RT-PCR or RNA gel blot analysis and were therefore considered to be due to experimental error (Saracco et al. 2007). The integrity of the diglycine motif is considered to be very important as ability to form SUMO conjugates is important for plant development: overexpression of conjugation-deficient forms of SUMO1 and SUMO2 led to a clear phenotype which was much more reminiscent of that of the *sum1-amiR SUM2* SUMO knockdown line than that of plants overexpressing conjugatable SUMO1/2 (van den Burg et al. 2010).

When sequence similarity was assessed through the construction of a neighbour-joining tree, two clades were immediately apparent. One cluster grouped the canonical SUMOs SUMO1/2/3/5, while the other grouped the non-expressed SUMO4/6/7/8. Within the first clade, “archetypal SUMOs” (Hammoudi et al. 2016) SUMO1 and SUMO2 formed a separate

group to the more variable SUMO3 and SUMO5. In this chapter, SUMO1/2/3/5, the SUMO paralogs whose expression has been widely confirmed, will be referred to as “canonical SUMOs” and the dominant SUMO1 and SUMO2 paralogs as “archetypal SUMOs”.

In the second clade, the four genes were easily split into two pair based on sequence similarity: *SUM4/6* and *SUM7/8* share 74 and 76 % sequence identity respectively are arranged in tandem on the respective chromosomes, they are thought to have emerged from tandem duplication events (Kurepa et al. 2003).

The evolutionary diversification process of *Arabidopsis* SUMO paralogs has recently been modelled (Hammoudi et al. 2016) and the resulting tree provides a far broader insight into the differences between the *Arabidopsis* SUMO paralogs. In this model, an ancient SUMO archetype is thought to have undergone triplication during the eudicot whole genome triplication event termed the ‘gamma polyploidy event’. Two forms, SUMO-γA and SUMO-γB are thought to have retained an archetypal character, possibly though subfunctionalisation or balancing of expression patterns, while the third form, termed SUMO-LIKE-ГC (SUL-γC), gave rise to a divergent family of paralogs (Hammoudi et al. 2016) encompassing homologs of the DI-SUMO-LIKE (DSUL) forms found in monocots (Augustine et al. 2016). In *Arabidopsis,* this lineage gave rise to SUMO5, which may explain why it is the most divergent of the four canonical SUMO paralogs. Indeed, phylogenetic analysis found the sequences of AtSUMO1/2/3 to be more similar to those of HsSUMO1/2/3 than to AtSUMO5 (Kurepa et al. 2003). Interestingly, very little SUMO5 sequence variation was found between *Arabidopsis* accessions, which points towards neofunctionalization of the paralog (Hammoudi et al. 2016).

After another whole genome duplication affecting *Brassicaceae,* the SUMO-γA lineage gave rise to SUMO4, which underwent tandem duplication to form the SUMO4-SUMO6 pair (Hammoudi et al. 2016). SUMO4/6 having arisen from a different lineage to the archetypal *Arabidopsis* SUMOs may explain key differences between SUMO4/6 and SUMO1/2: both SUMO4 and SUMO6 lack the diglycine motif necessary for conjugation, but they have also evolved an N-terminal consensus motif for SUMOylation at K15/K13 which is not found in any other SUMO paralog (Kurepa et al. 2003).

Meanwhile, this whole genome duplication gave rise to SUMO1 and SUMO2 in the SUMO-γB lineage. Further duplications are then thought to have generated SUMO7 from SUMO1 or SUMO2, while the final tandem duplication gave rise to the SUMO7/8 pair and more interestingly, to SUMO3 from SUMO2 (Hammoudi et al. 2016). After this event, SUMO3 is thought to have adapted to a distinct function as SUMO2 and SUMO3 now only share 48% identity as opposed to the more than 70% observed in the two other tandem duplication pairs.

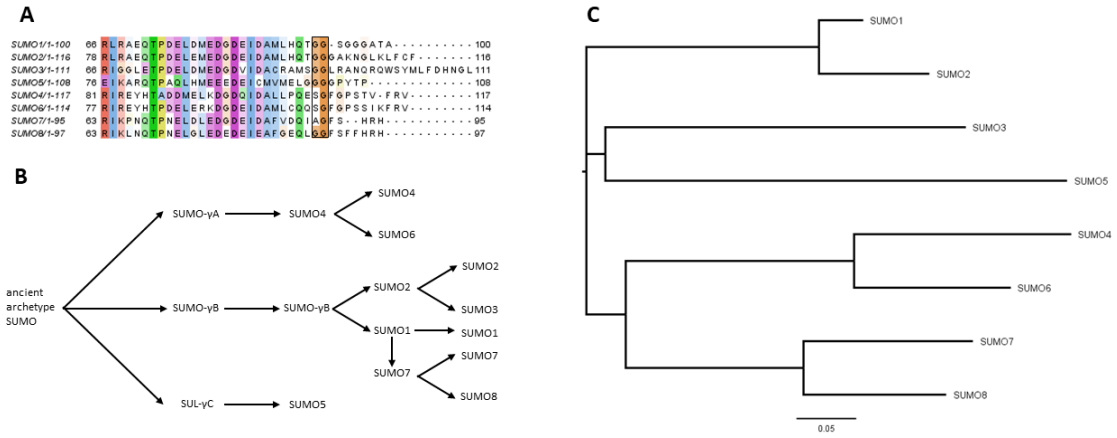

Figure 19: Similarity between SUMO paralogs. *A: Sequence alignment of SUMO paralog sequences. The diglycine motif necessary for conjugation in highlighted in black. Sequences were retrieved from the TAIR database, aligned using ClustalX and visualised in Jalview. Jalview's ClustalX-based colour scheme, which colour-codes amino acids by residue type, was applied. The intensity of the colouring reflects the degree of amino acid conservation. B: Model for the evolution of the Arabidopsis SUMO paralog from an ancient archetype SUMO as proposed by Hammoudi et al. Figure adapted from* (Hammoudi et al. 2016)*. C: Neighbour-joining tree of SUMO paralog sequences shows a dichotomy between canonical and non-canonical SUMOs. Neighbour-joining tree was produced using ClustalX and visualised using Figtree.*

In order for both copies to survive the evolutionary selection pressure, SUMO1 and SUMO2 must have functionally diversified quickly. As both paralogs are largely functionally redundant (Saracco et al. 2007), this is likely to have happened through the evolution of separate expression patterns. Indeed, distinct tissue-specific expression patterns have been described for *SUM1, SUM2* and *SUM3*. GUS staining showed that *SUM1* and *SUM2* are generally expressed in a complementary pattern (van den Burg et al. 2010). While *SUM1* was widely expressed in leaves, except in the vasculature, *SUM2* was mainly detected in the vasculature. This pattern is exceptionally clear in seedlings. However, in mature leaves *SUM2* expression - though still abundant in the vasculature- is more widespread leading to a certain degree of overlap between the two paralogs.

In roots, *SUM1* is again expressed ubiquitously, except for in the vasculature. Additionally, no expression was present in the root apices, or lateral root primordia. However, the expression complementarity is less complete in root tissue: *SUM2* is strongly expressed in lateral root primordia and in the vasculature of lateral root, but less so in the primary root vasculature. *SUM2* expression was also detectable in part of the rcanonioot apex, though it was not present throughout, nor was it detectable in root hairs.

In reproductive tissues, *SUM1* and *SUM2* have disparate expression patterns. *SUM1* expression was detected in immature anthers, at the base of mature flowers and in developing seeds. Meanwhile, *SUM2* expression was present in the vasculature of filaments and sepals, in immature anthers and at the base of siliques. In embryos, *SUM1* is expressed throughout development, while *SUM2* expression ceased after the late heart stage. This intense involvement in embryogenesis supports the non-viable phenotype of *sum1sum2* double mutants (Saracco et al. 2007).

Expression of *SUM3* was substantially less widespread, but very tissue-specific. In seedling leaves, *SUM3* expression was only detected in hydathodes while in mature leaves it was also weakly expressed in the vasculature. In roots, *SUM3* was only expressed in lateral root primordia and the vasculature of developing lateral roots. When the reproductive tissues were examined, *SUM3* expression was only observed in mature anthers and immature ovaries. No *SUM3* expression was detected in embryos, which tallies with the lack of embryonic development phenotype observed in *sum3* mutants (van den Burg et al. 2010).

The SUMO paralogs do not only differ in expression pattern, they also have differing affinities with SUMOylation enzymes. In a Yeast-2-Hybrid assay, SUMO1 and SUMO2 were able to interact non-covalently with the conjugating enzyme SCE1, while SUMO3 and SUMO5 were not (Lois et al. 2003; Castaño-Miquel et al. 2011). However, when the paralogs were incubated with the activating enzyme SAE1 and the conjugating enzyme SCE1, low rates of conjugation of both SUMO3 and SUMO5 to the C-terminal domain of SUMOylation substrate CATALASE 3 (CAT3Ct) were observed, indicating that they were both able to interact with SCE1, be it with a much lower affinity. Additionally, when SUMO1 amino acid residues predicted to interact with SAE1 or SCE1 (Lois & Lima 2005) were mutated to their non-conserved SUMO3 or SUMO5 equivalents, rates of SUMO conjugation were decreased. This indicates that interaction affinity with both SAE1 and SCE1 could play a role in regulating paralog specificity and therefore contribute to the low levels of SUMO3 and SUMO5 conjugation as opposed to SUMO1 and SUMO2 conjugation (Castaño-Miquel et al. 2011).

In order to better understand the transcriptional balance between SUMO isoforms and to assess any changes in that balance as a result of abiotic stress treatment, RT-qPCR was used to monitor SUM transcript abundance. Col-0 SOX lined and *ots1/2* mutants were also included in this analysis so as to establish whether SUMO1 overexpression or the lack of OTS1/2 function and the associated increase in SUMO conjugate accumulation affects this balance.

6.2 RESULTS

6.2.1 Differential transcription of SUM genes in 10 day-old seedlings

First, the transcript levels of the *SUM* genes in 10 day-old Col-0 WT seedlings were compared relative to one another other using -ΔCt as a metric. Though this metric can act as an approximate gauge to express whether a given transcript is more or less abundant than another, numerical differences should not be interpreted directly as they can be impacted by differences in primer efficiency. The highest transcript levels were found to correspond to the archetypal *SUM* genes *SUM1* and *SUM2.* Substantially lower transcript levels were detected for *SUM3* and *SUM5*. The *SUM4* transcript detected was of approximately equal abundance to that of *SUM3* and *SUM5*, while *SUM6* transcript abundance was much lower. (Figure 20) The extremely low transcript level of *SUM6* is consistent with its lack of record in the literature.

The difference in transcript level observed between *SUM1* and *SUM2* on the one hand and *SUM3* and *SUM5* on the other hand corresponds to that previously described in the literature (Kurepa et al. 2003; Saracco et al. 2007; Lois 2010). The expression of canonical SUMO-encoding genes *SUM1, SUM2, SUM3* and *SUM5* has previously been quantified though associated expressed sequence tags (ESTs) (Kurepa et al. 2003; Lois 2010). Their expression has also been confirmed though RT-PCR and microarray analysis. Interestingly, microarray results also pointed towards low levels of *SUM4* transcription, but as this result was not backed up by ESTs or RT-PCR, it was dismissed (Saracco et al. 2007).

However, RT-qPCR offered the possibility to measure *SUM* gene transcript abundance with a higher sensitivity. This method allowed for the detection of *SUM4* and *SUM6* transcript in 10 day-old seedlings in addition to that of the four canonical SUMO paralogs (Figure 20). Transcription of *SUM4* was within the same range as that of *SUM3* and *SUM5*, while that of *SUM6* was considerably lower. In fact, *SUM6* transcript levels were so low that it was not possible to confirm primer quality using the standard curve method detailed in the methods section. In the case of *SUM4*, it was only possible to obtain reliable standard curve data using

the two highest of three cDNA dilutions. This was not the case for *SUM3* and *SUM5,* indicating that *SUM4* transcript is likely to be less abundance than the results shown in figure 19 indicate. Those results must therefore be treated with the necessary caution. Additionally, designing specific primers for *SUM7* and *SUM8* using either translated or UTR sequences as templates proved impossible.

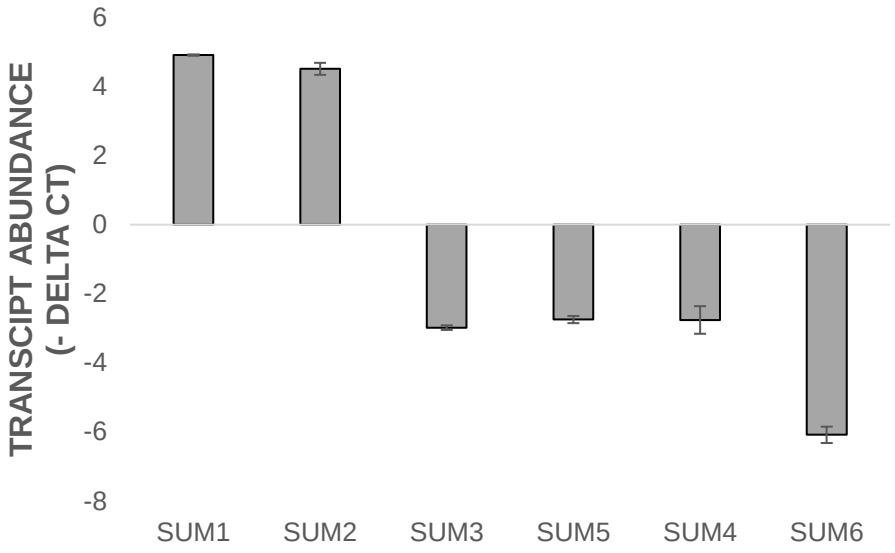

Figure 20: Differential transcription of SUMO paralogs in 10 day-old seedlings. *SUM1 and SUM2 transcripts are substantially more abundant than those of the other SUMO paralogs. Both SUM4 and SUM6 transcripts were detected. Transcript abundance was normalised to UBC9 and CBP20 housekeeping genes. Error bars represent standard error based on three biological and two technical repeats.*

6.2.2 Combined influence of SUMO overexpression and abiotic stress on transcription of the SUMO paralogs

RT-qPCR was used to monitor the transcription of the four canonical *SUM* genes and their transcriptional regulation in response to abiotic stress and SUMO overexpression. Due to the low transcript levels of *SUM4* and especially *SUM6* and to the lack of time to repeat experiments where needed, the data collected for *SUM4* and *SUM6* transcription in plants grown in stress conditions were deemed of insufficient quality to provide a basis for solid data analysis and were therefore omitted. In order to accurately portray the influence of SUMO1 overexpression in the Col-0 SOX lines on transcription of the native *SUM1* gene, primers which only detected native *SUM1* expression were used.

In order to gain an overall impression of canonical *SUM* transcription levels in the different genotypes and stress conditions tested, the Ct values obtained were compared to the Ct values

obtained for Col-0 WT plants grown in MS conditions. The resulting variations in transcript abundance can be seen in figure 21. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005.

In control conditions, all genotypes showed a significant decrease in *SUM5* transcription when compared to Col-0 WT, while none of them exhibited a difference in *SUM3* transcript levels. In the case of *SUM1* and *SUM2*, only Col-0 SOX L8 and *ots1/2* showed a significant decrease in transcript abundance.

Meanwhile, in the stress conditions tested, all samples but one showed a significant decrease ($p \leq 0.05$) in transcription when compared to Col-0 WT grown in MS medium. In the case of *SUM2* in Col-0 WT grown in the presence of 50 mM NaCl the p-value was $p < 0.1$. The decrease in transcription is seemingly more pronounced in severe stress conditions (100 mM NaCl) than under moderate stress (1 µM ABA and 50 mM NaCl), with few clearly visible differences between genotypes. In order to better assess the hypothesis that abiotic stress treatment has a larger effect on *SUM* gene transcript abundance than genotype, the effects of genotype and exposure to abiotic stress on transcript levels were separated out.

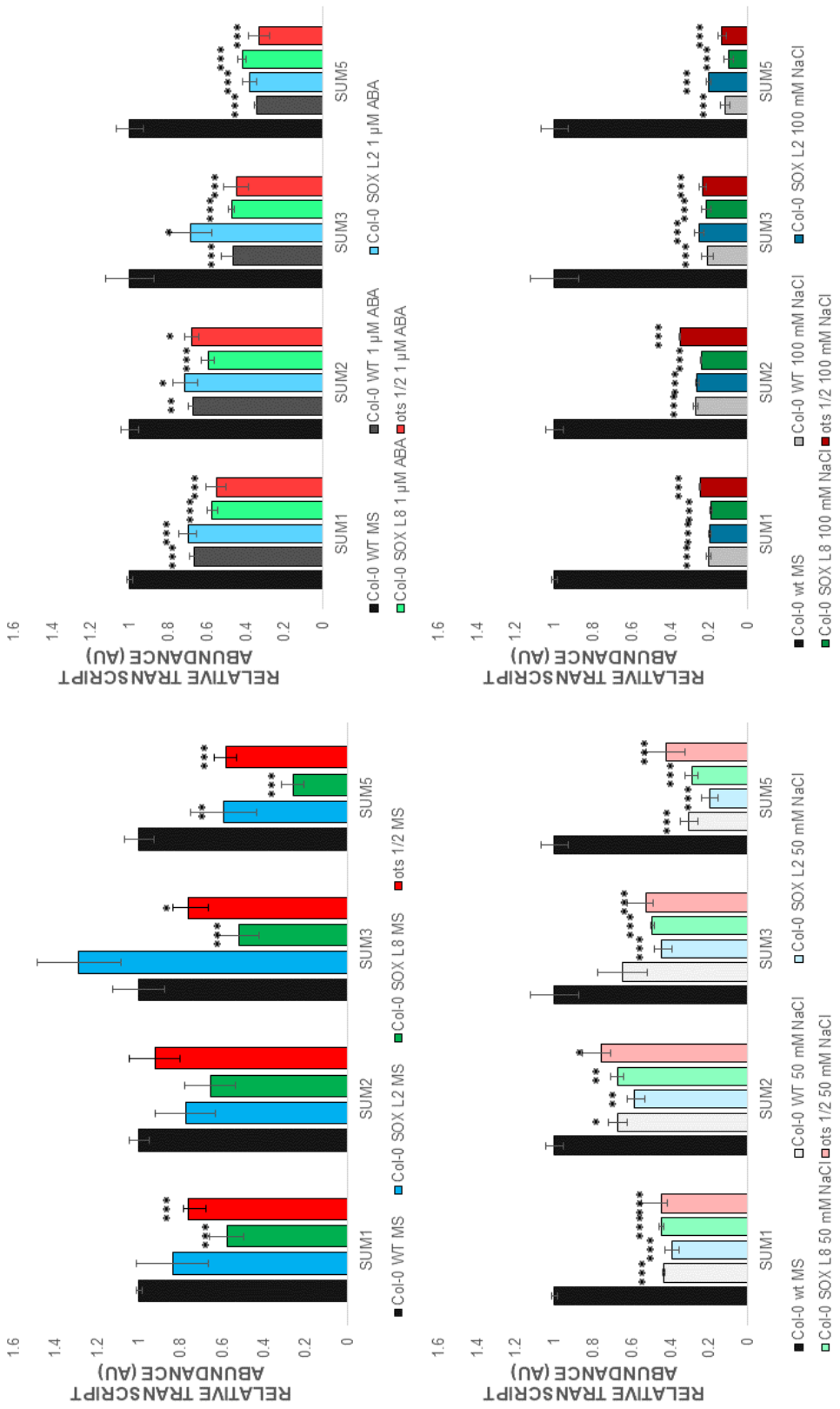

Figure 21: Combined influence of SUMO overexpression and abiotic stress on transcription of the SUMO paralogs. *Differences in SUM transcript abundance in 10 day-old seedlings grown on A: unsupplemented MS medium B: MS medium supplemented with 1 µM ABA C: MS medium supplemented with 50 mM NaCl D: MS medium supplemented with 100 mM NaCl.*

***Figure 21, continued**: All Ct values were compared to the Ct values obtained for Col-0 WT grown in MS conditions. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

6.2.3 Influence of SUMO overexpression on transcription of the SUMO paralogs

In order to assess the effect of SUMO levels on *SUM* gene transcription, Ct values for each growth condition were compared to values obtained for Col-0 WT plants grown in that condition. The resulting differences in transcript abundance are shown in figure 22. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005.

In untreated controls, the lower levels of SUMO overexpression in Col-0 SOX L2 only affected the transcript levels of *SUM5*, while the high SUMO conjugate levels in Col-0 SOX L8 affected levels of *SUM1, SUM2* and *SUM5*. Meanwhile, *ots1/2* mutants exhibit the same trend as Col-0 SOX L8 plants, though more mildly.

In plants treated with 1 µM ABA, only Col-0 SOX L8 plants showed slightly reduced levels of *SUM1* and SUM3 transcripts, while all other lines showed no difference in transcript abundance when compared to Col-0 WT plants. Treatment with 50 mM NaCl did not differentially affect transcript levels in any of the lines. When plants were subjected to severe salt stress (100 mM NaCl), SUM5 transcript levels increased in Col-0 SOX L2 but decreased in Col-0 SOX L8 when compared to Col-0 WT. Additionally, *ots1/2* mutants showed decreased transcript levels of *SUM1* and *SUM3* but not *SUM5*.

As expected, the effect of the seedling genotype on *SUM* transcript levels was minor in all conditions but the unsupplemented MS control condition. The next topic of investigation was the effect of abiotic stress on *SUM* transcript abundance in each of the studied genotypes.

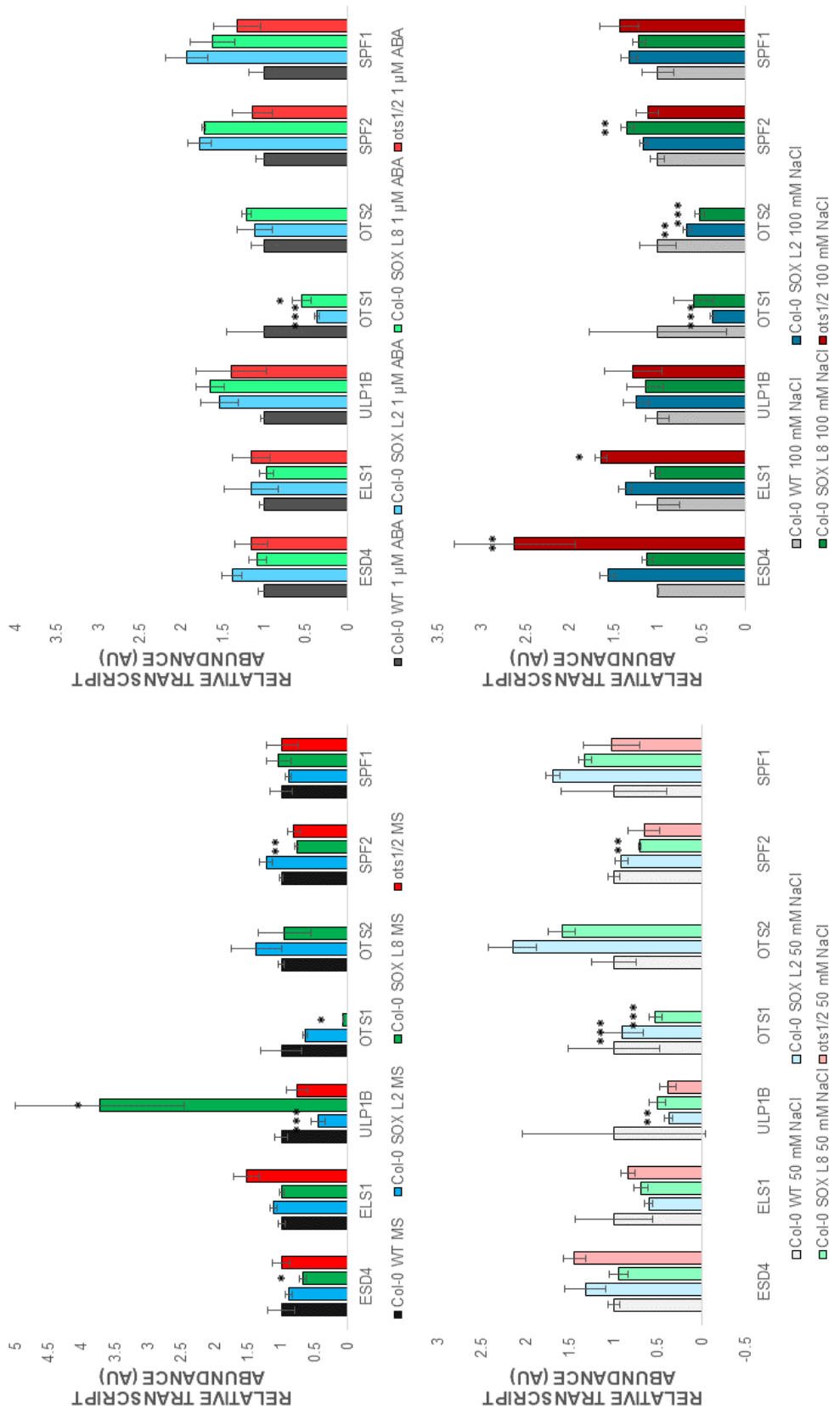

Figure 22: Effect of genotype on transcription of the SUMO paralogs. *Differences in SUM transcript abundance in 10 day-old seedlings grown on A: unsupplemented MS medium B: MS medium supplemented with 1 µM ABA C: MS medium supplemented with 50 mM NaCl D: MS medium supplemented with 100 mM NaCl.*

***Figure 22, continued**: All Ct values were compared to the Ct values obtained for Col-0 WT grown in MS conditions. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

6.2.4 Influence of abiotic stress on transcription of the SUMO paralogs

In order to assess the effect of abiotic stress on *SUM* gene transcription, Ct values for each growth condition were compared to values obtained for plants of that genotype grown on unsupplemented MS. The resulting changes in transcript abundance are shown in figure 23. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005.

In the case of *SUM1* and *SUM2,* both Col-WT and *ots1/2* mutant plants showed a moderate decrease in transcription in the presence of 1 µM ABA and 50 mM NaCl (the reduction in *SUM2* transcription in Col-WT plants grown on medium supplemented with 50 mM NaCl was significant to p<0.1) and stronger downregulation in the presence of 100 mM NaCl. However, this was not the case in the SOX lines. In Col-0 SOX L2, 50 mM NaCl treatment caused a mild reduction in *SUM1* but not *SUM2* transcription while 1 µM ABA treatment produced no significant effects. 100 mM NaCl treatment did cause a larger downregulation of *SUM1* and *SUM2*, with the reduction in *SUM2* transcript abundance significant to p<0.1. Meanwhile, only the severe stress of 100 mM NaCl treatment caused downregulation of *SUM1* and *SUM2*.

For *SUM3* a downregulation by 1 µM ABA or 50 mM NaCl treatment was only observed in Col-0 WT plants while 100 mM NaCl treatment caused a reduction in *SUM3* transcription in all genotypes.

The pattern of changes in *SUM5* transcript levels seems less clear. In the case of Col-0 WT and *ots1/2* mutant plants, the same downregulation previously described for the other *SUM* genes emerges, except for the transcription of *SUM2* in *ots1/2* mutants grown in the presence of 50 mM NaCl, where the decrease was not significant. However, Col-0 SOX L2 plants harboured only a mild downregulation significant to p<0.1 after 50 mM NaCl treatment while in Col-0 SOX L8 plants 1 µM ABA treatment led to one of the only upregulations observed in this dataset

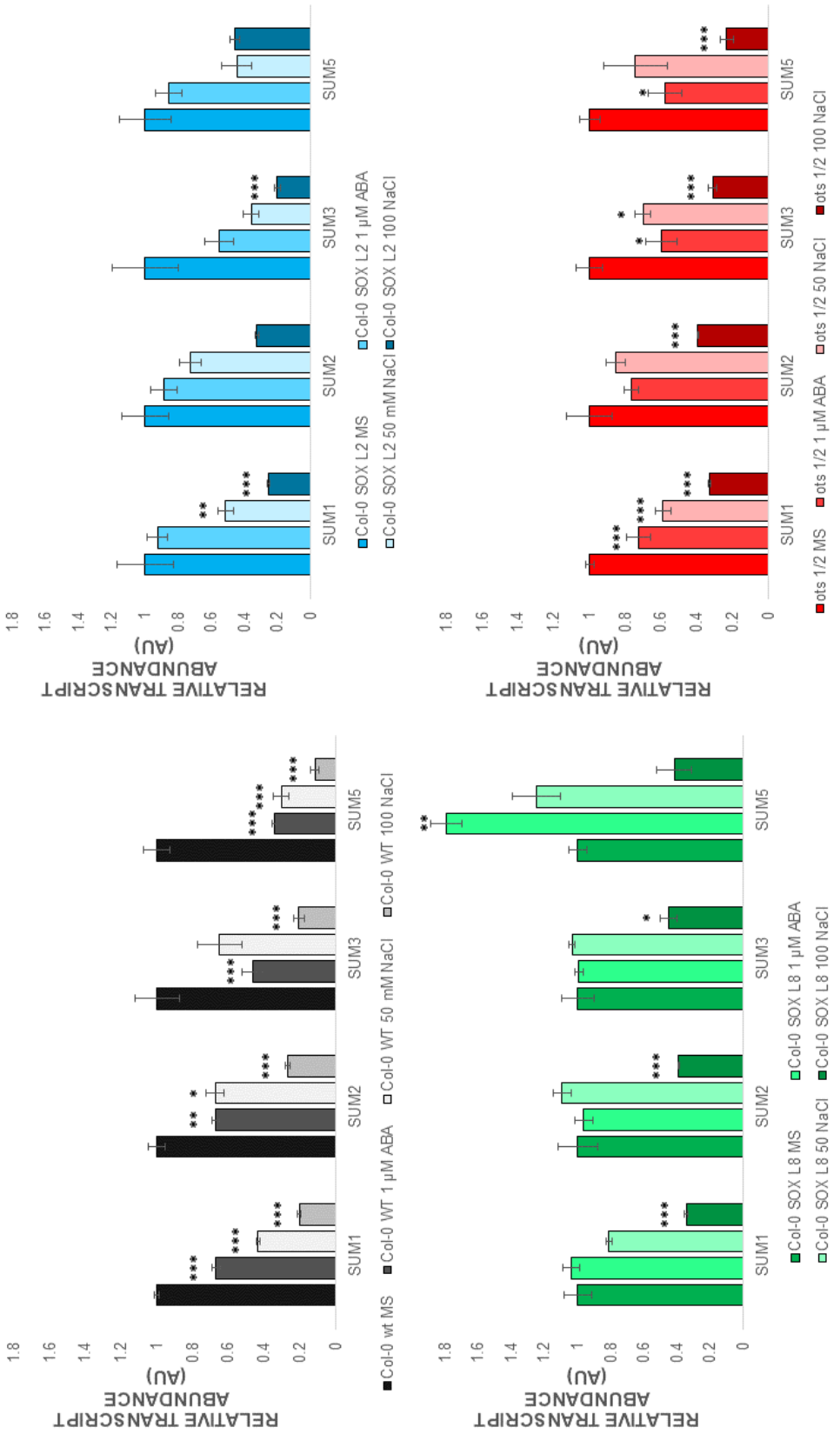

Figure 23: Influence of abiotic stress on transcription of the SUMO paralogs. *Changes in SUM transcript abundance in 10 day-old seedlings of several genotypes after 6 days of treatment with 1 µM ABA, 50 mM NaCl and 100 mM NaCl. A: Col-0 WT, B: ots1/2 mutant, C: Col-0 SOX L2, D: Col-0 SOX L8.*

Figure 23, continued: *Ct values for each genotype were compared to the Ct values obtained for that genotype when grown on unsupplemented MS. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

Overall, abiotic stress had a clear impact on the transcriptional regulation of the *SUM* genes, especially on those encoding the more readily conjugatable SUMO1/2 and SUMO3. There was also a clear distinction between the effects of moderate and severe abiotic stress, the former of which affected the genotypes differently according to SUMO overexpression levels. The similar transcript pattern seen in Col-0 SOX L8 and *ots1/2* seedlings suggests that plants attempt to counteract a perturbation of their SUMO balance caused by both overSUMOylation and defective deSUMOylation in the same way, namely by reducing the abundance of mRNA corresponding to the more highly conjugatable archetypal SUMOs. The observation that Col-0 SOX L2 seedlings showed no significant reduction in the quantity of *SUM1/2* transcript detected points toward the accumulation of SUMO conjugates rather than free SUMO as a possible cause for this downregulation effect, as Col-0 SOX L2 plants do accumulate high levels of free SUMO but the increase in SUMO conjugates levels they exhibit is substantially smaller than that seen in Col-0 SOX L8 plants (see figure 7). This similarity between Col-0 SOX L8 and *ots1/2* is interesting, as it is not clearly present in the regulation of the genes encoding either the ULP SUMO proteases (see chapter 5) or the SUMOylation enzymes (see chapter 7).

6.3 DISCUSSION

6.3.1 SUMO4 and SUMO6 may not be pseudogenes

The qPCR data confirm the expression of *SUM4,* which has previously been detected using micro array data. They also show the presence of a very low level of *SUM6* transcript. In both cases, variation between replicates was larger for *SUM4* and *SUM6* readings than for canonical *SUM* genes. Surprisingly, a search for previously obtained microarray data with which to compare the qPCR data obtained here yielded the finding that limited Genevesigator (Hruz et al. 2008) expression data is now available for both genes and that expression data for *SUM6* but not *SUM4* is available through the eFP Gene Browser (Winter et al. 2007). In general, this data confirmed the patterns seen in the qPCR analysis carried out in this study. However, while it also showed higher expression levels for *SUM4* than *SUM6*, seedling *SUM4* expression was markedly lower than *SUM3* and *SUM5* expression. As is was not possible to test the *SUM4* and *SUM6* primers used in this study to the same standard as the canonical SUMO primers due to low transcript levels, their quality is likely to be the cause of this discrepancy.

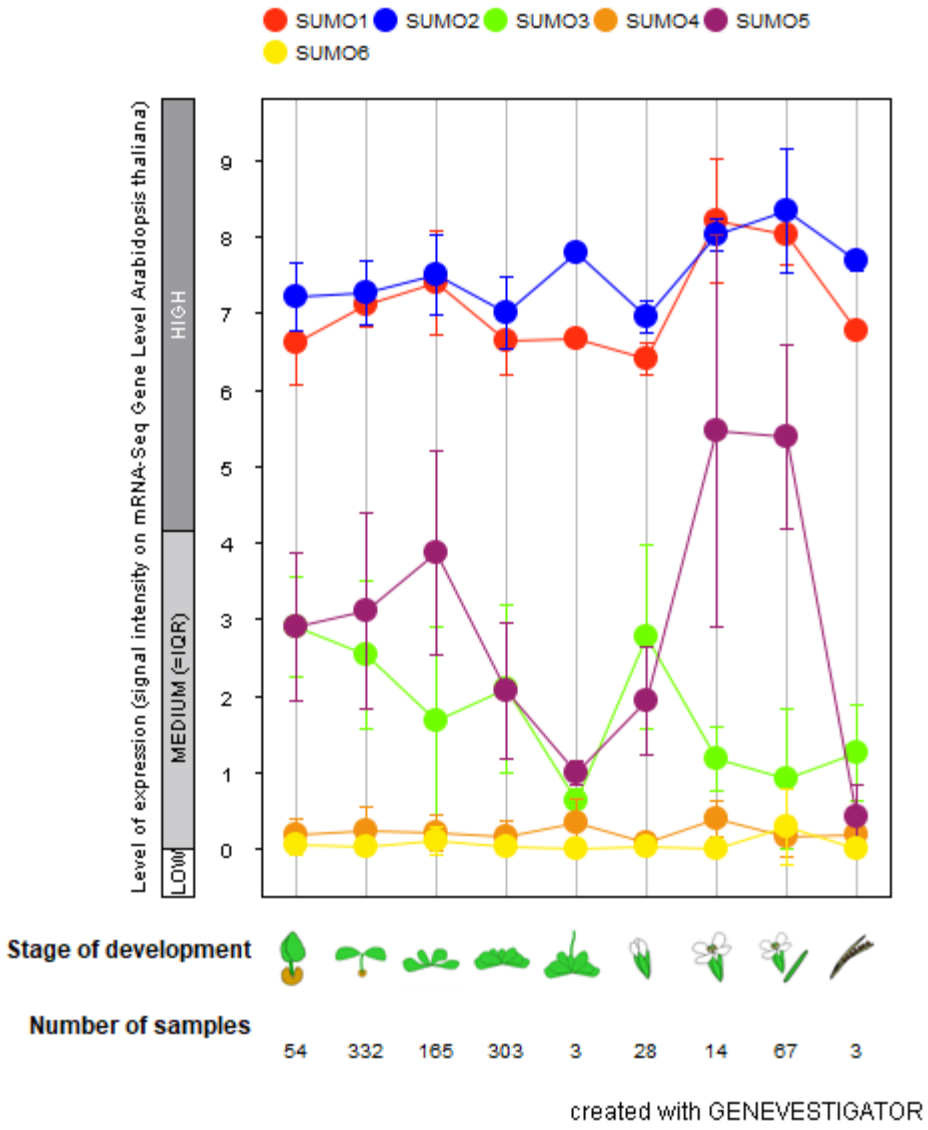

Figure 24: Variations in expression level of the six studied SUM genes throughout plant development. *mRNASeq data from the Genevestigator database show a similar balance in expression of the SUM genes in the seedling stage to the one found using qPCR. However, there is a discrepancy in the expression level of SUM4 (compare Figure 20).*

Tissue-specific expression data showed *SUM6* expression was markedly higher in mature pollen than in any other tissue, while floral tissue yielded the highest results for *SUM4*.

If the SUMO4 and SUMO6 paralogs are indeed expressed and biologically relevant, they are likely to perform different roles to those of the canonical SUMO paralogs. SUMO4 and SUMO6 lack the second part of the highly conserved diglycine motif. In place of the Gly-Gly sequence seen in the canonical SUMOs, SUMO4 and SUMO6 have a Gly-Ser sequence (see figure 25A). As the diglycine motif is necessary for conjugation (Saitoh et al. 1997), SUMO4 and SUMO6 may act through SIM motif interactions. Indeed, SUMO 4 and SUMO6 show conservation of the basic and hydrophobic character of the amino acid residues predicted to interact with SIM

motifs of target proteins (Castaño-Miquel et al. 2011), although the properties of the equivalents of two spatially adjacent residues, the hydrophobic SUMO1 V30 and the basic SUMO1 R50 are each replaced by a residue with the other property in SUMO4 and SUMO6. This modified SIM interaction surface may allow for paralog specificity in interactions.

Interestingly, SUMO4 and SUMO6 exhibit a greater conservation of the amino acid residues proposed to mediate interactions with SAE1 and SCE1 than SUMO3 and SUMO5 (Castaño-Miquel et al. 2011). If they do interact with the SUMOylating enzymes, they may play an inhibitory role. The same may be true for any interaction with ULPs: both SUMO4 and SUMO6 retain the SUMO1 Q90 equivalent (Q105 and Q101 for SUMO4 and SUMO6 respectively) which has been shown to mediate interaction with ULPs (Budhiraja et al. 2009), while SUMO3 and SUMO5 do not.

The most remarkable difference in SUMO4 and SUMO6 when compared to the canonical SUMOs is the emergence of a consensus SUMO site at their N-terminus. As it is the only consensus-type SUMO site present in any of the SUMO paralogs it is likely to be integral to any specific biological function SUMO4 and SUMO6 may have evolved in a neofunctionalization process.

6.3.2 The SUMO paralogs are themselves subject to post-translational modification

The presence of the consensus SUMO site in SUMO4 and SUMO6 prompted an investigation into the presence of SUMO sites in the other SUMO paralogs. Previous *in* vitro experiments have shown that when recombinant SUMO1 and SUMO2 were incubated with E1 and E2 enzymes, they were both able to form polySUMO chains, with SUMO1 forming higher molecular weight chains than SUMO2 (Colby et al. 2006). Additionally, the formation of these poly-SUMO chains was increased in the presence of a truncated form of the E4 ligase PIAL2. In the case of SUMO3, addition of the truncated PIAL2 was necessary for the formation of poly-SUMO chains of more than two monomers (Tomanov et al. 2014).

This requirement for PIAL2 may be due to the reduced SUMO3's reduced affinity for SAE1 and SCE1 (Castaño-Miquel et al. 2011). Although it was widely believed that SUMO3 could not polymerise (Colby et al. 2006; Miura & Hasegawa 2010; Castaño-Miquel et al. 2011), the formation of SUMO3 dimers had in fact been observable when SUMOylation of the yeast substrate ScPCNA was tested. While SUMO1 and SUMO2 formed a clear ladder of PCNA-SUMO conjugates, SUMO3 only showed a faint diSUMO3-PCNA band in addition to the stronger SUMO3-PCNA band (Colby et al. 2006). Faint di-SUMO3-sized bands were also present after

heat shock treatment of seedlings, becoming noticeably more intense with loss of *SUM1* and/or *SUM2* function (van den Burg et al. 2010).

SUMO site location has been investigated *in vivo* in SUMO1 and SUMO2. For SUMO2, the K10/R mutation abolished higher molecular weight SUMO chains and reduced triSUMO2 levels but did not affect diSUMO2 levels (Colby et al. 2006), suggesting K10 is not the only active SUMO site in SUMO2. It was also suggested that SUMO1 may have a SUMO site at K10 in homology with SUMO2 (Colby et al. 2006). This was later confirmed by mass spectrometry (Tomanov 2014).

More data is available for SUMO1. Mass spectrometry assays found *in vivo* SUMOylation at K42 and K23. Interestingly, K23 was found to serve as an attachment site for both SUMO1 and ubiquitin. Meanwhile, when di-SUMO levels in both K23 and K42 SUMO1 K/R mutants were examined, only K42R mutation led to a reduction in di-SUMO accumulation (Castaño Miquel 2015) (Figure 25).

These three lysine residues are highly conserved between SUMO paralogs and K23 in particular is conserved in all paralogs. The lysine residue K42 is conserved in all SUMO paralogs except SUMO7, while the lysine residues equivalent to K10 are conserved in all paralogs except SUMO3 and SUMO5.

In order to further investigate whether the SUMO sites associated with these conserved lysine residues are conserved between paralogs, the SUMO and SIM site production software GPS-SUMO (Zhao et al. 2014) was used to predict SUMO sites based on protein sequence. The associated probabilities for all SUMO sites can be found in supplementary table 4. A schematic illustrating the SUMO sites with highest conservation and predicted probability. In accordance with the experimental data, the algorithm returned SUMO2 K10 and SUMO1 K42, although it predicted the second SUMO site at K21 rather than K23. Two other remarkable predictions emerged.

Firstly, the algorithm predicted two SUMO potential SUMO sites at the C-terminus of SUMO2. As they are located past the diglycine cleavage site, they are likely to be biologically irrelevant, highlighting the limitations of sequence-based prediction methods.
Secondly, the algorithm also predicted a SUMO site in the central region of the protein sequence for both SUMO2 and SUMO3 at the lysines equivalent to SUMO1 K42, which are K54 for SUMO2 and K42 for SUMO3. Additionally, a second SUMO site in SUMO3 was predicted to be located at K54. As this prediction harboured the lowest p-value and the only consensus

sequence of all SUMO site predictions, it would be interesting to experimentally investigate whether these lysine residues functions as a SUMO site *in vivo*.

With a view to providing a more complete picture of possible SUMO sites in the consensus SUMO paralogs, the more divergent SUMO5 sequence was also examined for putative SUMO sites using GPS-SUMO. Here, the algorithm returned three possible SUMO sites with associated p-values in the same range as those discussed above. The only predicted SUMO acceptor lysine which was located at a conserved site was SUMO5 K52, the equivalent of SUMO1 K42. Both other candidates (K60 and K78) are lysine residues which are conserved as arginine residues in SUMO1/2/3, suggesting that it is their role as bearers of a positive charge rather than their character as a SUMO acceptor lysine which has been subject to conservation. Interestingly, the SUMO1 K23 equivalents (K23 and K33 respectively) were not strong SUMO site candidates in either SUMO3 or SUMO5, whereas the adjacent K21 and K31 returned much lower p-values. As the same pattern holds true for SUMO1, where the SUMO acceptor lysine was found to be K23 rather than the predicted K21 *in vivo* the validity of these predictions will need to be confirmed experimentally.

Taken together, the SUMO site predictions, sequence homology and experimental data suggest at least one SUMO site in the N-terminal region for both SUMO1 and SUMO2. Additionally, a more centrally localised SUMO site may be present in all canonical paralogs (figure 24). It is tempting to speculate that these two types of putative SUMO sites could have different functions, as has previously been reported for ubiquitin (Walsh & Sadanandom 2014). The case of SUMO1 K23, which was found to undergo both SUMOylation and ubiquitination, is especially interesting. On the one hand, the acceptor lysine for this SUMO site is the only one to be conserved in all eight SUMO paralogs, which may indicate ancestral origins. On the other hand, no reduction in the formation of SUMO chains was reported (Castaño Miquel 2015), which may indicate preferential binding of ubiquitin over SUMO. Indeed, the preferential binding of ubiquitin over SUMO may explain the high SUMO site p-values for the SUMO K23 equivalents in SUMO3 and SUMO5 This SUMO site may play a role in the formation of hybrid SUMO-ubiquitin chains. Alternatively, it may allow for polyubiquitination and subsequent degradation of SUMO1. However, the area of SUMO homeostasis is poorly studied. While the data presented in this chapter give some insight into the transcriptional regulation of SUMO homeostasis, further research is needed in order to understand it at the protein level.

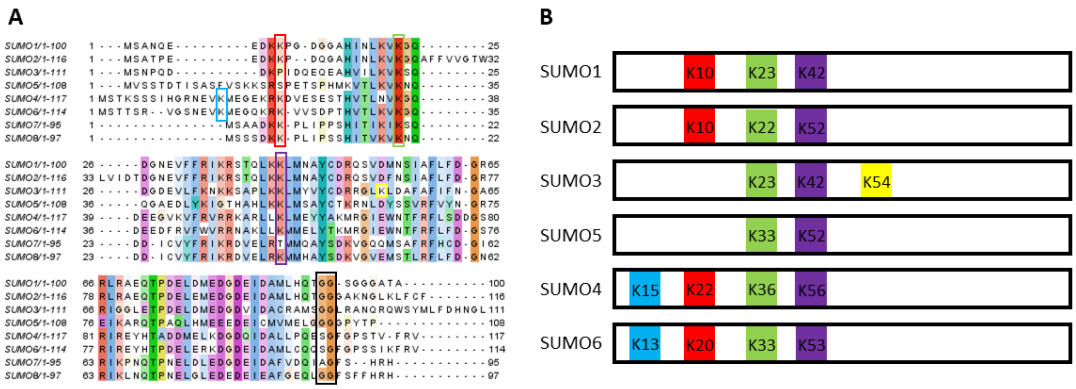

Figure 25: Localisation of putative SUMO sites in SUMO1/2/3. *A: Sequence alignment of the eight Arabidopsis SUMO paralogs showing the conservation of lysine residues serving as the acceptor for putative SUMO sites. The colours used correspond to B: schematic showing the respective amino acid numbering of lysine residues in putative SUMO sites in the six SUMO paralogs known to be expressed. Sequences were retrieved from the TAIR database, aligned using ClustalX and visualised in Jalview. Jalview's ClustalX-based colour scheme, which colour-codes amino acids by residue type, was applied. The intensity of the colouring reflects the degree of amino acid conservation.*

6.3.3 Changes in SUMO paralog expression may influence the conjugatability of non-archetypal SUMO paralogs

Previously, deSUMOylation was discussed as a mechanism for the modification of specific targets. However, plants must not only SUMOylate some proteins but not others, they must also modify the target with the right SUMO paralog. As the two non-archetypal paralogs have a series amino acid substitutions which impact their ability to interact with SAE1 and SCE1, the efficiency of *in vitro* conjugation of SUMO3 conjugation is reduced to approximately half - and that of SUMO5 to slightly more than a sixth of the conjugation rate of SUMO1 (Castaño-Miquel et al. 2011). However in some cases a substrate only interacts with a non-archetypal SUMO form, indicating that targets can provide a certain degree of paralog specificity (Zhang et al. 2017). Meanwhile, other proteins are preferentially SUMOylated with an archetypal SUMO paralog *in vitro* even though they are modified by a non-canonical paralog *in vivo*, indicating this is not always sufficient (Budhiraja et al. 2009).

Plants may have various mechanisms to overcome this barrier. Firstly, they may employ other enzymes to provide more complementary surface interactions and thus improve conjugation rates. Indeed, the addition of the SIZ1 to an *in vitro* conjugation assay containing SAE1 and SCE1 did not affect conjugation rates for wild-type SUMO1, but substantially increased conjugation rates of SUMO1 when the Q90A mutation was introduced (Budhiraja et al. 2009). This mutation made the protein more similar to SUMO3 and SUMO5 which sport methionine and leucine residues at this position respectively.

Secondly, there may be kinetic effects at play. There is no data available in this respect for SUMO5, but I propose these kinetic effects may have a role to play in the SA-mediated SUMOylation of NONEXPRESSER OF PR GENES 1 (NPR1) by SUMO3 (Saleh et al. 2015). It has been shown that SA treatment increases transcription of both *SUM1* and *SUM3*. However, while *SUM1* showed approximately nine-fold upregulation, upregulation of *SUM3* was approximately 90-fold, although variation between replicates was large for both genes (van den Burg et al. 2010). Van den Burg et al. do not state exactly which mature plant tissue was used for this RT-PCR, but considering other experiments in the same report, it is likely to have concerned rosette leaves. In these leaves, *SUM1* transcript abundance is approximately five-fold higher than that of *SUM3* in non-treated conditions (Saracco et al. 2007). If this ratio were to be taken as a baseline, *SUM3* would be twofold more upregulated than *SUM1* after SA treatment. If it is then adjusted for the approximately twofold higher affinity of the SUMOylation process for SUMO1 as measured by *in vitro* SUMOylation assay (Castaño-Miquel et al. 2011), the two paralogs come out on a roughly equal footing, making it easy to see how minor changes to this balance could allow SUMO3 to kinetically compete with SUMO1 for conjugation.

However, the above calculation does not factor in the influence of the highly tissue-specific expression pattern of *SUM3.* In SA-treated mature plant leaves, the increase in *SUM3* expression appears to leak from the vasculature, where it is already expressed in non-treated conditions, into the surrounding tissue (van den Burg et al. 2010). It is in this surrounding tissue that *SUM1* is most expressed and the above comparison could be made. As *SUM3* is most highly expressed in the vasculature, where *SUM2* is dominant over *SUM1* (van den Burg et al. 2010)*,* the possibility of SUMO3 competition becomes even more likely. In rosette leaves, the *SUM2* transcript level also approximately five times higher than that of *SUM3* (Saracco et al. 2007). However, *SUM2* is only approximately 3.5 times upregulated by SA treatment -as opposed to the 90-fold upregulation of SUMO3- (van den Burg et al. 2010), bringing the total transcript level of *SUM3* after SA treatment to just over five times higher than that of *SUM2*. It remains to be shown whether this balance hols true at the protein level. Though not explicitly tested in relation to SUMO3, the SUMOylation efficiency of SUMO2 is generally considered to be slightly inferior to that of SUMO1 (Colby et al. 2006; Castaño-Miquel et al. 2011), further pushing the kinetic balance in the favour of SUMO3. For example, a preferential conjugation rate of 1.8 compared to 2 of SUMO1 would still result in an approximately 2.85 fold kinetic advantage in favour of SUMO3.

Due to a lack of quality and quantity of data, this theory remains speculative: for example, the rate of *SUM* gene upregulation was not determined organ-specifically and showed considerable variation between replicates. However, it could provide a better insight into the conjugation of less-expressed paralogs SUMO3 and SUMO5.

Apart from the differences between archetypal and non-archetypal SUMO paralogs, there may also be differences in role between the archetypal SUMOs SUMO1 and SUMO2. A high-throughput interaction study found a number of proteins which interact with SUMO1 but not SUMO2 or the inverse (Mazur et al. 2017). Data on this phenomenon is very scarce as *in vitro* experiments are almost always performed with only one of the two archetypal SUMOs. The only other evidence of this phenomenon is in the maturase activity of SPF1 and SPF2, which were shown to mature SUMO1 but not SUMO2 (Liu et al. 2017). Though the expression patterns for SUMO1 and SUMO2 were found to be largely complementary, there are areas (such as parts of mature leaves) were both paralogues are expressed in overlapping areas (van den Burg et al. 2010). As the two paralogs are currently generally assumed to interchangeable in the literature, the importance to the plant of discerning between SUMO1 and SUMO2 is yet to be discovered.

6.3.4 SUMO overexpression and abiotic stress cause a dose-dependent downregulation of *SUM* genes to a set level

This chapter focussed on the examination of SUMO homeostasis through the transcriptional effects of SUMO overexpression and abiotic stress on the transcription of the four canonically expressed *SUM* genes.

Examination of the combined effect of both factors led to the finding that the decrease in transcript abundance in comparison to Col-0 WT plants grown in control conditions was more or less constant over the genotypes for plants grown in stress conditions. The moderate stress treatments (1 µM ABA and 50 mM NaCl) produced a downregulation to around half of wild-type levels for SUM1/2/3 and quarter of wild-type levels for *SUM5* (figure 21 B, C). In the more severely stress-inducing 100 mM NaCl treatment, *SUM1/2/3* levels were downregulated to around a quarter of wild-type levels while the *SUM5* transcript level was reduced to around an eighth (figure 21 D).

However, the differences in transcript abundance for plants grown in unsupplemented MS conditions show a different pattern. While Col-0 SOX L2 plants only exhibit lower levels of *SUM5* transcript, Col-0 SOX L8 and *ots1/2* mutant plants showed a reduction in the transcription of *SUM1, SUM2* and *SUM5* with a sharper decrease visible in Col-0 SOX L8 for all genes.

When the data was renormalised to show only the influence of abiotic stress on *SUM* transcript levels in the different genotypes, this pattern generally repeated itself. Nonetheless, there were some important differences between genotypes (figure 23). The largest decreases were present in Col-0 WT plants with downregulation of genes in the moderate stress conditions ranging from just under two times to four times downregulated, while severe stress conditions caused a downregulation of between four and eight times. Meanwhile, in *ots1/2* mutants this downregulation was slightly reduced with moderate stress causing up to two times downregulation and severe stress causing up to four times downregulation. The most important changes, however, were visible in the Col-0 SOX lines, where only the severe stress of 100 mM NaCl triggered downregulation of *SUM1/2/3* (figure 23 C, D).

These data suggest that abiotic stress treatment leads to a reduction in *SUM* transcript abundance. When the results obtained for the different genotypes grown on unsupplemented MS are interpreted in the light of this hypothesis, it becomes clear that Col-0 SOX L8 plants are showing the *SUM* signature of moderately stressed Col-0 WT plants when grown in control conditions. The same, though to a lesser extent, seems true for *ots1/2* mutants.
The differences in Col-0 SOX L2 are more puzzling, as the decreases in transcription were generally not significant in either plants grown on unsupplemented MS when compared to Col-0 WT (difference due to genotype) or plants grown in 1 µM ABA or 50 mM NaCl when compared to plants of the same line grown in control conditions (difference due to abiotic stress). However, when both effect were considered together, the *SUM* genes were downregulated to similar levels to all other genotypes when exposed to stress conditions. Taken together these data suggest that where the achievement of this downregulation was due to abiotic stress in Col-0 WT and was mainly contributed to by the high levels of SUMO overexpression in Col-0 SOX L8, Col-0 SOX L2 forms an intermediate between these two triggers. Differences due to increased SUMO levels and moderate stress treatment were in themselves not large enough to be statistically significant, but when put together they did account for the difference to be sufficiently large and significant.

I therefore propose a model in which the accumulation of SUMO conjugates is a proxy for stress in as far as its influence on the transcriptional regulation of *SUM* genes is concerned. The high levels of SUMO conjugate accumulation in Col-0 SOX L8 caused a substantial downregulation in *SUM* transcription in comparison to Col-0 WT plants when grown on unsupplemented MS medium and were only triggered to further downregulate *SUM* transcription by severe environmental stress. In the case of Col-0 WT plants, the changes can - by experimental design- only be due to environmental stress. In the other two genotypes, the balance is situated somewhere in between.

However, when compared the transcript levels seen in to Col-0 WT plants grown on unsupplemented MS, the overall downregulation of each of the *SUM* genes seemed to plateau at a certain level dependent on the intensity of the environmental stress condition, irrespective of genotype. These plateaus may serve as a barrier to the loss of the ability to react to a further increase in environmental stress or to protect the plant from the developmental cost of unnecessarily launching the severe stress metabolism. The exact trigger and signalling behind the transition from the moderate to the intense stress transcription plateau remain unknown. In any case, these data confirm the tight control of *SUM* expression previously proposed to be necessary for the correct control of the plant stress response (van den Burg et al. 2010).

In order to further elucidate the transcriptional regulation of the SUMO system by abiotic stress and SUMO overexpression, I investigated their influence on the third component of the SUMO system, the SUMOylation enzymes.

7 TRANSCRIPTIONAL ANALYSIS OF THE *ARABIDOPSIS* SUMO CONJUGATION ENZYMES

7.1 INTRODUCTION

As is briefly discussed in the introduction, six enzymes have currently been identified as forming part of the *Arabidopsis* SUMOylation machinery. These are the SUMO activating enzyme SAE1, the SUMO conjugating enzyme SCE1, the two E3 SUMO ligases HPY2 and SIZ1 and the two E4 SUMO ligases PIAL1 and PIAL2. The roles of these enzymes in the SUMO cycle are illustrated in figure 3, which can be found in the introductory chapter. In the following section, we provide an overview of current knowledge concerning each of these SUMOylation enzymes.

7.1.1 The E1 SUMO activating enzyme SAE1

SUMO ACTIVATING ENZYME 1 (SAE1) is the only known SUMO activating enzyme in *Arabidopsis.* It consists of two subunits: the large subunit, SAE2, and the small subunit, of which there are two isoforms, SAE1a and SAE1b (Kurepa et al. 2003). Interestingly, SAE1b is encoded by two identical loci which are part of a set of seven genes which have undergone tandem duplication. This duplication event is thought to be very recent as it is not present in all *Arabidopsis thaliana* accessions (Castaño-Miquel et al. 2013).

Though no data from GUS assays are available for the SAE1 subunits, microarray data suggest ubiquitous expression. *SAE1a* expression levels were highest in roots and shoot tips, while *SAE1b* levels were highest in seed and shoot tips. Overall, *SAE1a* expression was higher that *SAE1b* expression in all tissues but seeds, siliques and petals. Expression of *SAE2* is very low compared to that of *SAE1a* and *SAE1b,* but highest in roots, petals, seed and senescent leaves (Saracco et al. 2007). While a *sae1b* mutant has not yet been generated, the *sae1a* mutant shows no clear developmental phenotype (Castaño-Miquel et al. 2013). In contrast, *sae2* mutants are non-viable (Saracco et al. 2007).

Other than the identification of residues which interact with SAE2, little is known about the domain structure or function of the SAE1a/b proteins. As the residues which differ between the two isoforms are spread throughout the proteins rather than clustered to particular areas, sequence comparison does not provide any clues towards which mutations are responsible for

the difference in behaviour between the SAE1a and SAE1b. SAE1 holoenzyme which used only SAE1a as a small subunit was more proficient at establishing thioester bonds at high temperature and overall catalysed the in-vitro SUMOylation reaction more efficiently than SAE1 holoenzyme which used only SAE1b. This increased efficiency became more prominent with increasing temperature (Castaño-Miquel et al. 2013).
The other difference between SAE1a and SAE1b which has been reported is one concerning their cellular localisation. While both isoforms exhibited a mix of exclusive nuclear and nuclear and cytoplasmic localisations in transient expression assays, the balances between the number of cells showing each type of localisation was markedly different. When expressed alone, SAE1a localises to both the nucleus and the cytoplasm in more than 80 percent of assayed cells. Meanwhile, when SAE1b is expressed alone, it localises exclusively to the nucleus in more than 80 percent of assayed cells. However, when either of the isoforms is co-expressed with SAE2, both proteins localise exclusively to the nucleus, as does SAE2 when it is expressed alone (Castaño-Miquel et al. 2013).

As it is the subunit which carries the catalytic cysteine responsible for SUMO thioester formation, SAE2 has been more extensively studied than SAE1a/b. It was found to be composed of four domains. The protein was found to be composed of an adenylation, catalytic, ubiquitin fold and C-terminal domain. (Lois & Lima 2005). SAE2 is thought to act in assembly line fashion, first binding SUMO non-covalently at the adenylation domain and activating it with the help of ATP. It is then transferred to the catalytic cysteine, with which it forms a thioester bond and pyrophosphate is released. A second SUMO moiety can then be bound to the adenylation domain, taking the place of the first at the catalytic cysteine after this is transferred to the E2 SUMO conjugating enzyme SCE1. While a crystal structure indicating this behaviour has yet to be produced for SUMO, it does exist for the related ubiquitin-like modifier NEDD8 (Schulman 2011).

However, it is not only the adenylation and catalytic domains that are important for SAE2 activity. The ubiquitin fold domain is necessary and sufficient for SEA2-SCE1 binding and its expression as a peptide has been shown to knock down SUMOylation (Castaño-Miquel et al. 2017). The C-terminal domain contains two nuclear localisation signals which determine intracellular localisation of SAE2. While the localisation remains predominantly nuclear, loss of either or both nuclear localisation signals (NLS) causes leakage of SAE2 into the cytoplasm. This mechanism is biologically relevant as a short fragment including NLS2 is known to be proteolytically cleaved from SAE2 in siliques (Castaño Miquel 2015). As discussed above, the

subcellular localisation of SAE2 is important, because it determines the subcellular localisation of SAE1a and SAE1b.

Both SAE1a and SAE2 have also been confirmed as SUMOylation targets by mass spectrometry analysis. Four SUMO sites were found in SAE1a and six in SAE2. Two K/R mutations of the SAE2 SUMO sites yielded mutants which exhibited decreased SUMOylation: K234R (located in the catalytic domain) and K595R (located in the C-terminal domain) (Castaño Miquel 2015).

In human cells, SUMOylation of the E1 SAE2 equivalent in the catalytic domain affects the transfer of SUMO between the E1 and the E2 enzyme (Truong, Lee & Chen 2012). This mechanism may also be present in *Arabidopsis*, although the role of the catalytic domain in SCE1 binding is much smaller than in the human E1 (Castaño-Miquel et al. 2017).

In parallel, the role of C-terminal SAE2 SUMOylation has also been studied in human cells. Similarly to *Arabidopsis*, the deletion of both C-terminal nuclear localisation signals in human SAE2 led to the detection of SAE2 in the cytoplasm, though the majority if the protein remained in the nucleus. However, mutation of the five C-terminal SUMO sites led to a purely cytoplasmic localisation. Additionally, chemically blocking the nuclear import process caused SAE2 to gradually accumulate in the cytoplasm. These observations led to the hypothesis that it is C-terminal SUMOylation which keeps human SAE2 in the nucleus, retargeting it to the nucleus as it is constantly shuttled out (Truong, Lee, Li, et al. 2012). If this mechanism holds true in *Arabidopsis*, it would provide a mechanism for the part nuclear, part cytoplasmic localisation of the C-terminally cleaved SAE2 found in siliques (Castaño Miquel 2015).

7.1.2 The E2 SUMO conjugating enzyme SCE1

As with SAE1, SUMO CONJUGATING ENZYME 1 (SCE1) is the only E2 SUMO conjugation enzyme present in *Arabidopsis*. Microarray analysis showed SCE1 expression to be ubiquitous, with the highest levels recorded in shoot tips and petals and the lowest in seeds, siliques and senescent leaves. Overall, SCE1 expression is distinctly higher than that of any of the SAE1 subunits (Saracco et al. 2007).

Loss of function SCE1 through mutation is embryonically lethal (Saracco et al. 2007). However, dominant-negative transgenic lines expressing SCE1 C94S, a form of SCE1 which is catalytically inactive, are viable (Tomanov et al. 2013). Sufficiently high transgene expression caused early flowering and a severely stunted phenotype. Similarly to *siz1* mutants discussed below, the SCE1 C94S phenotype could be mildly alleviated by supplementation with ammonia (Tomanov et al. 2018). More moderate suppression of SCE1 function led to increased chlorosis in

response to ABA treatment (Lois et al. 2003) and to higher resistance against the turnip mosaic virus, which has been shown to encode an RNA polymerase which is SUMOylated by SCE1 using SUMO3 (Xiong & Wang 2013).

Transient expression of SCE1 C94S caused a reduction in the accumulation of SUMO conjugates, especially in the high molecular weight range. Constitutive SCE1 C94S expression strongly reduced high molecular weight SUMO conjugate accumulation, but also reduced the levels of free SUMO and altered the SUMOylation pattern observed in the low molecular weight range. *In vitro*, SCE1 C94S was unable to SUMOylate TATA-BOX-BINDING PROTEIN-ASSOCIATED FACTOR 7 (TAF7), which posed no problem to wild-type SCE1 (Tomanov et al. 2018).

While SCE1 localised to both the nucleus and the cytoplasm when expressed alone, co-expression with its interaction partners altered this pattern. Firstly, co-expression with the nuclear localised SAE2 led to the exclusively nuclear localisation of both proteins (Castaño-Miquel et al. 2017).
Secondly, co-expression with SUMO1, which itself localises to both the nucleus and the cytoplasm when expressed alone, caused both proteins to localise to the nucleus. However, this effect was dependent on the catalytic activity of SCE1 (Lois et al. 2003), suggesting the localisation of SUMO to the nucleus is due to its attachment to its substrates in the form SUMO conjugates, which are known to be most concentrated in the nucleus (Saracco et al. 2007).

SCE1 was also shown to have a number of SUMO sites, both N-terminally and C-terminally (Castaño Miquel 2015; Tomanov et al. 2018). Substitution of the acceptor lysine residues of the five SCE1 C-terminal SUMO sites had no visible influence on the plant SUMOylation profile (Tomanov et al. 2018). However, mutation of the three N-terminal SUMO sites of SCE1 bore more interesting results: both the K15R and K19R mutations decreased SUMO chain formation, while monoSUMOylation was affected to a much lesser extent. Conversely, the K28R mutation increases SUMO chain formation. Interestingly, while the SCE1 N-terminal SUMO sites seem to regulate SUMO chain formation, they are all located far from the catalytic cysteine (Tomanov et al. 2018).

In all cases, the addition of E4 SUMO ligase PIAL2 increased the efficiency of SUMO chain formation. However, as the relative extent of SUMO chain formation was unchanged between SUMO site mutants, the role of PIAL2 is seemingly to improve the efficiency of a process which SCE1 can fulfil autonomously (Tomanov et al. 2018).

7.1.3 The E3 SUMO ligases: HPY2 and SIZ1

HIGH PLOIDY 2 (HPY2), also known as METHYL METHANE SULFONATE SENSITIVITY 21 (MMS21), is the smaller of the two known *Arabidopsis* E3 SUMO ligases. It is localised to both the nucleus and the cytoplasm and has been shown to interact with SCE1 (Huang et al. 2009). In an *in vitro* assay, HPY2 was both mono- and diSUMOylated in the presence of E1 and E2 enzymes (Ishida et al. 2009).

When the SUMOylation profile of *hpy2* mutants was assessed, the high molecular weight SUMO conjugate smear was slightly less intense than that observed in Col-0 WT, but the reduction was not a prominent as that observed in *siz1* mutants. However, the heat shock SUMOylation response was strongly dampened in *hpy2* mutants, resulting in conjugation levels similar to those found in *siz1* mutants. Interestingly, in both control and heat shock conditions, *hpy2* mutants showed a number of differences in low molecular weight SUMO conjugate profile with respect to Col-0 WT. While *hpy2* mutants accumulated less mono- and diSUMO, a large band with an apparent molecular mass of just over 50 KDa was present in the *hpy2* SUMOylation profile, increasing in intensity after heat shock treatment (Ishida et al. 2009).

hpy2 mutants exhibit a strong, pleiotropic phenotype. Seedlings exhibit an increased curvature of cotyledons and an extremely short primary root (Huang et al. 2009). Roots of *hpy2* seedlings at the 7 day stage measured only 10% of the length of Col-0 WT seedling roots (Ishida et al. 2009), while 16% of germinated *hpy2* seedlings exhibited no primary root growth (Xu et al. 2013). When the roots were subjected to microscopic analysis, they were found to have reduced and distorted meristematic tissue (Ishida et al. 2009). As the plants mature, their stature is severely dwarfed and they often suffer growth arrest before bolting, while those which do bolt show fasciation and irregular leaf placement (Huang et al. 2009; Ishida et al. 2009). Meanwhile, HPY2 knockdown plants exhibited and early flowering phenotype (Kwak et al. 2016). When they were exposed to stress, they were found to be drought-tolerant (Zhang et al. 2013).

The hormone response is also affected in *hpy2* mutants: they were show be insensitive to both auxin and cytokinin (Huang et al. 2009) and to express ABA marker genes at increased levels (Zhang et al. 2013), while cytokinin marker genes were downregulated (Huang et al. 2009).

hpy2 plants exhibit incomplete fertility, which is manifested though the development of short siliques containing around 40% aborted seeds, with others showing abnormal seed coat development. Additionally, *hpy2* pollen is defective in development, germination and pollen

tube growth. This reduced ability to produce seeds is present in both Col-0 WT backcrosses with *hpy2* mutants, indicating that HPY2 plays a role in the development of both the male and the female gamete. However, certain properties, such as pollen tube growth and aborted seed count, showed asymmetry in the backcrosses. For these measures, the *hpy2* mutant phenotype was only present when *hpy2* was used as the male parent (Liu et al. 2014).

When both the root meristem and pollen growth phenotypes were examined more closely, they were found to be due to defective and erratic cell division. Specifically, *hpy2* mutants showed defects in chromosome condensation, sister chromatid segregation and homologous recombination, exhibition abnormal rates of chromosome fragmentation and double stranded breaks (Liu et al. 2014; Xu et al. 2013). Indeed, the meristematic cell death and associated erratic division pattern seen in *hpy2* mutants phenocopies the patterns observed in Col-0 WT plants subjected to DNA damaging treatments (Xu et al. 2013). The proposed role for HPY2 in DNA remodelling and repair is consistent with the colocalisation of HPY2 with condensed chromosomes in mitotically dividing cells (Ishida et al. 2009).

In stark contrast with the phenotypes of the SUMO protease mutants, the mechanisms which lie at the base of a number of the *hpy2* mutant's phenotype components have been identified. For example, HPY2 was shown to SUMOylate FLOWERING LOCUS C (FLC), thus stabilising it. Meanwhile, in *hpy2* mutants, the combination of *FLC* downregulation and a lack of FLC stabilisation though SUMOylation is thought to lead to early flowering (Kwak et al. 2016).

Three mechanisms are known to contribute to the short, disorganised root phenotype. Firstly, HPY2 is a subunit of the STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 (SMC5/6) complex, which plays and important role in DNA condensation, sister chromatid cohesion and homologous recombination, thus preventing DNA damage (Yuan et al. 2014).
Secondly, HPY2 SUMOylates the Brahma proteins, which aid the maintenance of the root stem cell niche (Zhang et al. 2017). This area, which is important for correct meristematic architecture, is known to be incorrectly developed in *hpy2* mutants (Yuan et al. 2014).
Thirdly, the lack of growth may be in part due the role of HPY2 in the cell cycle. SUMOylation of DPa by HPY2 causes the dissociation of the E2Fa/DPa complex, a transcription factor complex which leads the cell to transition from G1- to S-phase, although it has also been implicated in the G2/M transition (Liu et al. 2016). The effect of the lack of DPa SUMOylation in *hpy2* mutants are seen in two ways. On the one hand, cell cycle arrest leads to repression of growth, which is evident in the severely dwarfed phenotype.

Secondly, cell cycle arrest before the M-phase specifically leads to increased endocycling, which is also observed in *hpy2* mutants (Ishida et al. 2009; Liu et al. 2016).

SAR AND MIZ 1 (SIZ1) was the first SUMO E3 ligase to be identified in *Arabidopsis.* Though more SIZ1 targets than HPY2 targets have been identified, very few mechanisms have been fully elucidated. GUS assays showed SIZ1 localises to hydathodes and at lower level to the upper part of the cotyledons in 3 day-old seedlings. As the plants mature, expression expands to the vasculature and spreads through the leaves. In 10 day-old seedlings it is present in the upper half of leaves, while expression is ubiquitous in adult leaves. A similar process is observed in the roots, with expression first localising to the root vasculature, lateral root primordia and the tips of developing lateral roots and spreading further through the root as the plant grows. While SIZ1 was expressed in the lower part of the hypocotyl in 3 day-old seedlings, it was limited to the very base of the developing stem in 10 day-old seedlings. Immature siliques express SIZ1 at both ends of the structure, while expression is ubiquitous in mature siliques. In floral tissues, SIZ1 expression was observed throughout inflorescences, in the petal vasculature, style and ovules (Catala et al. 2007; Ling et al. 2012; Ishida et al. 2012). Interestingly, two studies also found expression in anthers (Catala et al. 2007; Ishida et al. 2012), while another did not (Ling et al. 2012). Inside the cell, SIZ1 localises to nuclear speckles (Miura et al. 2005), where it was found to interact with ubiquitin ligase CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) (J. Y. Kim et al. 2016).

SIZ1 consists of five domains: SAP, PHD, PINIT, SP-RING and SXS. The SP-RING is by far the most influential in SIZ1's function as an E3 SUMO ligase. Mutation of the SP-RING domain drastically affected the SUMOylation profile, especially in the lower molecular weight bracket, while loss of the PHD domain mildly affected SUMO conjugation. Together with the PINIT domain, the SP-RING domain mediates nuclear retention of SIZ1, while SUMO binding is mediated by the SXS domain (Cheong et al. 2009).

siz1 mutants exhibit a strong pleiotropic phenotype. At the molecular level, they accumulate markedly fewer SUMO conjugates than Col-0 wild-type plants and show a dampening of the SUMO conjugation response after heat shock treatment (Miura et al. 2005). At the macroscopic level, they are dwarfed, flower early and show increased resistance to the pathogen *Pseudomonas syringae*. As these aspects of the *siz1* phenotype could be mostly reversed with expression of the bacterial hydroxylase NahG, they are attributed to the large increase in salicylic acid accumulation observed in *siz1* mutants (Lee et al. 2006; Jin et al. 2008).

SIZ1 is involved in the metabolism of inorganic compounds. Indeed, *siz1* mutants were first described due to their phosphate hypersensitive phenotype. Though P_i levels in *siz1* mutants are comparable to those in wild type plants, they show reduced primary root growth and an increased number of lateral roots and root hairs, increased root/shoot mass ratio and increased anthocyanin content when not supplemented with phosphate (Miura et al. 2005). Nitrogen metabolism is also affected in *siz1* mutants: supplementation with a source of ammonia allowed for considerable recovery of the dwarfism and *Pseudomonas syringae* resistance phenotypes, decreasing the accumulation of salicylic acid to wild-type levels. The need for ammonia supplementation was found to be due to the lack of NITRATE REDUCTASE 1 and 2 (NIA1/2) activity in *siz1* mutants. These nitrogen reductases are SUMOylated by SIZ1, a modification which greatly increases their activity (Park et al. 2011).

Sugar metabolism is also affected in *siz1* mutants. Metabolite analysis showed they accumulate lower levels of xylose, maltose and especially glucose (Tomanov et al. 2014), while also showing glucose hypersensitivity (Castro et al. 2015). The root hairs of *siz1* mutants exhibit basal bulges characteristic of plants grown on high concentrations of glucose and show an increase in post-germination growth in response to glucose which was able to be decoupled from the lower osmotic tolerance which was also observed (Castro et al. 2015).

siz1 mutants also show reproductive defects. While pollen tubes develop normally, their guidance is impaired. The pollen tube therefore does not always reach the embryonic sac. Additionally, 20 % of embryonic sacs suffer distortion due to collapse of the structure. Both these effects lead to the production of smaller siliques which harbour an increased number of aborted seeds (Ling et al. 2012).

The germination of fresh *siz1* seeds is both late and scarce, a phenotype which was found not to be recovered by ammonia supplementation. However, either a four-week after-ripening period, a one week after-ripening period in combination with stratification treatment or exogenous gibberellin treatment did recover germination to wild-type levels. The mechanism behind this delayed germination has been elucidated in part. Firstly, *siz1* seeds show upregulated transcription of both *DELAY OF GERMINATION 1 (DOG1)* a number of *DELLA* genes, which are known to delay germination and inhibit growth. Secondly, the seeds showed both lower levels of both SLEEPY 1 (SLY1) protein and *SLY1* transcript. SLY1 is part of a ubiquitin E3 ligase complex which ubiquitinates the DELLA proteins, derepressing growth, but it required stabilisation by SIZ1 SUMOylation in order to escape degradation and mediate germination. However, as seeds expressing a non-SUMOylatable form of SLY1 exhibit and even

greater delay in germination than *siz1* mutants, other, as yet undiscovered mechanisms must play a role in the germination process of *siz1* mutants (Kim et al. 2015; S. Il Kim et al. 2016). For example, germination of *siz1* mutants was found to be significantly more repressed in response to osmotic stress than Col-0 and the mechanism regulating this is still unknown.

SIZ1 is undoubtedly involved in plant stress signalling, as is testified by its interactions with a number of proteins such as SnRK1 (Crozet et al. 2016), SUPPRESSOR OF *npr1-1* CONSTITUTIVE 1 (SNC1) (Hammoudi et al. 2018), ICE1 (K. Miura et al. 2007) and the TCP transcription factors (Mazur et al. 2017), all of which are known to be involved in plant immunity or stress responses.

However, the mechanistics of SIZ1's role in the plant response to abiotic stress is currently very much unclear. *siz1* mutants have been reported to be more (Kim et al. 2017; Miura et al. 2013) and less (Catala et al. 2007) tolerant to drought and to accumulate more (Miura et al. 2005) and fewer (Catala et al. 2007) anthocyanins. Depending on the heat treatment, *siz1* mutants have been reported to be less (Cheong et al. 2009), more, or equally tolerant to heat (Kim et al. 2017) when compared to wild-type plants.

Additionally, both *siz1* mutants and SIZ1 overexpressing lines have been found to be cold sensitive (K. Miura et al. 2007; Miura & Nozawa 2014) and salt tolerant (Miura et al. 2011; Miura & Nozawa 2014).

SIZ1 is itself a target for post-translational modifications: in the presence of E1 and E2 enzymes, SIZ1 self-SUMOylates (Garcia-Dominguez et al. 2008). SIZ1 is deSUMOylated by ESD4 and SUMOylation is thought to protect SIZ1 from ubiquitination by COP1 (J. Y. Kim et al. 2016; Kim et al. 2017). However, SIZ1 is also targeted for degradation by the SUMO-targeted ubiquitin ligase SYNTHETIC LETHAL OF UNKNOWN FUNCTION PROTEIN 5/8 (SLX5/SLX8) (Westerbeck et al. 2014).

7.1.4 The E4 SUMO ligases: PIAL1 and PIAL2

The last two *Arabidopsis* SUMO ligases to have been described are PIAL1 and PIAL2. These were identified based on the SP-RING motif they share with SIZ1 and HPY2. However, in contrast to SIZ1 and HPY2, they do not facilitate monoSUMOylation. Rather, they improve the formation of polySUMO chains and were therefore termed E4 ligases. Most of the experiments carried out in order to characterise PIAL1 and PAIL2 were performed using a truncated form of PIAL2, PIAL2M. In the presence of SAE1 and SCE1, PIAL2M, was able to increase both the quantity and length of SUMO1 and SUMO3 chains.

This *in vitro* interaction with both paralogs begs the question as to whether it is conserved *in vivo* and if so, whether mixed chains can be constructed (Tomanov et al. 2014).

The SP-RING domain, the known docking site for SCE1, and the region directly N-terminal to it are of great importance to the activity of PIAL2, although they are not strictly essential. However, deletion of either in combination with the loss of SIM1, one of the two SUMO sites present in PIAL2M, proved detrimental to PIAL2M activity. Interestingly, loss of either SIM site individually did not affect PIAL2M function. PIAL2M also has two confirmed SUMO sites. Surprisingly, K/R mutants for both these SUMO sites were more proficient at forming high molecular weight SUMO chains than wild-type PIAL2M, indicating that SUMOylation of PIAL2 might negatively regulate its activity (Tomanov 2014).

Remarkably, *pial1/2* mutants accumulate wild-type levels of SUMO conjugates as opposed to the lower levels of SUMO conjugates observed in E3 ligase mutants *siz1* and *hpy2.* This observation has led to the hypothesis that the addition of SUMO chains to a substrate is a mechanism for deSUMOylation -possibly though the SUMO targeted ubiquitin ligases (Elrouby et al. 2013)- rather than for SUMOylation (Tomanov et al. 2014).

Though no GUS data is available for *PIAL1* and *PIAL2*, PCR results indicate both ligases are widely expressed, with *PIAL1* transcript generally being more abundant. Both transcripts are present in cauline leaves, young rosette leaves, flowers and stems. Low levels of *PIAL1* transcript were additionally detected in siliques and old rosette leaves. Both genes are also stress responsive; with *PIAL1* transcription increasing in response to heat shock and *PIAL2* transcript levels increasing transiently in response to both osmotic and salinity stress (Tomanov et al. 2014).

Phenotypically, neither single *pial1* or *pial2*, or *pial1/2* double mutants exhibit a clear phenotype when grown in non-stress conditions. However, loss of either or both PIAL ligases in the *siz1* mutant does exacerbate the stunted *siz1* mutant phenotype. When confronted with stress, *pial* mutants are both salt tolerant and drought sensitive, with double mutants showing each trait to higher intensity than either single mutant (Tomanov et al. 2014). Interestingly this phenotype is the exact opposite to that observed in *ots1/2* mutants (Conti et al. 2008; Srivastava, Zhang, Yates, et al. 2016; Srivastava et al. 2017). Similarly to the observations in control conditions, additional loss of either or both PIAL proteins exacerbates the salt, drought and ABA sensitivity of *siz1* mutants (Tomanov et al. 2014).

While this is additive effect is expected in the case of drought tolerance, it is less straightforward in the case of salt tolerance. These phenotypes indicate that there is a complex interplay between SUMOylation and salinity tolerance.

The PIALs have also been shown to be implicated in sulphur metabolism: *pial1/2* mutants show increased sulphate, cysteine and glutathione content. As *siz1* mutants also show differential regulation of sulphur metabolism, though in a different way to *pial1/2* mutants, it is likely that there is an as yet undiscovered link between SUMOylation and the regulation of sulphur metabolism (Tomanov et al. 2014).

Another process which PIAL1 and PIAL2 have been implicated in is MORPHEUS MOLECULE 1 (MOM1)-mediated gene silencing through the formation of a PIAL-MOM1 complex. Interestingly, catalytic activity of PIAL1 and PIAL2 is not necessary for interaction with MOM1 or for silencing of at least a subset of MOM1 targets. However, a number of non-tested MOM1 targets are known to be SUMOylated. There may therefore be a role for PIAL1/2 as SUMO ligases rather than merely as interaction partners in the silencing of specific targets (Han et al. 2016).

As illustrated above, all four SUMO3 ligases are involved in stress tolerance. However, the data available mainly details interactions between the ligases and their substrates and regulation of expression at the protein level. I therefore chose to include the genes encoding SUMOylation enzymes in this study, examining the transcriptional effects of SUMO conjugate accumulation, either though SUMO overexpression in Col-0 SOX L2 and L8 lines, though loss of deSUMOylation in *ots1/2* mutants or though abiotic stress treatments on these genes.

7.2 RESULTS

7.2.1 Differential transcription of genes encoding SUMOylation enzymes in 10 day-old seedlings

RT-qPCR was used to compare the transcript levels of the genes encoding SUMOylation enzymes in 10 day-old seedlings grown on unsupplemented MS medium. The results, expressed using $-\Delta C_T$ as a metric, are shown in figure 26. Though this metric can act as a gauge to express whether a given transcript is more or less abundant than another, numerical differences should not be interpreted directly.

Remarkably, there was a large degree of variation in the transcription levels of the various SAE1 subunits. The gene encoding the large subunit, *SAE2,* exhibited the lowest transcription level of all SUMOylation genes. Additionally, there is a large disparity in transcription levels between the small subunit isoforms. Even though *SAE1B* transcript is generated by two identical loci, *SAE1A* is strongly dominant at the transcriptional level in 10 day-old seedlings.

The highest transcript levels within the SUMOylation system were obtained for *SCE1* and *HPY2*. Transcription of the other E3 ligase gene, *SIZ1,* is lower than that of *HPY2.* Meanwhile, though *PIAL2* seemingly has higher transcript levels than *PIAL1*, the transcript levels of both E4 ligase genes fall within the same ballpark.

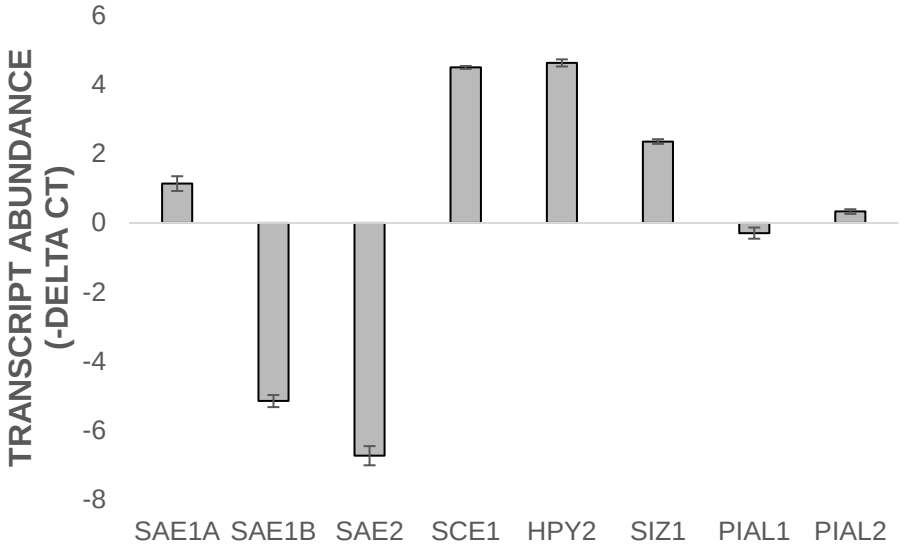

Figure 26: ***Differential transcription of genes encoding SUMOylation enzymes in 10 day-old seedlings****. Transcription levels of genes in the SUMOylation system vary widely. The largest differences are found between the genes encoding SAE1 subunits. A substantial difference in transcript abundance is also present between the two gene encoding E3 ligases. Error bars represent standard error based on three biological and two technical repeats. Ct values were normalised to UBC9 and CBP20 housekeeping genes.*

7.2.2 Combined influence of SUMO overexpression and abiotic stress on transcription in the SUMOylation system

In order to gain a general impression of SUM transcription levels in the different genotypes and stress conditions tested, all values were compared to the Ct values obtained for Col-0 WT in MS conditions. The resulting differences in transcript abundance are shown in figure 27. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005. However, variation between biological replicates was sometimes considerable and where highly significant data suggested a pattern, $p<0.1$ significance results are also described.

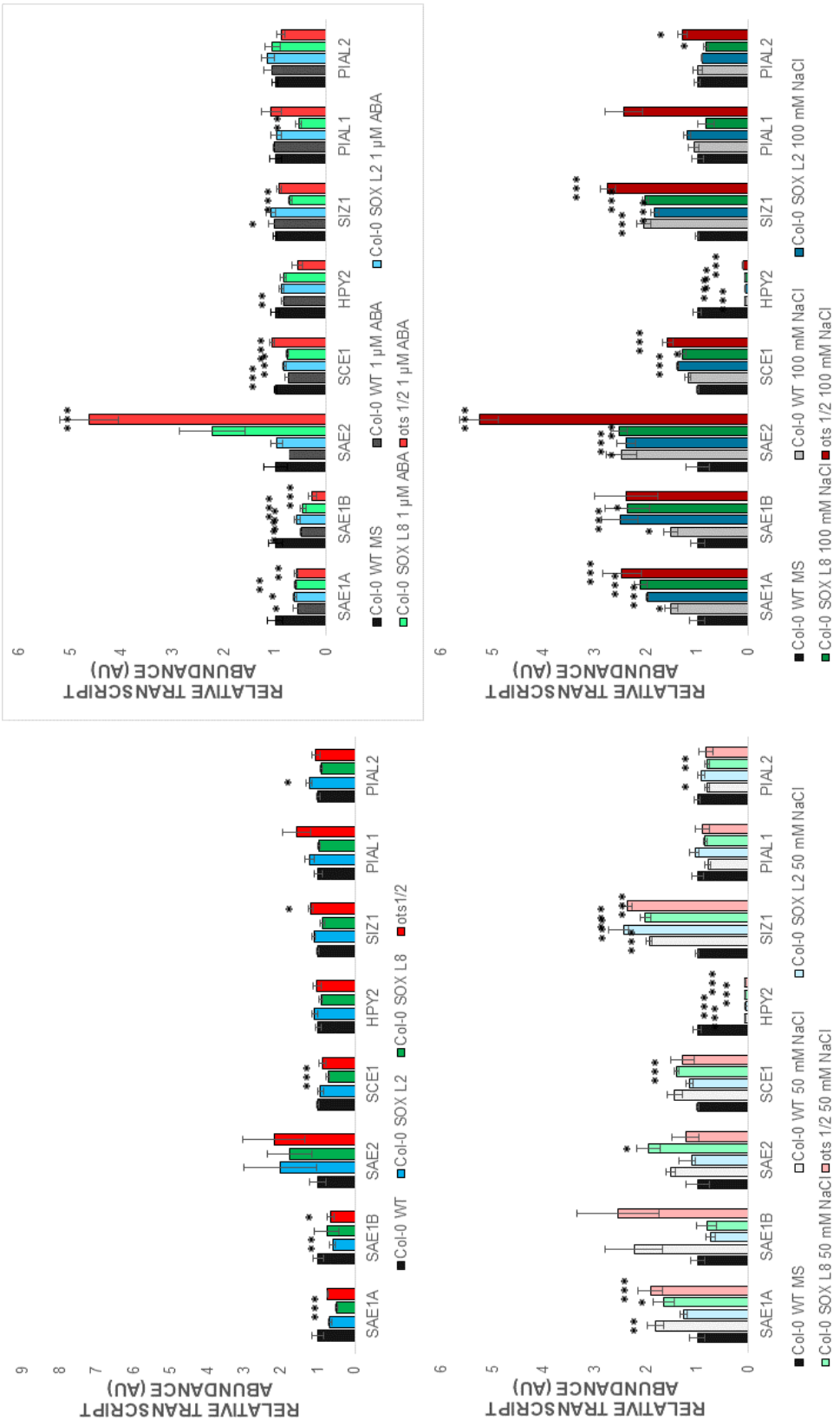

Figure 27: Combined effect of SUMO overexpression and abiotic stress on transcription in the SUMOylation system. *Differences in transcript levels of genes encoding SUMOylation enzymes in 10 day-old seedlings grown on A: unsupplemented MS medium B: MS medium supplemented with 1 µM ABA C: MS medium supplemented with 50 mM NaCl D: MS medium supplemented with 100 mM NaCl.*

***Figure 27, continued**: All Ct values were compared to the Ct values obtained for Col-0 WT grown in MS conditions. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

Overall, the genes encoding SUMOylation enzymes seem to form a heterogenous group as far as transcriptional regulation is concerned: some exhibit strong up- or downregulation, while others show few, if any, differences in transcript abundance.

In the case of *SAE1A,* Col-0 SOX L8 plants grown on unsupplemented MS showed a reduction in *SAE1A* transcription when compared to Col-0 WT plants. This reduction was also present in Col-0 SOX L2, although it was only significant to p<0.1. When plants were treated with 1 µM ABA, this downregulation was generalised to all genotypes. However, NaCl treatment showed the opposite effect. Plants grown in the presence of 50 mM NaCl showed upregulation of *SAE1A,* although the rise in transcript levels was not significant in Col-0 SOX L2. Meanwhile, 100 mM NaCl treatment caused an increase in the upregulation, which was now significant for all genotypes.

For *SAE1B,* Col-0 SOX L2 and *ots1/2* mutant plants in control conditions accumulated less *SAE1B* transcript than Col-0 WT plants. Col-0 SOX L8 plants showed a very large variation between replicates, possibly due to technical error. 1 µM ABA treatment caused a generalised downregulation of *SAE1B,* while 50 mM NaCl treatment had no effect. In contrast, 100 mM NaCl treatment led to upregulation of *SAE1B* in all genotypes except in the *ots1/2* mutant.

SAE2 showed no differences between genotypes for plants grown on unsupplemented MS, though there was a large amount of variation between replicates in all cases. 1 µM ABA treatment led to upregulation of the gene in *ots1/2* mutants, while 50 mM NaCl treatment upregulated it only in Col-0 SOX L8 plants. Meanwhile, plants grown in the presence of 100 mM NaCl showed a generalised increase in *SAE2* transcript abundance.

In control conditions, *SCE1* was only downregulated in Col-0 SOX L8 plants, but this downregulation expanded to Col-0 WT and both Col-0 SOX lines in the presence of 1 µM ABA. A similar pattern was visible in the salinity treatments: 50 mM NaCl treatment led to an increase in *SCE1* transcription only in Col-0 SOX L8 plants, while after 100 mM NaCl treatment it was present in both Col-0 SOX lines and in *ots1/2* mutants.

The most distinct change of all was visible for the gene encoding E3 ligases HPY2. For *HPY2* there were no differences in transcript levels between the genotypes in control conditions and 1 µM ABA treatment only affected *HPY2* transcription in Col-0 WT plants, which it downregulated. However, treatment with either 50 or 100 mM NaCl caused a large, 12 to 15-fold reduction in *HPY2* transcript abundance.

In the case of *SIZ1, ots1/2* plants grown on unsupplemented MS showed an increase in *SIZ1* transcript levels when compared to Col-0 WT plants, while 1 µM ABA treatment led to a marginal increase in transcription in Col-0 WT plants and a decrease in Col-0 SOX L8 plants. However, treatment with either 50 mM or 100 mM NaCl again caused a widespread impact, leading to an approximately twofold upregulation in all samples.

The only sample which showed a difference in *PIAL1* transcript abundance was that of Col-0 SOX L8 plants grown in the presence of 1 µM ABA. Meanwhile, transcriptional regulation of *PIAL2* was also limited. In control conditions, only Col-0 SOX L2 plants showed an increase in the transcript levels of *PIAL2*, while 1 µM ABA treatment did not cause any variation in transcript abundance. In the presence of 50 mM NaCl, a slight downregulation was visible in Col-0 WT and Col-0 SOX L8 plants. This slight downregulation was conserved for Col-0 SOX L8 plants in the presence of 100 mM. Additionally, it led to a slight increase in *PIAL2* transcript in *ots1/2* mutant plants.

7.2.3 Influence of SUMO overexpression on transcription in the SUMOylation system

In order to assess the effect of SUMO levels on the transcription of genes encoding SUMOylation enzymes, Ct values for each growth condition were compared to values obtained for Col-0 WT plants grown in that condition. The resulting differences in transcript abundance are shown in figure 28. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005. However, variation between biological replicates was sometimes considerable and where highly significant data suggested a pattern, $p<0.1$ significance results are also described.

In the case of *SAE1A*, when plants were grown on unsupplemented MS a reduction in transcript abundance was visible in Col-0 SOX L8 plants. In this growth condition, Col-0 SOX L2 plants also showed *SAE1A* downregulation, although this was only significant to $p<0.1$. Genotype had no influence on transcription in plants grown in the presence of 1 µM ABA. 50 mM NaCl treated Col-0 SOX L2 plants exhibited a reduction in *SAE1A* transcription when compared to Col-0 WT plants subject to the same treatment.

This is consistent with the lack of overall upregulation of *SAE1A* for Col-0 SOX L2 plants grown in the presence of 50 mM NaCl discussed above (figure 27C). Meanwhile, in 100 mM NaCl conditions, both Col-0 SOX lines and *ots1/2* mutants showed further upregulation of *SAE1A* when compared to Col-0 WT plants.

For *SAE1B*, a reduction in transcription was observed for Col-0 SOX L2 and *ots1/2* mutant plants grown in control conditions. Due to a large variability between replicates, any differences present in Col-0 SOX L8 were not significant. 50 mM NaCl treatment did not differentially affect the genotypes studied, while both 1 µM ABA and 100 mM NaCl treatment led to a higher level of *SAE1B* transcription in Col-0 SOX L2 plants than in Col-0 WT plants, reflecting a decrease in overall downregulation after 1 µM ABA treatment and an increase overall upregulation after 100 mM NaCl treatment (figure 27 B,D).

For *SAE2,* no differences between genotypes were visible in either control or 50 mM NaCl conditions. However, when plants were treated with 1 µM ABA, *SAE2* was upregulated both Col-0 SOX lines and in *ots1/2* mutants, though the upregulation was only significant to $p<0.1$ in Col-0 SOX L2. Meanwhile, 100 mM NaCl treatment led to an increase in *SAE2* transcription in *ots1/2* mutant plants.

In control conditions, *SCE1* transcription was lower in Col-0 SOX L8 plants than in Col-0 WT plants. In plants grown on medium supplemented with 1 µM ABA, both Col-0 SOX lines and *ots1/2* mutants showed less *SCE1* transcription than Col-0 WT plants grown in the same conditions, corresponding to a reduction in overall downregulation (figure 27 B). No differential transcription was observed in plants grown in the presence of 50 mM NaCl, while 100 mM treatment only led to an increase in transcription in *ots1/2* mutant plants.

No differences in the levels of *HPY2* transcript were observed between genotypes in either control or 50 mM NaCl treatment conditions. An increase in transcription when compared to Col-0 WT plants grown in the respective conditions was observed in Col-0 SOX L2 plant subjected to 1 µM ABA treatment and *ots1/2* mutant plants subjected to 100 mM NaCl treatment.

For *SIZ1*, transcription was upregulated in *ots1/2* mutant plants grown on both unsupplemented MS and MS supplemented with either 50 or 100 mM NaCl. However, 1 µM ABA treatment led to *SIZ1* upregulation in Col-0 SOX L2 but not *ots1/2* mutant plants.

Similarly to *HPY2,* no differences in the abundance of *PIAL1* mRNA were observed between genotypes in either control or 50 mM NaCl treatment conditions. In plants treated with 1 µM ABA, Col-0 SOX L8 plants showed a decrease in *PIAL1* transcription when compared to Col-0 WT plants, while *PIAL1* transcription was increased in *ots1/2* plants treated with 100 mM NaCl.

Meanwhile, *PIAL2* transcription was only significantly affected in Col-0 SOX L2, where it was increased. An increase in PIAL1 transcription was also visible *ots1/2* mutant plants grown in the presence of 100 mM NaCl. However, this increase was only significant to $p<0.1$.

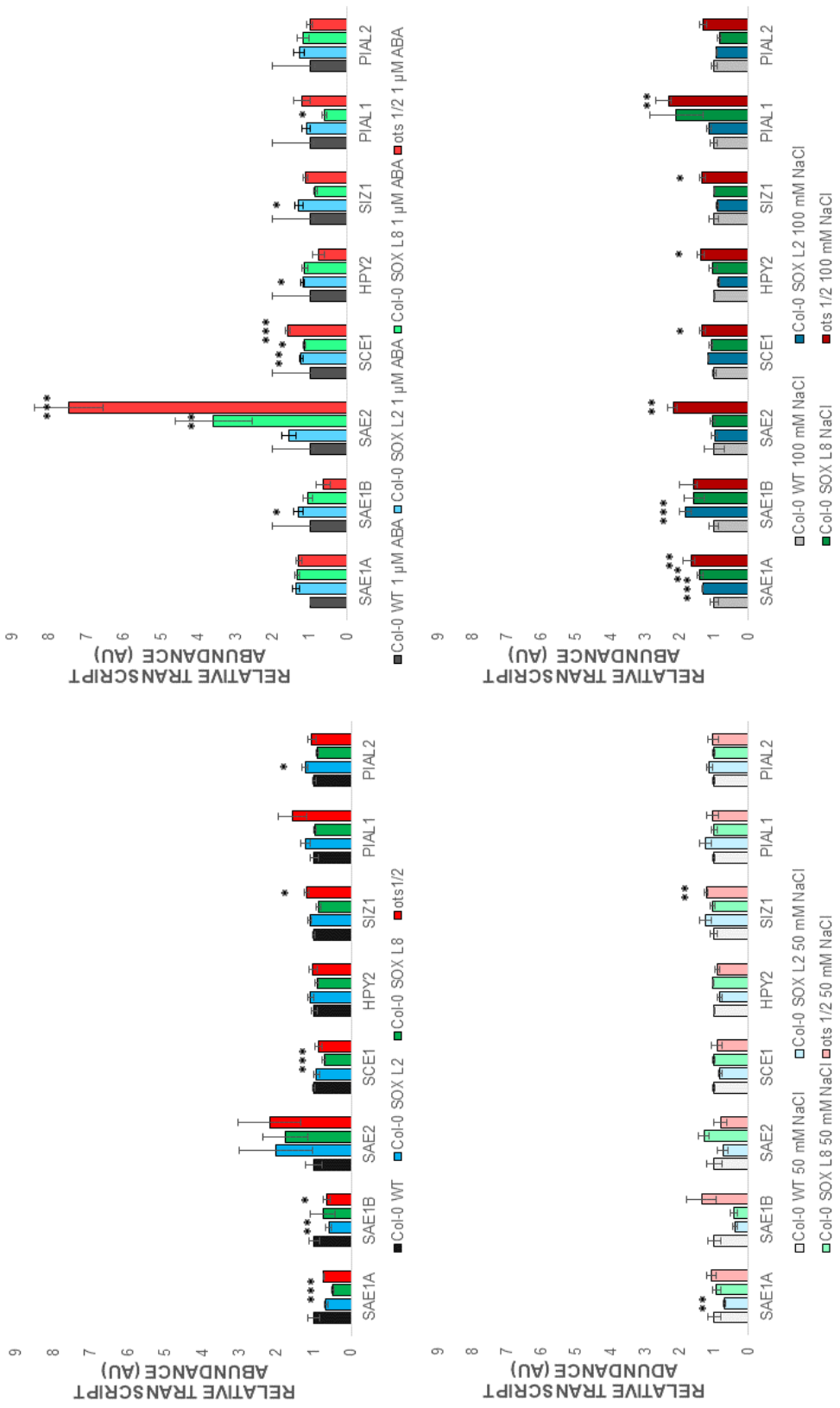

Figure 28: Effect of genotype on transcription in the SUMOylation system. *Differences in transcript levels of genes encoding SUMOylation enzymes in 10 day-old seedlings grown on A: unsupplemented MS medium B: MS medium supplemented with 1 µM ABA C: MS medium supplemented with 50 mM NaCl D: MS medium supplemented with 100 mM NaCl.*

Figure 28, continued: *All Ct values were compared to the Ct values obtained for Col-0 WT grown in MS conditions. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

7.2.4 Influence of abiotic stress on transcription in the SUMOylation system

In order to assess the effect of abiotic stress on *SUM* gene transcription, Ct values for each growth condition were compared to values obtained for plants of that genotype grown on unsupplemented MS. The resulting changes in transcript abundance are shown in figure 29. In general, significance was assessed to p-values of 0.05, 0.01 and 0.005.

A clear pattern was visible for *SAE1A.* 1 µM ABA treatment caused downregulation of *SAE1A* in Col-0 WT plants. In *ots1/2* mutants, this reduction in transcription was still present, though smaller. Meanwhile, Col-0 SOX L2 plants showed no significant change in comparison to plant grown on unsupplemented MS. In Col-0 SOX L8 plants, the increase in comparison to Col-0 WT plants subjected to the same 1 µM ABA treatment was large enough for it to result in a significant upregulation of *SAE1A* when compared to Col-0 SOX plants grown in control conditions. NaCl treatment caused upregulation of *SAE1A* in all genotypes, with 100 mM treatment leading to a slightly larger increase in all genotypes except Col-0 WT. Additionally, the extent of upregulation was smallest in Col-0 WT plants and progressively greater in Col-0 SOX L2, *ots1/2* and Col-0 SOX L8 plants.

For *SAE1B,* 1 µM ABA treatment led to a decrease of *SAE1B* transcription in Col-0 WT and *ots1/2* mutant plant but not in the Col-0 SOX lines. 50 mM NaCl treatment only affected *SAE1B* transcript levels in *ots1/2* mutants, while 100 mM NaCl treatment caused upregulation of *SAE2* in all genotypes.

Abiotic stress did not influence SAE2 transcription in either of the Col-0 SOX lines. However, in both Col-0 WT and *ots1/2* mutant plants, 100 mM NaCl treatment resulted in an increase in *SAE2* transcription.

In the case of *SCE1,* 1 µM ABA treatment only influenced transcription in Col-0 WT plants, which it decreased. The increase in transcription in Col-0 SOX and *ots1/2* plants described above (Figure 28 B) can therefore be more accurately described as a decrease in downregulation, as in visible when the overall effect is assessed (Figure 27 B). 50 mM and 100 mM NaCl treatment increased *SCE1* transcription in both Col-0 SOX lines, but not Col-0 WT, while in *ots1/2* mutants the increase in transcription was only significant the case of 100 mM NaCl treatment.

For *HPY2*, Col-0 WT and Col-0 SOX L2 exhibited a small decrease in HPY2 transcription in the presence of 1 µM ABA while the decreases in Col-0 SOX L8 and *ots1/2* mutant plants were not significant. NaCl treatment, both at 50 and 100 mM led to a sharp 12- to 15-fold decrease in *HPY2* transcription in all genotypes.

Similarly to *HPY2*, the transcriptional effect of NaCl treatment on *SIZ1* is similar for all genotypes under both NaCl concentrations: all undergo an approximately twofold upregulation. The effects of 1 µM ABA treatment were more diverse. In Col-0 WT plants, it caused a small increase in *SIZ1* transcription, while there was no significant effect in Col-0 SOX L2 plants. Meanwhile, Col-0 SOX L8 and *ots1/2* mutant plants both exhibited a small decrease in *SIZ1* transcription when grown in the presence of 1 µM ABA.

PIAL1 and PIAL2 undergo little transcriptional regulation in response to the abiotic stress treatments used in this study. PIAL1 transcription is only affected in Col-0 SOX L8 plants grown in the presence of 1 µM ABA, where it is decreased. In contrast, *PIAL2* transcription only shows a small reduction in Col-0 WT and Col-0 SOX L2 plants grown in the presence of 50 mM NaCl, while Col-0 SOX L2 plants show the same reduction after 100 mM NaCl treatment.

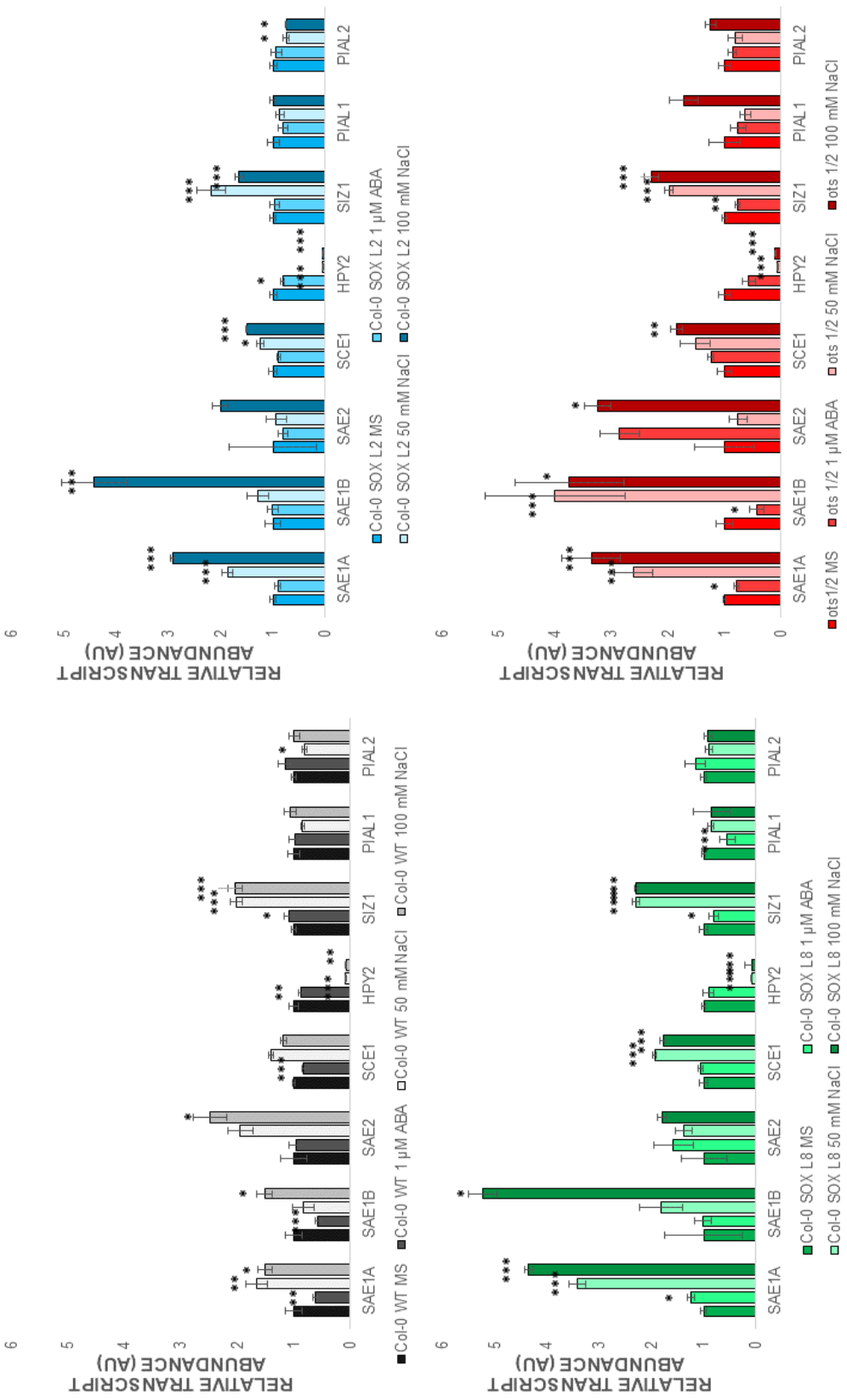

Figure 29: Influence of abiotic stress on transcription in the SUMOylation system. *Changes in transcript levels of genes encoding SUMOylation enzymes in 10 day-old seedlings of several genotypes after 6 days of treatment with 1 µM ABA, 50 mM NaCl and 100 mM NaCl. A: Col-0 WT, B: ots1/2 mutant, C: Col-0 SOX L2, D: Col-0 SOX L8.*

Figure 29, continued: *Ct values for each genotype were compared to the Ct values obtained for that genotype when grown on unsupplemented MS. Error bars represent standard error based on three biological and two technical repeats. Significance was assessed using the Mann-Whitney U-test. Significance values: * = P < 0.05; ** = P < 0.01; *** = P < 0.005. SOX: SUMO overexpressor. OTS: Overly tolerant to salt.*

7.3 DISCUSSION

7.3.1 SUMO overexpression and abiotic stress result in a variety of transcript level modifications in the SUMO system

Although the mechanisms regulating the transcription of genes encoding SUMOylation enzymes may not be as clear cut as those regulating transcription of the *SUM* genes, a number different patterns are visible.

Firstly and in parallel with the mechanism proposed for the *SUM* genes in the previous chapter, SUMO overexpression may lead to an earlier manifestation of the stress response. Col-0 SOX L8 plants show a reduction in *SCE1* transcription when grown in control conditions (Figure 27 A), a reaction which is present in all genotypes when plants were exposed to 1 µM ABA (Figure 27 B). A comparable reaction takes place after 50 mM NaCl treatment where *SCE1* and *SAE2* transcript levels are both increased only in Col-0 SOX L8 (Figure 27 C), while this increase is again generalised after 100 mM NaCl treatment, although the upregulation in Col-0 WT plants was only significant to $p<0.1$. These results again suggest that SUMO overexpression can function as a proxy for increased abiotic stress under certain conditions.

The regulation of *SAE1A* seems more complex overall. Both Col-0 SOX lines showed downregulation of *SAE1A* in control conditions (Figure 27 A) and this effect was generalised when the plants were grown in the presence of 1 µM ABA (Figure 27 B). This reduction in transcription appears to be fixed to a set level, as Col-0 WT and *ots1/2* mutant plants both achieved this by downregulating *SAE1A* transcript levels in with respect to levels present in plants grown on unsupplemented MS (Figure 27 A, B). Meanwhile, Col-0 SOX L8 plants reduced the downregulation of *SAE1A* from the higher levels of transcriptional repression present in control condition in order to meet the same plateau (Figure 27 D).

The regulation of *SAE1A* in response to NaCl treatment appears to follow a different pattern. When the influence of NaCl treatment on *SAE1A* transcript abundance is assessed independently, all genotypes show an increase in *SAE1A* transcription in response to both concentrations. Except for in Col-0 WT plants, the increase in *SAE1A* transcription was more pronounced in plants subjected to 100 mM NaCl. Indeed, these differences were also visible when the impact of genotype on plant reaction to 100 mM NaCl treatment was assessed

(Figure 28 D). Overall, the largest upregulation of *SAE1A* in response to NaCl treatment took place in Col-0 SOX L8 and *ots1/2* mutants and the smallest in Col-0 WT plants. However, when the overall effect of 50 mM treatment on *SAE1A* transcription was considered, the changes in the Col-0 SOX lines were both smaller and less significant than those in Col-0 WT and *ots1/2* mutant lines (Figure 27 C). It is tempting to speculate that this may indicate that SUMO overexpression leads to a slower transition from downregulation to upregulation compared to Col-0 WT plants grown on unsupplemented conditions in the case of *SAE1A* rather than the faster transition described above for *SCE1*.

Interestingly, the clearest case of generalised influence of genotype on the reaction to abiotic stress is visible in the response to 100 mM NaCl when *ots1/2* mutant plants showed upregulation of all genes encoding SUMOylation enzymes except *SAE2*, though the increase in *PIAL2* transcription was only significant to $p<0.1$. As this generalised upregulation was not present in either of the Col-0 SOX lines, both of which also accumulate SUMO conjugates, this effect is likely to be specific to the lack of OTS1/2 function or *OTS1/2* transcription in *ots1/2* mutants. When the *esd4* mutant was characterised, a similar upregulation of SUMOylation was postulated (Villajuana-Bonequi et al. 2014). If this pattern holds true generally, it opens up a new perspective on possible regulatory interactions between SUMOylation and deSUMOylation.

7.3.2 Altering the balance between E3 SUMO ligases may provide a mechanism for the response to salinity stress

When plants were grown on medium containing NaCl, whether it be the moderate 50 mM NaCl or the high 100 mM NaCl concentration, a change in the abundance of *HPY2* and *SIZ1* transcripts was observed (Figure 26, 28). This change, which was present in all genotypes, involved the drastic 12- to 15-fold downregulation of *HPY2*, combined with the approximately twofold upregulation of *SIZ1*. Indeed, though *HPY2* transcript is present at higher levels than *SIZ1* mRNA in plants grown on unsupplemented MS medium, supplementation with at least 50 mM NaCl shifts this balance so *SIZ1* becomes the transcriptionally dominant E3 ligase.

Of the two enzymes, only SIZ1 has been studied in the context of salinity stress. While *siz1* mutants did show increased tolerance to NaCl (Miura et al. 2011), overexpressing SIZ1 also led to increased salt tolerance. These results are easily comparable because *SIZ1* transcription in overexpressor lines only reached approximately twice the level observed on average in control plants (Miura & Nozawa 2014). This inability to produce lines expressing *SIZ1* transcript at high levels, coupled with the observation that induction plateau observed in this study was

independent of both genotype and NaCl concentration, suggest that *SIZ1* transcription may be tightly regulated *in planta* to avoid any deleterious effects caused by overSUMOylation, as it is already known that SIZ1 protein accumulation is regulated through multiple post-translational modifications (Garcia-Dominguez et al. 2008; Westerbeck et al. 2014; Kim et al. 2017).

Though there is no data available on the relationship between HPY2 and salt stress, there is some information on the relationship between both E3 ligases and drought stress. Most notably, *hpy2* mutants are known to be drought tolerant (Zhang et al. 2013). However, no data is available on the drought tolerance of SIZ1 overexpressor lines and as *siz1* mutants have been reported to both more and less sensitive to drought depending on the specific treatment (Catala et al. 2007; Miura et al. 2013; Kim et al. 2017), there is no basis upon which to speculate on what their behaviour may be.

However, as the osmotic stress which is also associated with drought is a major component of salinity stress, it is tempting to speculate that *hpy2* mutants may also exhibit increased tolerance to NaCl. If this is the case, it is tempting to speculate HPY2 negatively regulates salt tolerance, possibly though a stronger involvement in growth and developmental processes as evidenced by its role in meristematic maintenance. SIZ1 may positively regulate salt tolerance though an as yet unknown mechanism. Transcriptionally downregulating *HPY2* and upregulating *SIZ1* could therefore be a meta-level mechanism to increase plant NaCl tolerance by affecting which SUMOylation targets are more readily modified.

The arguments upon which this hypothesis is built have the caveat the role of ionic toxicity in salinity stress should not be overlooked: drought tolerance is not always a good proxy for salinity tolerance, as is demonstrated by the *Arabidopsis ots1/2* and rice *OTS1*-RNAi lines which are drought tolerant but salt sensitive (Conti et al. 2008; Srivastava, Zhang, Yates, et al. 2016; Srivastava, Zhang & Sadanandom 2016), while *pial1/2* mutants are salt tolerant but drought sensitive (Tomanov et al. 2014). It must also be noted that this mechanism appears to be NaCl-specific. 1 µM ABA treatment, which caused equivalent changes in transcription of the *SUM* genes to those elicited by 50 mM NaCl treatment (Figure 23), did not result in the same transcriptional changes when it comes to *HPY2* and *SIZ1* (Figure 28).

7.3.3 Changes is SAE2 integrity may allow SUMOylation outside the nucleus

A review of the literature, harbours a number of interesting findings related to the subcellular localisation of the SUMOylation enzymes. SAE1a, SAE1b and SCE1 all localise to both the nucleus and the cytoplasm, though all have an individual balance between the two. However, they all relocalise exclusively to the nucleus when co-expressed with SAE2 (Castaño-Miquel et al. 2013; Castaño-Miquel et al. 2017). While SIZ1 localised only to the nucleus (Miura et al. 2005), HPY2 is present in both the nucleus and the cytoplasm (Huang et al. 2009). It is unknown whether their localisations are influenced by the presence of SAE2 or SCE1. Additionally, the subcellular localisations of PIAL1 and PIAL2 are as yet unknown.
It is remarkable that the nuclear localisation of the very core SUMOylation components is determined by SAE2, the SUMOylation enzyme with the least abundant transcript (Figure 26). The relative levels of SAE1a/b, SAE2 and SCE1 have not been determined at protein level. However, if they do reflect the imbalance seen at the transcript level, the co-expression assay which do not take this possibly vastly different stoichiometry into account may not portray entirely accurate results.

Additionally, SAE2 is known to be processed *in vivo* in certain conditions, leading to the removal of one of the C-terminal nuclear localisation signals. This modified SAE2 localises mostly to the nucleus, but is also expressed at low levels in the cytoplasm (Castaño Miquel 2015). If SAE2 is indeed instrumental in determining the localisation of the core SUMOylation enzymes, this small change in subcellular distribution may also cause the transfer of some of the SUMOylation activity to the cytoplasm, where it might, for example, be responsible for the small SUMOylation signal that is present in the cytoplasm after heat shock treatment (Saracco et al. 2007).

8 GENERAL DISCUSSION

This study has revealed a number of mechanisms regulating the differences in transcript abundance in the *Arabidopsis* SUMO system in response to abiotic stress and SUMO overexpression. However, as it focusses on transcriptional effects, the response at protein level and subsequently its correlation – or lack of correlation – with the differences in transcription have not been explored.

Additionally, the choice of model system must be taken into account. This study focussed on seedlings which were exposed to the abiotic stress condition for six days and had thus reached a new metabolic equilibrium. Due to lack of time, it could not be expanded this to cover the initial "shock response" to stress treatment.

One of the main questions which remains open in the SUMOylation field is that of specificity. In the ubiquitin system, only one modifier is present and specificity is thought the be conferred by the plethora of E2 and E3 ligases encoded by *Arabidopsis* (Perry et al. 2008), with an additional layer of information encoded in the length and structure of ubiquitin chains (Walsh & Sadanandom 2014). Though this last aspect may also play a role in the SUMOylation system, it differs greatly with respect to the first. Further studies should address this issue of specificity, as lies at the core of our understanding of post-translational modification systems.

Using transcriptional analysis to characterise the complete SUMOylation and deSUMOylation system in *Arabidopsis* and its response to stress and SUMO overexpression, the three components of the system were found to show distinct degrees of transcriptional regulation.

First of all, the *ULP*s, which encode the ULP SUMO proteases, showed very few instantly recognisable patterns in response to either stress or SUMO overexpression (Figure 17, 18). Interestingly, even within the phylogenetically related groups (ESD4, OTS, ULP2 see figure 12), there were generally few similarities between the regulation patterns. Additionally, the variability between biological repeats was the largest in this set of genes. Each *ULP* gene appears to have its own individual regulation pattern, an observation which strengthens the case for deSUMOylating enzymes as a source specificity within the SUMO system (Yates et al. 2016). However, very little is currently known about the proteins each ULP targets. High-throughput target interaction studies such as the one recently performed for ESD4 could provide a base for further studies into the conditions governing the deSUMOylation process.

Secondly, the canonical *SUM* genes showed very tight transcriptional regulation and a strong, clear downregulation pattern in their response to both abiotic stress and SUMO overexpression (Figure 22, 23). The data presented here indicated that in the case of *SUM* gene regulation, SUMO overexpression can act as a proxy for abiotic stress (Figure 21, 23).

Finally, the genes encoding SUMOylation enzymes exhibit an intermediate level of transcriptional regulation. For example, the E3 ligase genes *HPY2* and *SIZ1* show a strong transcriptional regulation pattern in response to salinity stress, while the E4 ligase genes *PIAL1* and *PIAL2* show very little transcriptional regulation across the board. It would be interesting to establish whether *PIAL1* and *PIAL2* are transcriptionally regulated by other abiotic stress conditions or whether their regulation takes place purely at the protein level. Again, for *SCE1* and some of the *SAE1* subunit genes, SUMO overexpression appears to act as a partial proxy for abiotic stress intensity (Figure 27, 29). Interestingly, although NaCl is known to lead to the degradation of OTS1 (Conti et al. 2008), the Col-0 WT and SOX lines did not necessarily show the same differences in transcription as *ots1/2* mutants (Figure 17 C, D; 22 C, D; 28 C, D). This may be due to an insufficiently high NaCl concentration for complete degradation of OTS1 or to semi-redundant OTS2 activity, as it is not known whether OTS2 suffers the same degradation.

Overall, the regulation of the SUMO system as a whole in response to stress is a complex balance. The most severe stress condition assayed, treatment with 100 mM NaCl, caused a generalised downregulation of SUMO paralog transcription in combination with an almost generalised upregulation of genes encoding SUMOylation enzymes. Interestingly, this upregulation was greater still in *ots1/2* mutants, which are deficient in deSUMOylation. This example proves that each of the three branches of the SUMO system forms an integral part of its functioning in the plant stress response and the integration of data on each of these branches and their interactions will be necessary in order to further elucidate the mechanisms behind the plant SUMOylation response.

Additionally, all three branches of the *Arabidopsis* SUMO machinery (SUMO paralogs, SUMOylation and deSUMOylation enzymes) include proteins which have been shown by mass spectrometry to be SUMOylation targets in themselves, opening up possibilities for the discovery of even more complex and fine-tuned processes within the SUMO system.

9 BIBLIOGRAPHY

Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), pp.3389–402.

Augustine, R.C. et al., 2016. *Defining the SUMO System in Maize: SUMOylation Is Up-Regulated during Endosperm Development and Rapidly Induced by Stress*, American Society of Plant Biologists.

Bachmair, A. et al., 2001. Ubiquitylation in plants: A post-genomic look at a post-translational modification. *Trends in Plant Science*, 6(10), pp.463–470.

Bailey, M. et al., 2015. Stability of small ubiquitin-like modifier (SUMO) proteases OVERLY TOLERANT TO SALT1 and -2 modulates salicylic acid signalling and SUMO1/2 conjugation in Arabidopsis thaliana. *Journal of experimental botany*, 67(1), pp.353–363.

Budhiraja, R. et al., 2009. Substrates Related to Chromatin and to RNA-Dependent Processes Are Modified by Arabidopsis SUMO Isoforms That Differ in a Conserved Residue with Influence on Desumoylation. *PLANT PHYSIOLOGY*, 149(3), pp.1529–1540.

van den Burg, H.A. et al., 2010. Arabidopsis Small Ubiquitin-Like Modifier Paralogs Have Distinct Functions in Development and Defense. *The Plant Cell*, 22(6), pp.1998–2016.

Bylebyl, G.R., Belichenko, I. & Johnson, E.S., 2003. The SUMO isopeptidase Ulp2 prevents accumulation of SUMO chains in yeast. *The Journal of biological chemistry*, 278(45), pp.44113–20.

Campanaro, A. et al., 2016. SUMO proteases OTS1 and 2 control filament elongation through a DELLA-dependent mechanism. *Plant Reproduction*, 29(4), pp.287–290.

Casey, M., Srivastava, M. & Sadanandom, A., 2017. Posttranslational Modifications in Plant Disease Resistance. *eLS*, pp.1–7.

Castaño-Miquel, L. et al., 2013. Diversification of SUMO-activating enzyme in arabidopsis: Implications in SUMO conjugation. *Molecular Plant*, 6(5), pp.1646–1660.

Castaño-Miquel, L. et al., 2017. SUMOylation Inhibition Mediated by Disruption of SUMO E1-E2 Interactions Confers Plant Susceptibility to Necrotrophic Fungal Pathogens. *Molecular Plant*, 10(5), pp.709–720.

Castaño-Miquel, L., Seguí, J. & Lois, L.M., 2011. Distinctive properties of Arabidopsis SUMO

paralogues support the in vivo predominant role of AtSUMO1/2 isoforms. *Biochemical Journal*, 436(3), pp.581–590.

Castaño Miquel, L., 2015. *Caracterización molecular y funcional de Small-Ubiquitin-related MOdifier en Arabidopsis thaliana*. Universitat de Barcelona.

Castro, P.H. et al., 2015. SIZ1-Dependent Post-Translational Modification by SUMO Modulates Sugar Signaling and Metabolism in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 56(12), pp.2297–2311.

Castro, P.H. et al., 2012. SUMO,a heavyweight player in plant abiotic stress responses. *Cellular and Molecular Life Sciences*, 69(19), pp.3269–3283.

Castro, P.H. et al., 2016. SUMO proteases ULP1c and ULP1d are required for development and osmotic stress responses in Arabidopsis thaliana. *Plant Molecular Biology*, 92(1–2), pp.143–159.

Castro, P.H.A.R.F., 2013. *Functional analysis of the SUMO conjugation/deconjugation system during the development and stress response of Arabidopsis thaliana*. Universidade do Minho.

Catala, R. et al., 2007. The Arabidopsis E3 SUMO Ligase SIZ1 Regulates Plant Growth and Drought Responses. *the Plant Cell Online*, 19(9), pp.2952–2966.

Chen, L. & Hellmann, H., 2013. Plant E3 Ligases: Flexible Enzymes in a Sessile World. *Molecular Plant*, 6(5), pp.1388–1404.

Cheong, M.S. et al., 2009. Specific Domain Structures Control Abscisic Acid-, Salicylic Acid-, and Stress-Mediated SIZ1 Phenotypes. *PLANT PHYSIOLOGY*, 151(4), pp.1930–1942.

Chosed, R. et al., 2006. Evolution of a signalling system that incorporates both redundancy and diversity: Arabidopsis SUMOylation. *Biochemical Journal*, 398(3), pp.521–529.

Colby, T. et al., 2006. SUMO-Conjugating and SUMO-Deconjugating Enzymes from Arabidopsis. *Plant Physiology*, 142(1), pp.318–332.

Conti, L. et al., 2009. OTS1 and OTS2 SUMO proteases link plant development and survival under salt stress. *Plant Signaling and Behavior*, 4(3), pp.225–227.

Conti, L. et al., 2008. Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in Arabidopsis. *The Plant cell*, 20(10), pp.2894–908.

Conti, L. et al., 2014. Small Ubiquitin-like Modifier Protein SUMO Enables Plants to Control Growth Independently of the Phytohormone Gibberellin. *Developmental Cell*, 28(1), pp.102–110.

Crozet, P. et al., 2016. SUMOylation represses SnRK1 signaling in Arabidopsis. *The Plant Journal*, 85(1), pp.120–133.

Dill, A., 2004. The Arabidopsis F-Box Protein SLEEPY1 Targets Gibberellin Signaling Repressors for Gibberellin-Induced Degradation. *THE PLANT CELL ONLINE*, 16(6), pp.1392–1405.

Elrouby, N. et al., 2013. Identification of Arabidopsis SUMO-interacting proteins that regulate chromatin activity and developmental transitions. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), pp.19956–61.

Flotho, A. & Melchior, F., 2013. Sumoylation: A Regulatory Protein Modification in Health and Disease. *Annual Review of Biochemistry*, 82(1), pp.357–385.

Fuller, D.Q. & Allaby, R., 2009. Seed Dispersal and Crop Domestication: Shattering, Germination and Seasonality in Evolution under Cultivation. In *Fruit Development and Seed Dispersal*. pp. 238–295.

Garcia-Dominguez, M., March-Diaz, R. & Reyes, J.C., 2008. The PHD domain of plant PIAS proteins mediates sumoylation of bromodomain GTE proteins. *The Journal of biological chemistry*, 283(31), pp.21469–77.

Gareau, J.R. & Lima, C.D., 2010. The SUMO pathway: Emerging mechanisms that shape specificity, conjugation and recognition. *Nature Reviews Molecular Cell Biology*, 11(12), pp.861–871.

Geiss-Friedlander, R. & Melchior, F., 2007. Concepts in sumoylation: a decade on. *Nature Reviews Molecular Cell Biology*, 8(12), pp.947–956.

Ghassemian, M. et al., 2000. Regulation of abscisic acid signaling by the ethylene response pathway in Arabidopsis. *The Plant cell*, 12(7), pp.1117–26.

Gill, G., 2004. SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes & development*, 18(17), pp.2046–59.

Guzzo, C.M. & Matunis, M.J., 2013. Expanding SUMO and ubiquitin-mediated signaling through hybrid SUMO-ubiquitin chains and their receptors. *Cell Cycle*, 12(7), pp.1015–1017.

Hammoudi, V. et al., 2018. The Arabidopsis SUMO E3 ligase SIZ1 mediates the temperature

dependent trade-off between plant immunity and growth M. Quint, ed. *PLOS Genetics*, 14(1), p.e1007157.

Hammoudi, V. et al., 2016. Whole-genome duplications followed by tandem duplications drive diversification of the protein modifier SUMO in Angiosperms. *New Phytologist*, 211(1), pp.172–185.

Han, Y.-F. et al., 2016. The SUMO E3 Ligase-Like Proteins PIAL1 and PIAL2 Interact with MOM1 and Form a Novel Complex Required for Transcriptional Silencing. *The Plant cell*, 28(5), pp.1215–29.

Hatfield, P.M. et al., 1997. The ubiquitin-activating enzyme (E1) gene family in Arabidopsis thaliana. *The Plant journal : for cell and molecular biology*, 11(2), pp.213–26.

Hay, R.T., 2007. SUMO-specific proteases: a twist in the tail. *Trends in Cell Biology*, 17(8), pp.370–376.

Hermkes, R. et al., 2011. Distinct roles for Arabidopsis SUMO protease ESD4 and its closest homolog ELS1. *Planta*, 233(1), pp.63–73.

Hickey, C.M., Wilson, N.R. & Hochstrasser, M., 2012. Function and regulation of SUMO proteases. *Nature Reviews Molecular Cell Biology*, 13(12), pp.755–766.

Hochstrasser, M., 2009. Origin and function of ubiquitin-like proteins. *Nature*, 458(7237), pp.422–429.

Hoen, D.R. et al., 2006. Transposon-mediated expansion and diversification of a family of ULP-like genes. *Molecular Biology and Evolution*, 23(6), pp.1254–1268.

Hruz, T. et al., 2008. Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. *Advances in Bioinformatics*, 2008, pp.1–5.

Hua, Z. & Vierstra, R.D., 2011. The Cullin-RING Ubiquitin-Protein Ligases. *Annual Review of Plant Biology*, 62(1), pp.299–334.

Huang, L. et al., 2009. The Arabidopsis SUMO E3 ligase AtMMS21, a homologue of NSE2/MMS21, regulates cell proliferation in the root. *Plant Journal*, 60(4), pp.666–678.

Hunter, T., 2007. The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond. *Molecular Cell*, 28(5), pp.730–738.

Ishida, T. et al., 2012. MMS21/HPY2 and SIZ1, Two Arabidopsis SUMO E3 Ligases, Have Distinct

Functions in Development S. Strack, ed. *PLoS ONE*, 7(10), p.e46897.

Ishida, T. et al., 2009. SUMO E3 Ligase HIGH PLOIDY2 Regulates Endocycle Onset and Meristem Maintenance in Arabidopsis. *THE PLANT CELL ONLINE*, 21(8), pp.2284–2297.

Isono, E. & Nagel, M.-K., 2014. Deubiquitylating enzymes and their emerging role in plant biology. *Frontiers in Plant Science*, 5.

Itoh, H. et al., 2005. Dissection of the phosphorylation of rice DELLA protein, SLENDER RICE1. *Plant and Cell Physiology*, 46(8), pp.1392–1399.

Jacobsen, J. V. & Beach, L.R., 1985. Control of transcription of α-amylase and rRNA genes in barley aleurone protoplasts by gibberellin and abscisic acid. *Nature*, 316(6025), pp.275–277.

Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science*, 293(5532), pp.1074–1080.

Jin, J.B. et al., 2008. The SUMO E3 ligase, AtSIZ1, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through affects on FLC chromatin structure. *Plant Journal*, 53(3), pp.530–540.

Johnson, E.S. et al., 1997. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *The EMBO journal*, 16(18), pp.5509–19.

Joosen, R.V.L. et al., 2012. Visualizing the genetic landscape of Arabidopsis seed performance. *Plant physiology*, 158(2), pp.570–89.

Karssen, C.M. et al., 1983. Induction of dormancy during seed development by endogenous abscisic acid: studies on abscisic acid deficient genotypes of Arabidopsis thaliana (L.) Heynh. *Planta*, 157(2), pp.158–165.

Kawahara, Y. et al., 2013. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice*.

Kim, J.Y., Jang, I.-C. & Seo, H.S., 2016. COP1 Controls Abiotic Stress Responses by Modulating AtSIZ1 Function through Its E3 Ubiquitin Ligase Activity. *Frontiers in plant science*, 7(August), p.1182.

Kim, J.Y., Song, J.T. & Seo, H.S., 2017. Post-translational modifications of Arabidopsis E3 SUMO ligase AtSIZ1 are controlled by environmental conditions. *FEBS Open Bio*, 7(10), pp.1622–1634.

Kim, S.-I. et al., 2015. E3 SUMO ligase AtSIZ1 positively regulates SLY1-mediated GA signalling and plant development. *The Biochemical journal*, 469(2), pp.299–314.

Kim, S. Il et al., 2016. The E3 SUMO ligase AtSIZ1 functions in seed germination in Arabidopsis. *Physiologia Plantarum*, 158(3), pp.256–271.

Kong, X. et al., 2017. Arabidopsis SUMO protease ASP1 positively regulates flowering time partially through regulating FLC stability. *Journal of Integrative Plant Biology*, 59(1), pp.15–29.

Kurepa, J. et al., 2003. The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of sumo1 and -2 conjugates is increased by stress. *Journal of Biological Chemistry*, 278(9), pp.6862–6872.

Kwak, J.S. et al., 2016. Arabidopsis HIGH PLOIDY2 Sumoylates and Stabilizes Flowering Locus C through Its E3 Ligase Activity. *Frontiers in Plant Science*, 7, p.530.

Lee, J. et al., 2006. Salicylic acid-mediated innate immunity in Arabidopsis is regulated by SIZ1 SUMO E3 ligase. *The Plant Journal*, 49(1), pp.79–90.

Li, S.-J. & Hochstrasser, M., 1999. A new protease required for cell-cycle progression in yeast. *Nature*, 398(6724), pp.246–251.

Li, S.-J. & Hochstrasser, M., 2003. The Ulp1 SUMO isopeptidase: distinct domains required for viability, nuclear envelope localization, and substrate specificity. *The Journal of cell biology*, 160(7), pp.1069–81.

Li, S.J. & Hochstrasser, M., 2000. The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Molecular and cellular biology*, 20(7), pp.2367–77.

Lima, C.D. & Reverter, D., 2008. Structure of the human SENP7 catalytic domain and poly-SUMO deconjugation activities for SENP6 and SENP7. *The Journal of biological chemistry*, 283(46), pp.32045–55.

Ling, Y. et al., 2012. Mutation in SUMO E3 ligase, SIZ1, Disrupts the Mature Female Gametophyte in Arabidopsis E. Newbigin, ed. *PLoS ONE*, 7(1), p.e29470.

Liu, L. et al., 2017. Two SUMO Proteases SUMO PROTEASE RELATED TO FERTILITY1 and 2 Are Required for Fertility in Arabidopsis. *Plant Physiology*, 175(4), pp.1703–1719.

Liu, M. et al., 2014. SUMO E3 ligase AtMMS21 is required for normal meiosis and gametophyte

development in Arabidopsis. *BMC Plant Biology*, 14(1), p.153.

Liu, Y. et al., 2016. The Arabidopsis SUMO E3 Ligase AtMMS21 Dissociates the E2Fa/DPa Complex in Cell Cycle Regulation. *The Plant Cell*, 28(9), pp.2225–2237.

Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2^-DDCT Method. *Methods*, 25, pp.402–408.

Locascio, A. et al., 2014. Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. *Frontiers in Plant Science*, 5, p.412.

Lois, L.M., 2010. Diversity of the SUMOylation machinery in plants. *Biochemical Society Transactions*, 38(1), pp.60–64.

Lois, L.M. & Lima, C.D., 2005. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO Journal*, 24(3), pp.439–451.

Lois, L.M., Lima, C.D. & Chua, N.-H., 2003. Small ubiquitin-like modifier modulates abscisic acid signaling in Arabidopsis. *The Plant cell*, 15(6), pp.1347–59.

Mazur, M.J. et al., 2017. Arabidopsis TCP Transcription Factors Interact with the SUMO Conjugating Machinery in Nuclear Foci. *Frontiers in Plant Science*, 8, p.2043.

Melchior, F., Schergaut, M. & Pichler, A., 2003. SUMO: ligases, isopeptidases and nuclear pores. *TRENDS in Biochemical Sciences*, 28(11).

Merrill, J.C. et al., 2010. A Role for Non-Covalent SUMO Interaction Motifs in Pc2/CBX4 E3 Activity S. Agarwal, ed. *PLoS ONE*, 5(1), p.e8794.

Miller, M.J. et al., 2010. Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in Arabidopsis. *Proceedings of the National Academy of Sciences*, 107(38), pp.16512–16517.

Miura, K. et al., 2011. Increased tolerance to salt stress in the phosphate-accumulating Arabidopsis mutants siz1 and pho2. *Planta*, 234(6), pp.1191–1199.

Miura, K. et al., 2007. SIZ1-Mediated Sumoylation of ICE1 Controls CBF3/DREB1A Expression and Freezing Tolerance in Arabidopsis. *THE PLANT CELL ONLINE*, 19(4), pp.1403–1414.

Miura, K. et al., 2013. *SIZ1* deficiency causes reduced stomatal aperture and enhanced drought tolerance via controlling salicylic acid-induced accumulation of reactive oxygen species in Arabidopsis. *The Plant Journal*, 73(1), pp.91–104.

Miura, K. et al., 2009. Sumoylation of ABI5 by the Arabidopsis SUMO E3 ligase SIZ1 negatively regulates abscisic acid signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13), pp.5418–5423.

Miura, K. et al., 2005. The Arabidopsis SUMO E3 ligase SIZ1 controls phosphate deficiency responses. *Proceedings of the National Academy of Sciences*, 102(21), pp.7760–7765.

Miura, K. & Hasegawa, P.M., 2010. Sumoylation and other ubiquitin-like post-translational modifications in plants. *Trends in Cell Biology*, 20(4), pp.223–232.

Miura, K., Jin, J.B. & Hasegawa, P.M., 2007. Sumoylation, a post-translational regulatory process in plants. *Current Opinion in Plant Biology*, 10(5), pp.495–502.

Miura, K. & Nozawa, R., 2014. Overexpression of SIZ1 enhances tolerance to cold and salt stresses and attenuates response to abscisic acid in Arabidopsis thaliana. *Plant Biotechnology*, 31, pp.167–172.

Mossessova, E. et al., 2000. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Molecular cell*, 5(5), pp.865–76.

Mukhopadhyay, D. & Dasso, M., 2007. Modification in reverse: the SUMO proteases. *Trends in Biochemical Sciences*, 32(6), pp.286–295.

Murtas, G. et al., 2003. A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. *The Plant cell*, 15(10), pp.2308–19.

Nambara, E. et al., 2010. Abscisic acid and the control of seed dormancy and germination. *Seed Science Research*, 20(2), pp.55–67.

Novatchkova, M. et al., 2004. SUMO conjugation in plants. *Planta*, 220(1), pp.1–8.

Nukarinen, E. et al., 2017. Protein sumoylation and phosphorylation intersect in Arabidopsis signaling. *Plant Journal*, 91(3), pp.505–517.

Olsen, O.-A., 2004. Nuclear endosperm development in cereals and Arabidopsis thaliana. *The Plant cell*, 16 Suppl(suppl 1), pp.S214–S227.

Owerbach, D. et al., 2005. A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochemical and Biophysical Research Communications*, 337(2), pp.517–520.

Panse, V.G. et al., 2002. Unconventional tethering of Ulp1 to the transport channel of the nuclear pore complex by karyopherins. *Nature Cell Biology*, 5(1), pp.21–27.

Park, B.S., Song, J.T. & Seo, H.S., 2011. Arabidopsis nitrate reductase activity is stimulated by the E3 SUMO ligase AtSIZ1. *Nature Communications*, 2, p.400.

Peng, J.R. et al., 1999. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, 400(6741), pp.256–261.

Perry, J.J.P., Tainer, J.A. & Boddy, M.N., 2008. A SIM-ultaneous role for SUMO and ubiquitin. *Trends in Biochemical Sciences*, 33(5), pp.201–208.

Pieterse, C.M.J., Pierik, R. & Van Wees, S.C.M., 2014. Different shades of JAZ during plant growth and defense. *The New phytologist*, 204(2), pp.261–4.

Raghavendra, A.S. et al., 2010. ABA perception and signalling. *Trends in Plant Science*, 15(7), pp.395–401.

Reeves, P.H., 2002. early in short days 4, a mutation in Arabidopsis that causes early flowering and reduces the mRNA abundance of the floral repressor FLC. *Development*, 129(23), pp.5349–5361.

Rodriguez, M.S., Dargemont, C. & Hay, R.T., 2001. SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *The Journal of biological chemistry*, 276(16), pp.12654–9.

Saitoh, H., Robert T, P. & Dasso, M., 1997. SUMO-1: Wrestling with a new ubiquitin-related modifier. *Trends in Biochemical Sciences*, 22(10), pp.374–376.

Saleh, A. et al., 2015. Posttranslational modifications of the master transcriptional regulator NPR1 enable dynamic but tight control of plant immune responses. *Cell Host and Microbe*, 18(2), pp.169–182.

Saracco, S.A. et al., 2007. Genetic Analysis of SUMOylation in Arabidopsis: Conjugation of SUMO1 and SUMO2 to Nuclear Proteins Is Essential. *PLANT PHYSIOLOGY*, 145(1), pp.119–134.

Schulman, B.A., 2011. Twists and turns in ubiquitin-like protein conjugation cascades. *Protein Science*, 20(12), pp.1941–1954.

Schweighofer, A., Hirt, H. & Meskiene, I., 2004. Plant PP2C phosphatases: emerging functions in stress signaling. *Trends in Plant Science*, 9(5), pp.236–243.

Schwienhorst, I., Johnson, E.S. & Dohmen, R.J., 2000. SUMO conjugation and deconjugation. *Molecular and General Genetics MGG*, 263(5), pp.771–786.

Soon, F. et al., 2012. Molecular Mimicry Regulates ABA Signaling by SnRK2 Kinases. *Science (New York, N.Y.)*, 335(January), pp.85–89.

Spoel, S.H., 2018. Orchestrating the proteome with post-translational modifications. *Journal of Experimental Botany*, 69(19), pp.4499–4503.

Srilunchang, K., Krohn, N.G. & Dresselhaus, T., 2010. DiSUMO-like DSUL is required for nuclei positioning, cell specification and viability during female gametophyte maturation in maize. *Development*, 137(2), pp.333–345.

Srivastava, A.K. et al., 2017. Rice SUMO protease Overly Tolerant to Salt 1 targets the transcription factor, OsbZIP23 to promote drought tolerance in rice. *Plant Journal*, 1, pp.1031–1043.

Srivastava, A.K., Zhang, C., Yates, G., et al., 2016. SUMO Is a Critical Regulator of Salt Stress Responses in Rice. *Plant Physiology*, 170(4), pp.2378–2391.

Srivastava, A.K., Zhang, C. & Sadanandom, A., 2016. Rice OVERLY TOLERANT TO SALT 1 (OTS1) SUMO protease is a positive regulator of seed germination and root development. *Plant Signaling and Behavior*, 11(5), pp.1–3.

Sydorskyy, Y. et al., 2010. A novel mechanism for SUMO system control: regulated Ulp1 nucleolar sequestration. *Molecular and cellular biology*, 30(18), pp.4452–62.

Tomanov, K. et al., 2014. Arabidopsis PIAL1 and 2 Promote SUMO Chain Formation as E4-Type SUMO Ligases and Are Involved in Stress Responses and Sulfur Metabolism. *The Plant Cell*, 26(11), pp.4547–4560.

Tomanov, K., 2014. *Mutant Analysis and Biochemical Characterisation of the Arabidopsis SUMO Ligase PIAL2 "* Unversität Wien.

Tomanov, K. et al., 2013. Small Ubiquitin-Like Modifier Conjugating Enzyme with Active Site Mutation Acts as Dominant Negative Inhibitor of SUMO Conjugation in Arabidopsis. *Journal of Integrative Plant Biology*, 55(1), pp.75–82.

Tomanov, K. et al., 2018. SUMO chain formation relies on the amino-terminal region of SUMO-conjugating enzyme and has dedicated substrates in plants. *The Biochemical journal*, 475(1), pp.61–74.

Truong, K., Lee, T.D., Li, B., et al., 2012. Sumoylation of SAE2 C terminus regulates SAE nuclear localization. *The Journal of biological chemistry*, 287(51), pp.42611–9.

Truong, K., Lee, T.D. & Chen, Y., 2012. Small ubiquitin-like modifier (SUMO) modification of E1 Cys domain inhibits E1 Cys domain enzymatic activity. *The Journal of biological chemistry*, 287(19), pp.15154–63.

Ulrich, H.D., 2005. Mutual interactions between the SUMO and ubiquitin systems: A plea of no contest. *Trends in Cell Biology*, 15(10), pp.525–532.

Vierstra, R.D., 2003. The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends in Plant Science*, 8(3), pp.135–142.

Vierstra, R.D. & Callis, J., 1999. Polypeptide tags, ubiquitous modifiers for plant protein regulation. *Plant Molecular Biology*, 41(4), pp.435–442.

Villajuana-Bonequi, M. et al., 2014. Elevated salicylic acid levels conferred by increased expression of ISOCHORISMATE SYNTHASE 1 contribute to hyperaccumulation of SUMO1 conjugates in the Arabidopsis mutant *early in short days 4*. *The Plant Journal*, 79(2), pp.206–219.

Walsh, C.K. & Sadanandom, A., 2014. Ubiquitin chain topology in plant cell signaling: a new facet to an evergreen story. *Frontiers in plant science*, 5, p.122.

Wei, W. et al., 2008. A stress-dependent SUMO4 sumoylation of its substrate proteins. *Biochemical and Biophysical Research Communications*, 375(3), pp.454–459.

Weissman, A.M., 2001. Themes and variations on ubiquitylation. *Nature Reviews Molecular Cell Biology*, 2(3), pp.169–178.

Westerbeck, J.W. et al., 2014. A SUMO-targeted ubiquitin ligase is involved in the degradation of the nuclear pool of the SUMO E3 ligase Siz1. *Molecular biology of the cell*, 25(1), pp.1–16.

Winter, D. et al., 2007. An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets I. Baxter, ed. *PLoS ONE*, 2(8), p.e718.

Xie, Z. et al., 2006. Interactions of two abscisic-acid induced WRKY genes in repressing gibberellin signaling in aleurone cells. *The Plant Journal*, 46(2), pp.231–242.

Xiong, R. & Wang, A., 2013. SCE1, the SUMO-conjugating enzyme in plants that interacts with NIb, the RNA-dependent RNA polymerase of Turnip mosaic virus, is required for viral

infection. *Journal of virology*, 87(8), pp.4704–15.

Xu, P. et al., 2013. AtMMS21, an SMC5/6 complex subunit, is involved in stem cell niche maintenance and DNA damage responses in Arabidopsis roots. *Plant physiology*, 161(4), pp.1755–68.

Xu, X.M. et al., 2007. NUCLEAR PORE ANCHOR, the Arabidopsis Homolog of Tpr/Mlp1/Mlp2/Megator, Is Involved in mRNA Export and SUMO Homeostasis and Affects Diverse Aspects of Plant Development. *THE PLANT CELL ONLINE*, 19(5), pp.1537–1548.

Yates, G., Srivastava, A.K. & Sadanandom, A., 2016. SUMO proteases: Uncovering the roles of deSUMOylation in plants. *Journal of Experimental Botany*, 67(9), pp.2541–2548.

Yu, F., Wu, Y. & Xie, Q., 2016. Ubiquitin-Proteasome System in ABA Signaling: From Perception to Action. *Molecular plant*, 9(1), pp.21–33.

Yuan, D. et al., 2014. AtMMS21 regulates DNA damage response and homologous recombination repair in Arabidopsis. *DNA Repair*, 21, pp.140–147.

Zhang, J. et al., 2017. A SUMO Ligase AtMMS21 Regulates the Stability of the Chromatin Remodeler BRAHMA in Root Development. *Plant physiology*, 173(3), pp.1574–1582.

Zhang, S. et al., 2013. SUMO E3 Ligase AtMMS21 Regulates Drought Tolerance in Arabidopsis thaliana. *Journal of Integrative Plant Biology*, 55(1), pp.83–95.

Zhao, Q. et al., 2014. GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Research*, 42(W1), pp.W325–W330.

10 SUPPLEMENTARY DATA

Supplementary table 1: primer sequences (5' to 3')

Gene	Forward	Reverse
qPCR		
qESD4	GTACCCTAATGGGTGCCGTAG	GAGGAGCTTTAGCATCAGGGT
qELS1	TGCGGATACTCAATAGCGGTC	CCATAGCCACGTTTTCGACAC
qULP1b	CATCTCGGACAAACCAGTCTC	ATTGTGAACACCTTGCTGCTT
qOTS2	AGTGCAGGTTCCACAACAGAA	CCCTTGCGGAATAGATCAACG
qOTS1	TGGTTTAGACCCGATGAAGCC	AACCTATGGAGTCTGTAGCCA
qULP2a	TGGAGAATCTTCAGGGATTCA	CTTGGCCATCACATTTGAACT
qULP2b	ATGCAGAGCACGTTAATCAGG	GTGAGTTTAGCTGGCGTTGAC
qubc9	TGGATTGGTTTTCGATTGCAGA	TGGATGGAACACCTTCGTCC
qcbp20	TCG TAA GGT GGG CCA TGA AA	CTG ATA GCT TTG CTT GCT CCT TG
qSUM1	TCC AGG TTT AGG GCT TGT GTG	GGA TCC GAT ACC AAA CGA ACA
qSUM2	TGG CAG GCT TTG TTT TAC TTT GA	CGG CAA CTT GTG CGT TTA CTT
qSUM3	CTT CGT CCA GGT TTA GGG CTT	AGG ATC CGA TAC CAA ACG AAC A
qSUM4	GAG AGA GAC AGT TGA GCC ACC	ACA GAT CAT GAC CGA TCA CTC G
qSUM5	GAT CTC GCC GGA GTT TCT TCA	ACG CTT TCT CGA TAT TCC GGT
qSUM6	TCA AAA CCC AAA AGC AGC CAC	ATC TCA ACC GGT CGA TCA CTC
qSAE1a	CCC AGT GGG AGG ACT AGG TTT	TAA TCT GGC GGT CGT ACA AGG
qSAE1b	ACT TGG CTC CTT GTC TTT TGT	TGC CAC TCA AAG AGC CAG AAA
qSAE2	CTG AAT CAC GAA ATC AAG ATG GGT	GGA GAT GGT GTC CAT CCA GAG
qSCE1	ATC CGA GCA AAC CTC CGA AAT	CGG GAT TCG GTG TGT CAA GTA
qSIZ1	ATA GCG CCT CTG GGA ATC AT	GCC TTG TCT TGT CTA CTG TCA TTC ATA C
qHPY2	ACAGTTCCACGCTTGCTGATA	TGC TGT CGA ACG GTG ATT ACA
qPIAL1	CCGAACGCGTATCAAACTTCC	TTT CCG TCT CAA CCA TCC AGG

qPIAL2	ACCATACCGAATCCCTCTCCA	ACT GGC TCG TTA CTG GAA CTG
sequencing		
ulp1b	CAC CAT GCC TCT TGT ACC CAA GAT TC	GTC AGC TCG CAG TCT CAG TAT CT

Supplementary table 2: ULP SUMO proteases

Accession number	Systematic name	Mutant name
AT4G15880		ESD4
AT3G06910	ULP1a	ELS1
AT4G00690	ULP1b	
AT1G10570	ULP1c	OTS2
AT1G60220	ULP1d	OTS1
At4g33620	ULP2a	SPF2
AT1G09730	ULP2b	SPF1, ASP1

Supplementary table 3: Chapter 3 ULP accession codes

Label	Accession code	Label	Accession code	Label	Accession code
AtULP1D	OAP13197.1	OsULP1	LOC_Os01g25370.1	ZmULP1D	ONM38459.1
AtULP2B	NP_001184951.1	OsULP2	LOC_Os03g22400.1	ZmULP1C	XP_020408528.1
AtULP2A	OAO99507.1	OsULP3	LOC_Os04g30860.1	ZmULP3	XP_008644224.1
AtULP1A	NP_187347.2	OsULP4	LOC_Os04g54680.1	ZmULP4	XP_020397614.1
AtESD4	OAO98717.1	OsULP5	LOC_Os05g11770.1	ZmULP5	NP_001147104.1
AtLP1B	OAO98233.1	OsULP6	LOC_Os05g34520.1	ZmULP6	NP_001150238.1
AtULP3	NP_190417.4	OsULP7	LOC_Os05g40660.1	ZmULP7	NP_001241799.1
AtULP1C	NP_172527.2	OsULP8	LOC_Os06g29310.1	ZmULP8	XP_020406897.1
BrULP1C	XP_009148385.1	OsULP10	LOC_Os07g13010.1	ZmESD4	ONM26145.1
BrULP1D	XP_009113217.1	OsULP11	LOC_Os09g08450.1	ZmULP9	NP_001132719.1
BrULP3	XP_018511509.1	OsULP12	LOC_Os09g12480.1	ZmULP10	AQK61393.1
BrULP2A	XP_018511049.1	OsULP13	LOC_Os10g24954.1	ZmULP11	XP_020393394.1
BrULP2B	XP_009110806.1	OsULP14	LOC_Os11g01180.1	ZmULP1B	XP_020393398.1
BrULP1A	XP_009124609.1	OsULP15	LOC_Os12g01290.1	ZmULP2B	AQK99349.1
BrESD4	XP_009136721.1	OsULP16	LOC_Os12g41380.1	ZmULP12	XP_020399622.1
BrULP1B	XP_009124236.1	OsULP17	LOC_Os03g29630.1	ZmULP13	XP_020402837.1
BoULP1C	XP_013586313.1	OsULP18	LOC_Os11g10780.1	ZMULP14	ONM06236.1
BoULP1D	XP_013612591.1	OsULP19	LOC_Os01g63040.1	ZmULP15	ONL99895.1
BoULP2A	XP_013606968.1	OsULP20	ABA97984.2	ZmULP16	ONM06627.1
BoULP2B	XP_013605145.1	OsULP21	BAS95897.1	ZmULP17	ONL97048.1
BoULP1A	XP_013619667.1	OsULP22	BAD81124.1	ZmOTS1b	ANH10625.1
BoESD4	XP_013593423.1	OsULP23	ABF95757.1	ZmOTS1d	NP_001150833.1
BoULP1B	XP_013632030.1	OsULP1D	XP_015635087.1	SbULP1	KXG19998.1
BnULP1C	XP_013586313.1	BdULP1D	XP_010227697.1	SbUPL2	EES01448.2
BnULP1D	XP_013725317.1	BdULP1C	XP_010233511.2	SbULP3	EES06202.2
BnULP3	CDY28152.1	BdULP2A	XP_010236177.1	SbULP4	OQU79548.1
BnULP2A	XP_013735196.1	BdESD41	XP_003567671.1	SbULP5	OQU79546.1
BnULP2B	XP_013750844.1	BdESD42	XP_003559810.1	SbULP6	XP_002442491.1
BnULP4	CDY65540.1	BdULP1B	XP_003557935.1	SbULP7	CAZ96035.1
BnULP1A	XP_013640810.1	BdULP2B	XP_010239228.1	SbULP2B	EES03901.2

BnESD4	XP_013742926.1
BnULP5	CDY49966.1
BnULP1B	XP_013632030.1

BdUlp3	KQJ88966.1
BdULP4	XP_010237055.1
BdULP5	XP_014753317.1
TaULP1	CDM84034.1
TaULP2	CDM82971.1
TaULP3	CDM84034.1
TaULP4	CDM83319.1
TaULP5	CDM86144.1
TaULP6	CDM86296.1
TaULP7	CDM86575.1

SbULP1A	KXG36922.1
SbULP8	KXG28247.1
SbULP9	XP_002449387.1
SbESD4	XP_002457813.1
SbULP10	OQU79584.1
HVULP1	MLOC_18163.4
HVULP2	BAK05429.1
HVULP3	BAK00185.1

Supplementary table 4: SUMO paralog GPS-SUMO values

paralog	K number	p-value	paralog	K number	p-value
SUMO1	9	0.805	SUMO3	23	0.978
SUMO1	10	0.313	SUMO3	33	0.339
SUMO1	21	0.159	SUMO3	35	0.62
SUMO1	23	0.908	SUMO3	36	0.364
SUMO1	35	0.332	SUMO3	41	0.639
SUMO1	41	0.354	SUMO3	42	0.266
SUMO1	42	0.206	SUMO3	54	0.044
SUMO2	9	0.811	SUMO5	15	0.755
SUMO2	10	0.09	SUMO5	16	0.833
SUMO2	20	0.616	SUMO5	27	0.8
SUMO2	22	0.751	SUMO5	28 - 32	0.957
SUMO2	47	0.342	SUMO5	31	0.538
SUMO2	53	0.343	SUMO5	33	0.952
SUMO2	54	0.245	SUMO5	43	0.088
SUMO2	108	0.047	SUMO5	51	0.766
SUMO2	112	0.051	SUMO5	52	0.348
SUMO3	8	0.329	SUMO5	60	0.692