# Investigations of *Arabidopsis thaliana* WHIRLY1 function in early seedling development

Dissertation

To obtain the academic degree
doctor rerum naturalium (Dr. rer. nat.)

submitted to
Faculty of natural sciences I - Biosciences -
Martin-Luther-University Halle-Wittenberg

submitted by Thuy Linh Nguyen

Reviewer 1: Prof. Dr. Klaus Humbeck
Reviewer 2: Prof. Dr. Sascha Laubinger
Reviewer 3. Prof. Dr. Karin Krupinska

Date of public defense: 12/12/2023

# TABLE OF CONTENT

# ABBREVIATIONS

| | |
|---|---|
| $^{1}O_2$ | Singlet oxygen |
| A3SS | Alternative 3' splicing site |
| A5SS | Alternative 5' splicing site |
| aa | Amino acid |
| ABA | Abscisic acid |
| aGSL | Aliphatic glucosinolate |
| AS | alternative splicing |
| cDNA | complementary DNA |
| DAG | Days after germination |
| DASE | Differentially alternative splicing event |
| DASG | Differentially alternative spliced gene |
| DEG | Differentially expressed gene |
| dsDNA | Double-stranded DNA |
| fpkm | fragments per kilobase of transcript per million mapped reads |
| GA | gibberellins |
| GO | Gene Ontology |
| GSL | Glucosinolate |
| $H_2O_2$ | Hydrogen peroxide |
| iGSL | Indole glucosinolate |
| OE | overexpressing |
| PCR | Polymerase chain reaction |
| PSI | Percent spliced in |
| PSII | Photosystem II |
| qRT-PCR | quantitative real-time PCR |
| RE | Restriction enzyme |
| RI | Retained Intron |
| RNA-seq | RNA-sequencing |
| ROS | Reactive oxygen species |
| RT-PCR | Reverse-transcriptase PCR |
| SE | Skipped exon |
| ssDNA | Single-stranded DNA |
| TF | transcription factor |
| TP | transit peptide |

# SUMMARY

WHIRLY1 (WHY1) belongs to a plant-specific transcription factor family, which is characterized by a conserved single-stranded-DNA(ssDNA)-binding domain that supports the binding of DNA or RNA. Besides, many WHIRLIES are located in not only the nucleus but also other nucleic acid-containing organelles, i.e., chloroplast and mitochondrion. In *Arabidopsis thaliana*, AtWHIRLY1 was reported to play significant roles in later stages of development, including flowering and leaf senescence, as well as biotic and abiotic stress responses, as results from studying a T-DNA insertion line.

In this work, the involvement of AtWHIRLY1 in the early developmental stage, i.e., five days after germination, was investigated using two knock-out mutants of *AtWHIRLY1* created by CRISPR/Cas9 technique, *why1-4* and *why1-5.* The loss-of-function of *WHIRLY1* at this stage did not cause macrophenotypic differences compared to the wildtype. However, knock-out of *WHIRLY1* affected expression of a small but specific set of genes during this critical phase, in which many of these genes have functions in biotic and abiotic stress responses, as well in processes involved in early post-germination plant development. Noticeably, most of genes involved in first steps of aliphatic glucosinolate biosynthesis were suppressed by lacking of WHIRLY1, leading to a reduction in aliphatic glucosinolate content in the knock-out line *why1-5*. Another striking result was that knock-out of *AtWHIRLY1* affected alternative splicing. Due to loss-of-function of *WHIRLY1*, 1367 differential alternative splicing events in 1043 genes were recorded, most of them were retained introns. Interestingly, a large number of alternatively spliced genes in the *why1-5* line are related to DNA and RNA-related functions, especially RNA splicing/processing and DNA-repair.

Plants during early development have to assert themselves in a constantly changing environment. A common consequence of such menacing changes in the abiotic and biotic surrounding is oxidative stress. To investigate whether AtWHIRLY1 is involved in protective measures against this kind of stress, an established experimental set-up to create oxidative stress by using elevated light intensity under low temperature was used. By analyzing the stress-associated transcriptome of the WT seedling, two co-occurant independent layers of nuclear gene regulation were revealed, the well-known transcriptional regulatory network and in addition, a post-transcriptional modulation via RNA splicing. Oxidative stress caused differential regulation of thousands of genes in both the knock-out line *why1-5* and the WT seedlings. These two genetic backgrounds showed a similar stress-induced transcriptome, though the missing of a functional WHIRLY1 appeared to affect the time-scale and the magnitude of differential expression in response

to oxidative stress. On the other hand, the post-transcriptional regulation acted on a different set of target genes. The changes in alternative splicing patterns of these genes provide an additional regulatory level protecting the seedlings against this oxidative stress. Many differentially spliced genes function themselves in RNA splicing and also in DNA repair, in gene expression regulation including transcription factors and chromosome remodellers, and also encode enzymatic proteins. Interestingly, the loss of *WHIRLY1* function significantly affectted the alternative splicing pattern upon stress, as different variants and genes were deregulated compared to the WT.

Taken together, AtWHIRLY1 functions in the post-germination early development, balancing growth and defense processes at different regulatory levels. Together with previous knowledge, it is proposed that AtWHIRLY1 acts in a central regulatory hub and may serve as a scaffold, where multiple regulatory proteins can bind to and thereby modulate gene expression. These proteins are components of chromatin remodelling (affecting epigenetic markers), of transcriptional complex (affecting gene expression level), and alternative splicing complex (affecting transcript variants' ratio).

# ZUSAMMENFASSUNG

WHIRLY1 (WHY1) gehört zu einer pflanzenspezifischen Transkriptionsfaktor-Familie, die durch eine konservierte ssDNA-Bindungsdomäne gekennzeichnet ist, die die Bindung von DNA oder RNA unterstützt. Außerdem sind WHIRLIES nicht nur im Zellkern, sondern auch in anderen nukleinsäurehaltigen Organellen (Chloroplasten und Mitochondrien) lokalisiert. In *Arabidopsis thaliana* spielt AtWHIRLY1 eine wichtige Rolle in späteren Entwicklungsstadien, einschließlich Blüte und Blattseneszenz, sowie bei biotischen und abiotischen Stressreaktionen. Dies konnte durch frühere Untersuchung einer T-DNA-Insertionslinie belegt werden.

In dieser Arbeit wurde die Funktion von AtWHIRLY1 im frühen Entwicklungsstadium, d. h. fünf Tage nach der Keimung, anhand von zwei Knock-out-Mutanten von *AtWHIRLY1* (*why1-4* und *why1-5*) untersucht, die mit der CRISPR/Cas9-Technik erzeugt wurden. Der Funktionsverlust von *WHIRLY1* bewirkte keine phänotypischen Unterschiede, beeinflusste jedoch die Expression einer kleinen, aber spezifischen Gruppe von Genen in dieser kritischen Phase, in der viele Stress- und Entwicklungsgene exprimiert werden. Auffallend ist, dass die Expression der meisten Gene, die an frühen Schritten der Biosynthese aliphatischer Glucosinolate beteiligt sind, durch das Fehlen von WHIRLY1 unterdrückt wurde, was zu einer Verringerung des Gehalts an aliphatischen Glucosinolaten in der Knock-out-Linie *why1-5* führte. Knockout von *AtWHIRLY1* beeinflusste zusätzlich das alternative Spleißen. Es wurden 1367 unterschiedliche alternative Spleißereignisse in 1043 Genen detektiert, wobei in den meisten Fällen Introns behalten wurden. In der Knock-out-Mutante wurden im Vergleich zum Wildtyp ähnlich viele erhöhte oder verringerte Einschlusssequenzen gefunden. Interessanterweise steht eine große Anzahl von Genen, die in der *why1*-Linie alternativ gespleißt werden, mit DNA und RNA-bezogenen Funktionen in Zusammenhang, insbesondere mit RNA-Spleißen/RNA-Verarbeitung und DNA-Reparatur.

Pflanzen müssen sich in der frühen Entwicklung in einer sich ständig verändernden Umgebung behaupten. Eine häufige Folge solch bedrohlicher Veränderungen in der abiotischen und biotischen Umgebung ist oxidativer Stress. Um zu untersuchen, ob AtWHIRLY1 an Schutzmaßnahmen gegen diese Art von Stress beteiligt ist, wurde ein etablierter Versuchsaufbau zur Erzeugung von oxidativem Stress durch eine erhöhte Lichtintensität bei niedriger Temperatur verwendet. Durch die Analyse der stressassoziierten Transkriptome der WT-Keimlinge wurden zwei nebeneinander bestehende unabhängige Ebenen der nukleären stressabhängigen Genregulation

aufgedeckt: das bekannte transkriptionelle Regulationsnetzwerk und das Alternative Spleißen als eine andere posttranskriptionelle Modulation. Der oxidative Stress bewirkte in beiden Genotypen eine ähnliche differentielle Regulation tausender Stress-abhängiger Gene. Das Fehlen von funktionellem WHIRLY1 beeinflusste allerdings im Vergleich zum Wildtyp den Beginn und die Intensität der Expressionsveränderungen als Reaktion auf oxidativen Stress. Zusätzlich zeigten sich deutliche Unterschiede zwischen Knock-out-Linien and dem Wildtyp im alternativen Spleißmuster, wobei viele dieser alternativ gespleißten Gene Schutzproteine gegen vor abiotischen Stress kodieren. Außerdem sind viele der alternativ gespleißten Gene ihrerseits an der DNA-Reparatur und dem RNA-Spleißen, an der Genregulation als Transkriptionsfaktoren und an epigenetischen Prozessen, sowie an enzymatischen Reaktionen beteiligt. Der Verlust der WHIRLY1-Funktion wirkt sich signifikant auf das alternative Spleißmuster bei Stress aus, da verschiedene Varianten und Gene im Vergleich zum WT dereguliert wurden.

Unter Einbeziehung früherer Erkenntnisse wird vorgeschlagen, dass AtWHIRLY1 in der frühen Entwicklungsphase eine regulatorische Funktion zwischen Stressantwort und Wachstum ausübt. Es wird vorgeschlagen, dass WHIRLY1 als Gerüstprotein dienen kann, an das mehrere regulatorische Proteine binden und dadurch die Genexpression modulieren können. Diese Proteine sind Bestandteile von Komplexen epigenetischer Kontrolle (beeinflussen die Chromatinstruktur), des Transkriptionskomplexes (beeinflussen das Niveau der Genexpression) und des alternativen Spleißkomplexes (beeinflussen das Verhältnis der Varianten).

# 1. INTRODUCTION

## 1.1. The plant-specific ssDNA-binding protein family WHIRLY

### 1.1.1. Discovery of WHIRLY family and number of WHIRLIES in higher plant

WHIRLY (WHY) is a small and plant-specific single-stranded DNA-binding protein family (Desveaux et al., 2000, 2002) with a general structure as shown in Figure 1A. The name of the family derives from the appearance of the tetramere structure resembling a whirligig (Desveaux et al., 2002). The innitial discovery of WHIRLIES occurred in potato, where it was named p24, due to its molecular weight of 24 kDa (Desveaux et al., 2000). The protein p24 is a component of the nuclear factor PBF-2 (PR-10a binding factor 2) that binds to 30-bp elicitor reponse element (ERE) sequence in the promoter of *PR-10a*. Upon *Phytophthora infestans* infection, PBF-2 induces PR-10a gene expression (Després et al., 1995).

The number of genes encoding WHIRLIES varies, but at least two *WHIRLY* genes are present in higher plants. In most of monocots, two genes were identified, so called *WHIRLY1* and *WHIRLY2* (Oetke et al., 2022; Qiu et al., 2022). Phylogenetic analysis based on the Whirly domain sequence reveals that WHIRLIES can be classified into two distinct groups, which assemble to either monocot WHIRLY1 or monocot WHIRLY2 (Figure 1B). In some dicots, additional WHIRLY protein(s) encoded by new gene(s) show(s) similarity to either monocot WHIRLY1 (as in the case of *A. thaliana*, (Desveaux et al., 2005)) or monocot WHIRLY2 (as in the case of *M. cultiva*, (Ruan et al., 2022)). The expanding of WHIRLY family in specific plant lineages is likely a result of independent gene duplication events. In *A. thaliana*, for instance, AtWHIRLY1 and AtWHIRLY3 protein share around 67% similarity as they are originated from an ancestral gene, whereas AtWHIRLY2 shows less homology with AtWHIRLY1 (35%) and AtWHIRLY3 (33%) (Appendix Figure S1).
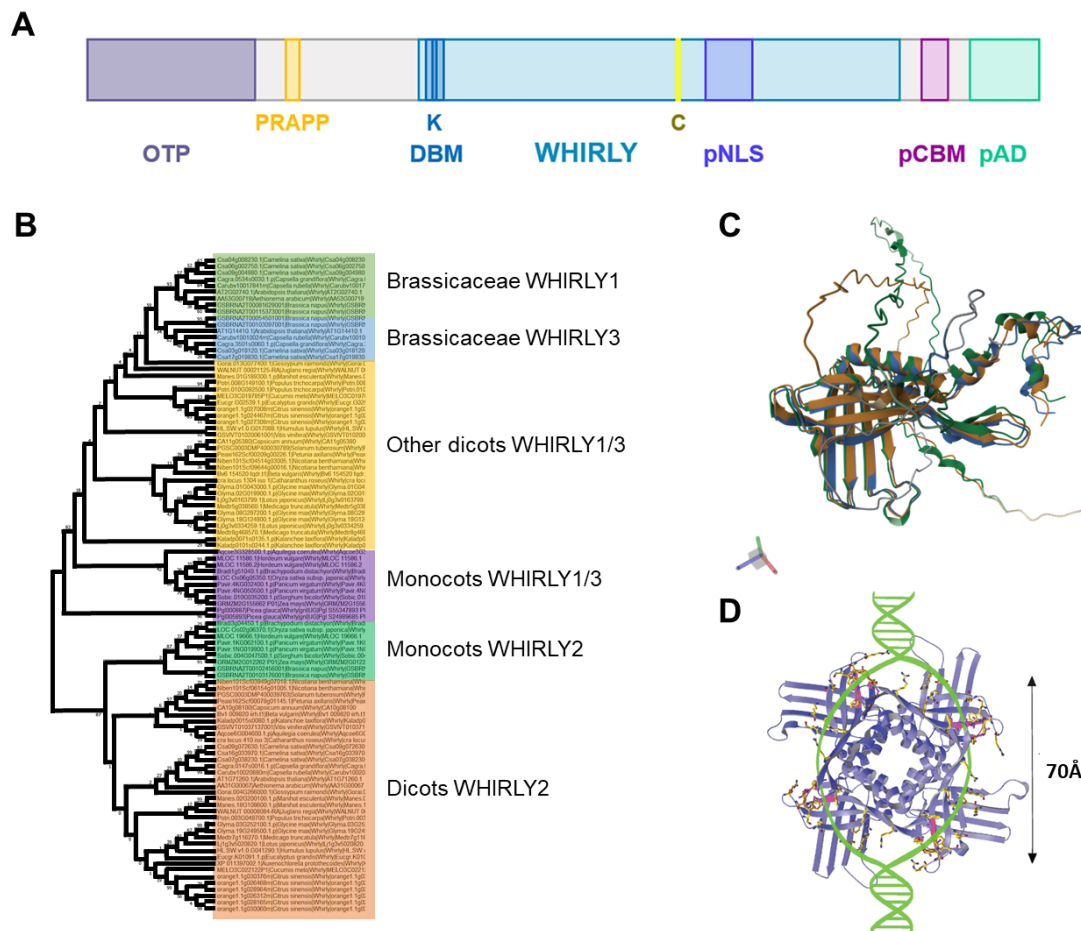
**Figure 1. The WHIRLY family in plants.** (**A**) Scheme of WHIRLIES with different motifs. An organelle targeting peptide (OTP) is located at N-terminal of all WHIRLIES, A PRAPP motif is identified in only monocot WHIRLY1. The conserved Whirly domain (WHIRLY) with around 180 aa contains the conserved DNA-binding motif (DBM) with a highly conserved Lysine (K) residue, a putative nuclear localization signal (pNLS), and a conserved Cystein (C) residue that is important for tetrameric assembly. A putative copper binding motif is predicted in all WHIRLIES except dicot WHIRLY1. A putative activation domain (pAD) is defined at C-terminal. Adapted from (Krupinska et al., 2022). (**B**) Phylogenetic tree created using maximum likelihood method with 114 WHIRLY sequences from 35 species in MEGA-X. (**C**) Alignment of three AtWHIRLY sequences using models of AlphaFold prediction (Jumper et al., 2021; Varadi et al., 2022) and Pairwise Structure Alignment web tool (Li et al., 2020) on RCSB PBD platform (Berman, 2000). AtWHIRLY1, AF-Q9M9S3-F1, is shown in yellow; AtWHIRLY2, AF-Q8VYF7-F1, is presented in blue; AtWHIRLY3, AF-Q66GR6-F1, is depicted in green. Three AtWHIRLIES share high homology in conformations. (**D**) A model of tetrameric form of WHIRLY1 interacting to melted ssDNA according to (Desveaux et al., 2002).

### 1.1.2. Structure of WHIRLIES determines their function

Every WHIRLY protein contains a single Whirly domain, which consists of approximately 180 amino acids. The Whirly domain is composed of two layers of four-β-sheets and two α-helices that are oriented perpendicular to each other (Figure 1C) (Desveaux et al., 2002). It is worth noting that the Whirly domain is one of four structural topologies that are characterized with ssDNA-binding capacity. The other three are oligonucleotide/ oligosaccharide/ oligopeptidebinding (OB) folds, K homology (KH)

domains, and RNA recognition motifs (RRMs) (Dickey et al., 2013). All these structural domains are comprised of β-sheets and α-helices coupled with loops, arranging in a manner that β-sheets form a ssDNA-binding pocket. Toward these pockets, the bases of the ssDNA face the protein surface and the backbone is exposed to the solvent. Unlike other three domains, the Whirly domain forms relatively few hydrogen bonds with bases and electrostatic interactions with the phosphate backbone of the ssDNA (Dickey et al., 2013 and references within). Interestingly, the interaction between the Whirly domain and ssDNA only occurs on the edge and in between β-sheets of adjacent monomers, suggesting the need for proper positioning between the ssDNA residues and the Whirly domains (Figure 1D). Perhaps, this arrangement allows WHIRLIES to bind ssDNA with high affinity but with little sequence specificity, as the bases are not in close proximity with the protein surface (Cappadocia et al., 2010, 2012). An early study demonstrated that StWHIRLY1 binds to ssDNA with very high affinity, though it can also weakly bind to double-stranded DNA (dsDNA) (Desveaux et al., 2000). Furthermore, WHIRLIES tend to preferentially bind the noncoding strand (NCS) rather than the coding strand (CS) of the ERE. Within the Whirly domain, a short sequence called DNA-binding motif (DBM), with the most conserved sequence KGKAAL, is crucial for ssDNA-binding activity (Figure 1A). This region forms a loop between two antiparalle β-sheets, and mutation in this sequence abolished the DNA-binding activity (Desveaux et al., 2002). Still, it cannot be ruled out that the mutated version was improperly folded. Nevertheless, it appears that mutations in the KGKAAL motif do not affect the polymerization of StWHIRLY2 *in vitro* (Desveaux et al., 2002).

WHIRLIES exhibit high affinity for binding to various ssDNA sequences, though the specificity seems to be low. Indeed, the reported binding sites of WHY proteins show very litte similarity. For example, *A. thaliana* WHIRLY1 binds to the PB element (GTCAAAAA/T), which overlaps with the opposite strand sequence of *cis*-elements such as the W-box [T/G]TGAC[C/T] and the TGACG (Desveaux et al., 2004). AtWHIRLY1 can also bind to an AT-rich region of the KP1 gene promoter (Xiong et al., 2009). In barley, HvWHIRLY1 binds to the sequence TGTCAn$_6$GGTCAA in the promoter region of *HvS40* gene, acting as a negative regulator of leaf senescence (Krupinska et al., 2014a). Additionally, AtWHIRLY1 directly regulates *AtWRKY53*, a senescence marker gene, by binding to the ERE-like motif (GTNNNAAT) and an AAAT rich region in a developmental-stage dependent manner (Miao et al., 2013). AtWHIRLY1 can also bind to the specific plant telomere sequence (TTTAGGG) (Yoo et al., 2007). Recently, tomato SlWHIRLY1 was reported to bind to the motif GTTACCCT in the promoter region of the *psbA* gene (Zhuang et al., 2019). Additionally, the binding of WHIRLIES to organelle DNA was

reported to be nonspecific (Maréchal et al., 2008; Prikryl et al., 2008). Interestingly, all the mentioned binding sites have been verified through experimental approaches, such as electrophoretic mobility shift assays (EMSA) or chromatin immunoprecipitation (ChIP) assays, to investigate the interaction between WHIRLIES and the target sequences. However, two genome-wide studies analyzing the binding sites of plant transcription factors (TF) failed to identify specific binding sequence for WHIRLIES (Franco-Zorrilla et al., 2014; O'Malley et al., 2016). Nevertheless, it was reported that WHIRLIES share binding sites with other TFs, indicating their interaction with other TFs to regulate gene function. For instance, the PBF-2 element is also used by other defense-related TFs (Desveaux et al., 2005). Moreover, W-boxes serve as binding sites for various WRKY TFs, playing significant roles in senescence as well as in abiotic and biotic responses (Jiang et al., 2017b).

It is common that ssDNA binding proteins function in oligomeric forms (dimers to tetramers), and often these multiple domains can be found within the same polypeptide with linkers between them or as multiple single subunits (Dickey et al., 2013 and references within). Similarly, WHIRLIES usually form whirligig-like homo-tetramers by using a conserved lysine residue (Figure 1D). Six tetrameres can further assemble to a higher-order structure called 24-mers, where the second lysine residue of the KGKAAL motif (K67) stabilizes the interactions between tetrameres (Cappadocia et al., 2012). *In vitro*, WHIRLIES can form 24-mers at very high local concentrations or upon binding long ssDNA molecules (Cappadocia et al., 2012). Moreover, WHIRLIES also were found *in vivo* in large complexes with DNA as the result of size exclusion chromatography studies in maize (Prikryl et al., 2008) and also in *A. thaliana* (Cappadocia et al., 2012). Interestingly, there have been no reports thus far regarding hetero-tetramers formed between different WHIRLIES within cell. It is also unclear whether WHIRLIES can function in monomeric or dimeric forms, as most of the existing research has focused on their tetrameric strutures.

All up-to-date known WHIRLIES possess an organelle transit peptide (OTP) (Figure 1A), which allows them to target into either plastids or mitochondria, or both (Huang et al., 2020; Isemer et al., 2012b; Krause et al., 2005). Importing WHIRLIES into organelles is crucial for their subsequent functions, as indicated by the detection of only one isoform of *AtWHIRLY1* transcript, and the similar size of WHIRLY1 in the nucleus and in plastid, both without the transit peptide (Isemer et al., 2012b). The subcellular localization of WHIRLIES in various species has been studied using different methods, including *in vitro* import assays and fluorescence protein imaging. Many studies showed that in all higher plants, WHIRLY1 proteins target to the chloroplasts and WHIRLY2 proteins are sorted to the mitochondria (Krupinska et al., 2022 and references within). However, the localization of WHIRLY2 and WHIRLY3 appears to be more complicated. In barley, a *WHIRLY1*

knock-out line exhibited an albino phenotype in young leaves, which turned green in older leaves, suggesting compensation by WHIRLY2 in plastid (personal communication). *AtWHIRLY3* was also upregulated in the root of the *AtWHIRLY2* knock-out mutant, indicating its potential role in replacing WHIRLY2 function in mitochondria (Golin et al., 2020). Recently, it was demonstrated that WHIRLY2 in *A. thaliana* can localize to all three genome-containing compartments (Huang et al., 2020). Nevertheless, the WHY2:GFP construct was observed only in the mitochondria but not in the other compartments.

Some WHIRLIES contain putative nuclear localization signals (NLS) that may aid their movement into the nucleus (Figure 1A). However, there is currently no experimental evidence confirming functionality of the NLS or explaining how WHIRLIES can commute between organelles or move from organelles to the nucleus (Krupinska et al., 2020). In fact, AtWHY1:GFP and HvWHY1:GFP fusion proteins were unable to be imported into nucleus (Melonek et al., 2010; Krause et al., 2005). Nonetheless, the presence of WHIRLY1 in the nucleus has been shown using immunoblot and immunohistochemical approaches (Ren et al., 2017; Melonek et al., 2010; Krause et al., 2005). Furthermore, WHIRLY1 as part of structural components in the nucleus was reported, for instance, as part of PBF-2 and KPRE binding factor 1 binding to promoter of *StPR-10a* (Desveaux et al., 2000) and *AtKP1* (Xiong et al., 2009), respectively, as well as one of proteins associated to single stranded telomere sequence (Luo et al., 2020; Yoo et al., 2007).

The Whirly domain is flanked by additional highly-disorder motifs at both ends. These flanking sequences show considerable diversity between different WHIRLIES and between species, indicating that they may confer different functions. One example is the presence of a stretch of 11 glutamine residue with one proline in the N-terminal half of WHIRLY1, which can also be classified as a proline/glutamine class of transcripton activation domains (Desveaux et al., 2002). In monocots WHIRLY1, a proline-rich PRAPP motif was recently shown to function in the compaction of organelle nucleoids (Oetke et al., 2022) (Figure 1A). At the C-terminal, a putative activation domain was identified also with some conserved residues that are important for ssDNA-binding activity, such as Glu271 and Trp272 in StWHIRLY1 (Desveaux et al., 2004). Furthermore, a putative copper binding motif (pCBM) is shared by all WHIRLIES in monocots as well as by WHIRLY2 and WHIRLY3 in *A. thaliana*, though AtWHIRLY1 does not contain this motif (Figure 1A). This suggests that AtWHIRLY3, rather than AtWHIRLY1, is more closely related to monocot WHIRLY1 (Krupinska et al., 2022).

### 1.1.3. WHIRLIES' functions are related to their nucleic acid binding activity and subcellular localization

In plant cells, DNA mostly exists in double-stranded helix conformation to stablize and protect the genome. However, ssDNA is often formed as consequence of multiple cellular processes, e.g., replication or transcription, or due to DNA breakage, or at telomeres of chromosomes. The WHIRLY family with the conserved ssDNA binding domain has been reported to be involved in these processes by binding to different ssDNA sequences. Furthermore, WHIRLIES can bind to dsDNA but with a lower affinity (Desveaux et al., 2000, 2002). Besides, RNA and ssDNA are chemically similar and same in flexibility and conformation. Therefore, WHIRLIES, as ssDNA binding proteins, can bind to ssRNA (Prikryl et al., 2008), and numerous of WHIRLIES' function were reported relating to their capacity of binding to nucleic acid in different gene-containing cell compartments (Figure 2).
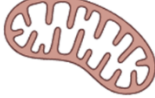
| | Nucleus | Plastid | Mitochondria |
|---|---|---|---|
| double helix dsDNA | | • Nucleoids compaction | • Nucleoids compaction |
| melted dsDNA | • Gene expression regulation<br>• Histone modification | | |
| ssDNA | • Telomere elongation | • Protect ssDNA overhang<br>• Promote precise repair | • Protect ssDNA overhang<br>• Promote precise repair<br>• mtDNA replication |
| RNA | | • RNA splicing | |

**Figure 2. Functions of WHIRLIES according to their localization and bound nucleic acids.** Details are discussed in Chapter 1.1.3. Created in BioRender.com

Chloroplasts and mitochondria are constantly producing various factors that can cause dsDNA breaks, such as reactive oxygen species (ROSs). Therefore, to maintain the integrity of organellar genomes, two repairing mechanisms can be implemented, including homologous recombination (HR) and nonhomologous DNA end joining (NHEJ), in which the later prones to cause mutations (Maréchal & Brisson, 2010). In *A. thaliana*, WHIRLY1 and WHIRLY3 have function in chloroplast genome maintaince as they accumulate to plastid unstable DNA regions to prevent error-prone NHEJ mechanism and to promote HR repair. Here, WHIRLY1 and WHIRLY3 work synergistically with other HR repair enzymes,

such as POL IB, RNase H1, and RecA1 (Wang et al., 2021; Zampini et al., 2015). Additionally, AtWHIRLY2 and AtODB1 are shown to be rapidly recruited to ssDNA regions in mitochondria to aid mtDNA replication as well as repair more efficiently (Janicka et al., 2012). Another type of ssDNA present in the cell is the telomere sequence. AtWHIRLY1 was found in telomeric binding protein complex and can interact with POT1a, component of the telomerase complex (Fulnečková et al., 2022; Yoo et al., 2007). There, AtWHIRLY1, by binding specifically to plant telomere sequence, modulates telomere length by negatively influencing telomerase activity (Yoo et al., 2007).

WHIRLIES also can be a part of organellar genomic structure. In *A. thaliana*, AtWHIRLY2 was identified in DNA/RNA-protein complexes in mitochondria, as it induces DNA nucleoid compaction and the knock-out mutant *why2-1* contained unpacked DNA in mitochondria (Golin et al., 2020; Janicka et al., 2012). Meanwhile, AtWHIRLY1 and AtWHIRLY3 were found in the transcriptionally active chromosome (TAC) in the chloroplast and therefore were also named pTAC1 and pTAC11, respectively (Pfalz et al., 2006). TAC is where various DNA-associated processes occur, including plastid DNA replication and transcription. However, these two AtWHIRLIES did not have impact on plastid nucleoid compaction, though AtWHIRLY2 showed this function in mitochondria (Krupinska et al., 2014b). In contrary, monocots WHIRLY1 clearly affect the nucleoid morphology as the knock-down RNAi barley mutant and knock-out maize line contain unpacked plastid DNA and very few visible nucleoids (Krupinska et al., 2014b; Prikryl et al., 2008). On the other hand, WHIRLY2 of monocot does not function in nucleoid compaction, at least in *E. coli* (Oetke et al., 2022). It was reported that the PRAPP motif, which has been only identified in monocot WHIRLY1, is responsible to this nucleoid related function (Oetke et al., 2022).

In nucleus, it was proposed that WHIRLIES can bind to melted promoter regions of target genes (Desveaux et al., 2002). WHIRLIES with a high affinity toward ssDNA, especially the non-coding strand, can stablize the promoter and allow other TFs interacting with them and regulate gene expresion in both a positive and a negative manner. A number of nuclear genes involved in different pathways was reported under the control of WHIRLY1, for instance pathogen-responsive genes, e.g., *StPR10a, AtPR1, AtKP1*; senescence-related genes, e.g., *AtWRKY53, HvS40*; development-related genes, e.g., *SlRbcS, AtICS1*; and hormone-associated genes, e.g., *HvNCED1* (Krupinska et al., 2022 and references within).

The binding of WHIRLY1 to the promoter region of target genes can additionally promote histone modifications. It was shown that AtWHIRLY1 repressed methylation and induced acetylation of histone 3 on promoter region of *AtWRKY53*, and consequently spatio-temporally affected the expression level of this gene during leaf sensecence (Huang

et al., 2018). Besides, in drought-induced senescence, HvWHIRLY1 affected histone markers on promoters of senescence-associated genes, including *HvS40* ( Manh et al., 2023; Janack et al., 2016) and *HvNCED1* (Manh et al., 2023). Interestingly, the overexpression and the knock-down mutants of *HvWHIRLY1* exerted a similar effect on chromatin markers, i.e., reduction of euchromatic markers upon drought-induced senescence, suggesting that an optimum level of WHIRLY1 is needed to modulate gene expression at epigenetic level. Perhaps, via impacting histone markers, WHIRLY1 affects the accessibility of TFs to promoters of target genes, thereby regulating gene expression.

The association between WHIRLY1 and RNA was reported in different species. Firstly, WHIRLY1 was found in the plastid transcriptionally active complex in *A. thaliana* and barley (Melonek et al., 2010; Prikryl et al., 2008; Pfalz et al., 2006), suggesting the involvement of WHIRLY1 in transcriptional and post-transcriptional regulations in the organelle. In monocots, WHIRLY1 was found in association with intron group II-containing RNAs (Qiu et al., 2022; Melonek et al., 2010; Prikryl et al., 2008). Barley WHIRLY1 knock-down mutant showed a decreased level of RNA splicing in chloroplasts (Melonek et al., 2010). Furthermore, ZmWHIRLY1 interacted with ZmCRS1, a plastidial group IIA intron splicing facilitator, and mediated *atpF* splicing process via modulating intron folding (Prikryl et al., 2008). Interestingly, the missing of *OsWHIRLY1* in rice did not cause abberant in splicing of *atpF* but of another chloroplast transcript, *ndhA* (Qiu et al., 2022). In the same study, it was shown for the first time that OsWHIRLY1 by interacting with OsTRXz manipulates the editing of two specific transcripts, *ndhA-C1070* and *rps14-C80*, but not of the rest of plastid transcripts.

### 1.1.4. WHIRLIES' functions in different plant developmental stages

*WHIRLIES* are expressed with high level in all stages of plant development, in accordance with numerous reported functions in every aspects of plant life, from embryo development to germination, and from flowering to senescence. WHIRLIES regulate plant development firstly by modulating hormone synthesis and signaling, e.g., of ABA and SA. In *A. thaliana*, the T-DNA insertion lines *why1-1* and *why1-2* exhibited reduced sensitivity to SA and ABA during germination (Isemer et al., 2012a). Overexpression of *AtWHIRLY1* with different localization of WHIRLY1 revealed that the sensitivity to these hormones is mediated via plastid-localized WHIRLY1 but not via the nuleus-located protein (Isemer et al., 2012a). Recently, Manh et al. (2023) showed the role of barley HvWHIRLY1 in ABA biosynthesis during drought-induced senescence by repressing *HvS40* and *HvNCED1* gene expresion level and consequently lowering ABA content during senescence. On the other hand, it was reported that cassava WHIRLIES bind to the promoter of *NCED1* to

induce its expression and to increase level of ABA and drought tolerance (Yan et al., 2020). Meanwhile, nuclear-located AtWHIRLY1 can repress SA accumulation during senescence while the plastid version cannot, and a T-DNA insertion line *why1* had higher SA content and early senescence (Lin et al., 2020).

   *WHIRLY1* was shown to have high expression levels in young leaves, suggesting the function in early development. Indeed, WHIRLY1 positively affects chloroplast formation in monocots, such as maize, barley, and rice, as the deficiency of *WHIRLY1* leads to an albino phenotype (Qiu et al., 2022; Prikryl et al., 2008) with severe reduction in chlorophyll content. Missing of WHIRLY1 affected intron splicing of plastid genes, including *atpF* in both barley (Comadira et al., 2015; Melonek et al., 2010) and maize (Prikryl et al., 2008), and *ndhA* in rice (Qiu et al., 2022). Altogether, this leads to the delay in chloroplast development at early seedling stage of monocot mutants of *WHIRLY1*. Still, there is no apparent phenotype in most of dicot mutants, including both gain-of-function or loss-of-function of *WHIRLIES* (Maréchal et al., 2008, 2009; Yoo et al., 2007). However, a small ratio of the double mutant *why1why3* showed variegated leaves due to disturbance of choloroplast development (Maréchal et al., 2009). Furthermore, WHIRLY2, as located in mitochondria, plays a key role in maintainance of genome stability in this organelle (Cai et al., 2015; Maréchal et al., 2008). Knock-out of *WHIRLY2* affected a majority of mitochondria with fewer cristae, larger and varying nucleoids, and low matrix density (Golin et al., 2020).

   The function of WHIRLY1 in regulating senescence is well documented. WHIRLY1 acts as a transcription factor of many nuclear leaf-senescence-associated genes, e.g., *AtWRKY53* (Miao et al., 2013) and *HvS40* (Krupinska et al., 2014a). Interestingly, the function of HvWHIRLY1 in age-dependent or dark-induced senescence is not experimentally proven. But a function under high-light was clearly shown, as the RNAi knock-down of *WHIRLY1* showed delay in senescence (Kucharewicz et al., 2017). The reason could be that WHIRLY1 plays role in light sensing and oxidative stress. The localization of WHIRLY1 also contributes to the senescence-related function. The *Arabidopsis* T-DNA insertion line *why1-1* showed an early senescence phenotype and can be reverted by introducing the WHIRLY1 only in nucleus, suggesting the function of nuclear WHIRLY1 in regulating senescence gene expression (Lin et al., 2019). Meanwhile, in a line that active WHIRLY1 locates only in the plastid, more hydrogen peroxide ($H_2O_2$) accumulated in leaves resulting in early senescence (Lin et al., 2019). Additionally, WHIRLY3 also is involved in senescence regulation. A recent study showed that miRNA840 can post-transcriptionally repress *WHIRLY3* and *PPR* expression, and consequently accelerate senescence (Ren et al., 2022). On other hand, overexpression of *WHIRLY2* in *A. thaliana* led to early senescence with shorter siliques and less seeds

compared to the WT (Maréchal et al., 2008). Recently, the involvement of WHIRLY1 in *A. thaliana* flowering transition was reported (Huang et al., 2022). WHIRLY1 interacts with the histone deacetylase HDA15 to change histone markers of many senescence and flowering related genes. The *WHIRLY1* knock-out line developed flower earlier than the WT, making a connection between flowering transition and nutrient relocalization in leaf senescence.

### 1.1.5. WHIRLIES' functions in stress response

There is a growing body of evidence that WHIRLIES are critical players in response to different environmental stresses, including abiotic stresses, e.g., salt and drought stress (Manh et al., 2023; Ruan et al., 2022; Akbudak & Filiz, 2019; Janack et al., 2016), temperature (Zhuang et al., 2019, 2020), and biotic stresses, e.g., fungus (Akbudak & Filiz, 2019; Desveaux et al., 2000, 2002). WHIRLIES are capable of binding to promoters of a number of genes that encode proteins involved in stress responses and modulating their expression, most likely via interacting with other TFs and/or changing chromatin structure.

The first report of WHIRLIES was StWHIRLY1 activating *StPR-10a* gene in response to an oomycete, *Phytophthora infestans* (Després et al., 1995). After that, *AtWHIRLY1* was shown to be induced upon both avirulent and virulent fungus pathogen infection and it is necessary for the resistance of plant agianst *Peronospora parasitica* (Desveaux et al., 2002, 2004). The knock-out mutant of *ZmWHIRLY1* in maize also had higher susceptibility to the fungus *Ustilago maydis* than the WT (Kretschmer et al., 2017). Meanwhile, tomato WHIRLY1 increases the tolerance of transgenic tobacco against a bacterial pathogen *Pseudomonas solanaceum* (Zhao et al., 2018). All three *WHIRLY* genes in cassava were induced upon flg22 and *Xanthomonas* treatment (Liu et al., 2018). However, expression of all *WHIRLIES* in strawberry was repressed by crown rot infection (Hu & Shu, 2021).

In terms of abiotic stresses, WHIRLY1 in tomato enhanced cold-resistance of the transgenic plant as it induced expression level of genes in starch synthesis and degradation (Zhuang et al., 2019). Furthermore, casava WHIRLIES promote ABA synthesis upon drought stress by interacting with CIPK23 (Yan et al., 2020). Barley WHIRLY1 modifies epigenetic markers on ABA biosynthetic genes and affects drought tolerance (Manh et al., 2023; Janack et al., 2016). Meanwhile, cassava WHIRLIES can bind to *NCED1* promoter and increase ABA as well as drought tolerance (Yan et al., 2020). In potato, WHIRLY2 was necessary for drought resistant as the reduction of SlWHIRLY2 led to more wilting phenotype and reduced fresh weight and chlorophyll content as well as photosynthetic performance (Meng et al., 2020). In *Arabidopsis*, it was also hypothesized that WHIRLY1 can sense redox state of plastids and act in retrograde signaling to regulate

gene expression in nucleus (Foyer et al., 2014). Recently, *AtWHIRLY1* and *AtMYB29* were reported to be  co-expressed in response to UV-B (Zhou et al., 2023).

## 1.2. Post-germination seedling development

### 1.2.1. Germination and seedling development are critical stages of plant's life cycle

Seeds are a highly resistent form in the plant's life that can stay dormant for years, to protect the embryo through harsh or unfavorable conditions. Seeds become active in the process called germination when they perceive positive external cues, demonstrating that the condition is suitable for plant development, e.g., enough water, light, nutrients, and moderate temperature (Bewley et al., 2013). These environmental cues interact with plant genetic landscape in a complex network, leading to reprogramming of gene expression and being the basis for ending dormancy, onset of germination, and eventually seedling development. Physiologically speaking, germination begins when seed rapidly takes up water and becomes swelling, leading to the rupture of membrane structure and solute leakage, characterizing the phase I of germination. Phase II includes the slow and stable water absorbtion, and the initiation of embryo expansion as well as seed-storage nutrient mobilization, resulting in the emergence of the radical at the end of the phase II. Seedling establishment as phase III is the stage where water uptake is resumed and all resources are used up, together with intensive cell division, to promote seedling growth (Figure 3).

During this process, the concerted interplay of phytohormones is critical, in which  two main phytohormones, gibberellins (GA) and abscisic acid (ABA), are involved. Dormant seed contains a high level of ABA and the reduction in ABA level leads to the release of seed from dormancy. Whereas GAs, as counterplayers of ABA, gradually increase the concentration in the germinating seed and young seedling. GAs were reported to be essential in seed-seedling transition as the GA-deficient mutants were unable to germinate without GA supplement (Hauvermale & Steber, 2020). Therefore, it is well-known that the ABA/GA ratio is critical in regulation of metabolic transition during germination. Other hormones, such as ethylene (ET), auxin (AUX), and cytokinin (CK) also influence germination, especially early seedling development (Hu et al., 2017; Corbineau et al., 2014; Wang et al., 2011).
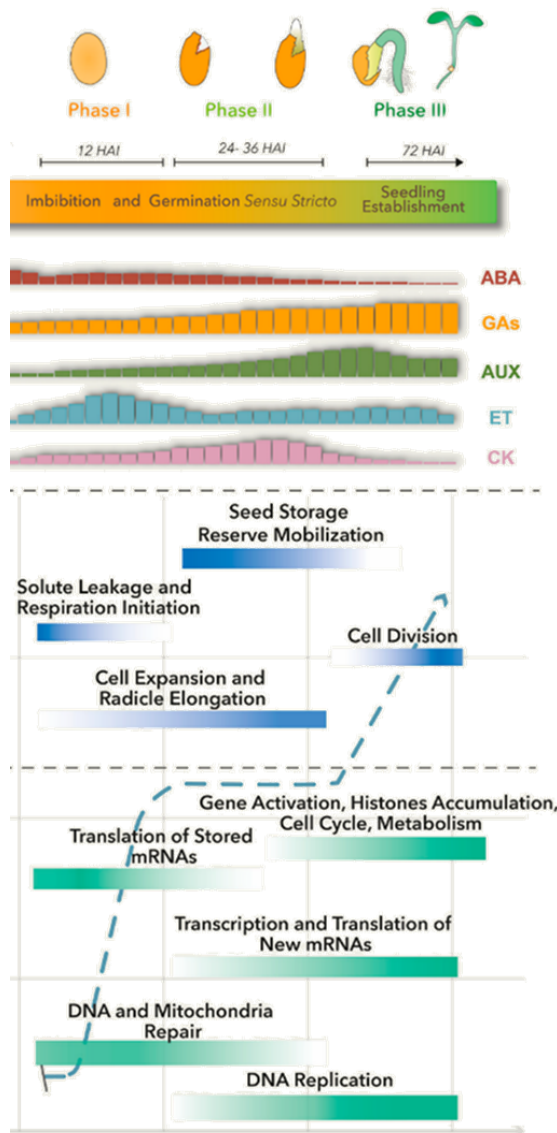
**Figure 3. Molecular events, physiological processes, and phytohormone changes during seed germination and seedling early development.** The upper panel shows germination to seedling establishment stages. The lower panel indicates changes in hormone level with bars indicating hormonal concentration. ABA negatively regulates germination while GA promotes release from seed dormancy and stimulates seedling development. ET production increases after seed absorbing water until the emergence of radical. Bellowing hormone level are indication of different physiological processes and molecular events during these critical stages .The dash blue line illustrates the water level throughout the process. HAI: hour after imbibition. Adapted from Luján-Soto & Dinkova (2021).

In coordination with hormone signaling, various molecular changes happen during seed germination and consequently seedling establishment. At the imbibition stage, proteins involved in redox reaction, glycolysis, and translation are synthesized from seed-stored mRNA, even before re-activating transcription (Sano et al., 2020). On the other hand, pre-existing mRNA that are related to ABA response and dormancy are degraded. Additionally, early and precise DNA repair is essential for germination as well as early seedling development (Waterworth et al., 2015). Meanwhile, phase II is characterized by the re-activation of transcription and also novel protein synthesis, however it was shown that novel proteins are not require for completion of the phase II (Rajjou et al., 2004). This stage also witnesses the mobilization of several metabolic intermediates, such as lipids, sugars, organic acids, and amino acids, which are accumulated after seed desication. These breakdown products provide energy and components to newly synthesis nucleic acid, protein, and other metabolites, before an autotrophic organism is formed. Finally,

seedling establishment (phase III) requires the transcription activation of numerous genes associated to various processes, such as DNA replication, transcription and translation, cell division, and metabolic pathways. Seedling growth is driven by cell replication and expansion, and needs both endogenous storages and exogenous resources to successfully establish in the new environment.

### 1.2.2. Gene regulation in post-germination seedling development

In most species, seed germination happens in the dark and consequently forms an etiolated seedling. Upon light exposure, seedling undergoes light-induced morphogenesis, chlorophyll production, and therefore becomes green. Thanks to these physiological changes, young seedlings can self-sustain as an autophototropic organism. Hence, under normal growth condition, seedling establishment is affected mostly by light signaling. Additionally, during seedling development, any dramatic changes in environmental conditions, e.g., light intensity, temperature, etc., can be perceived by a fine-tuning concert of different hormones, such as ABA, auxin, ethylene, etc., to consequently activate a complex but harmonizing gene regulatory network, so that seedlings can response precisely.

Different regulatory levels together affect a gene functionality, from epigenetical, transcriptional, and post-transcriptional, to translational regulation. Transcriptional regulation is most intensively studied, as it directly affects the abundance of a gene product. In this regulatory mechanism, transcription factors (TFs) play a central role; and especially for young seedlings, TFs which intergrate light and hormone signaling are crucial for successful seedling establishment. These regulators include well-known light-mediated TFs, such as ELONGATED COTYLEDON 5 (HY5) and PHYTOCHROME-INTERACTING FACTORs (PIFs). These TFs are regulated by an universal negative controller of light response, a RING E3 ubiquitin liagse CONSTITUTIVE PHOTOMORPHOGENIC (COP1), which can ubiquitinate these TFs and affect their stability and functionality (Kim et al., 2017).

HY5, a master TF of positive response to various light conditions (Xiao et al., 2022; Lau & Deng, 2010), promotes auxin signaling by transcriptional downregulation of *IAA7* and *IAA14*, which are negative auxin signaling genes. HY5 also interacts with BRASSINAZOLE-RESISTANT 1 (BZR1), a downstream TFs of brassinosteroids pathway, to antagonisticallly regulate genes involved in cotyledon opening (Li & He, 2016). It was reported that many HY5 target genes are involved in GA catabolism. The reduction in GA levels in response to light stablizes DELLA protein, leading to photomorphogenesis (Gangappa & Botto, 2016). On the other hand, HY5 together with RSM1 induce *ABI5*

expression, a master TF in ABA metabolism, and affect germination and early seedling development (Yang et al., 2018).

However, PIFs are bHLH TFs that were innitially identified as phytochrome-interacting partners. PIFs repress photosynthesis/chloroplast-related genes but induce cell wall and hormone signaling related genes (Leivar & Monte, 2014). Similar as HY5, PIFs are also in the cross talk between hormonal and environmental signaling. PIF4 interacts with auxin response factor ARF6, and also BZR1 of brassinosteroid signaling, to induce their common cell-elongation-related target genes (Oh et al., 2014). Meanwhile, GA signaling influence PIFs activity via DELLA proteins, and the interaction between DELLAs and PIFs prevents PIFs from binding to target genes and induces PIF degradation (Leivar & Monte, 2014). Additionally, PIF1 also represses GA biosynthetic genes, while induces DELLA genes, as well as ABA biosynthetic genes (Leivar & Monte, 2014). It was reported that PIFs can bind to G-box motifs on *ABI5* promoter and induce its transcription (Yadukrishnan & Datta, 2021). Under the light, these PIFs were inhibited, leading to lower the expression level of *ABI5*, and thereby suppressed the influence of ABA toward young seedling development.

Alternative splicing (AS), as a post-transcriptional regulatory level can generate multiple mRNA transcripts from a single multi-exon gene by using splice sites differently (Kashkan et al., 2022). There are four main types of AS: exon skipping (ES), retained intron (RI), alternative 5' splice sites (A5SS), and alternative 3' splice sites (A3SS), in which each can have different impact on functionality of gene products (review by Kashkan et al., 2022). Therefore, beyond transcription regulation, AS plays a big role in quantitative and qualitative regulation of mRNA variants derived from a gene. AS during early seedling development can regulate responses to environmental cue, such as light, heat, nutrient avaiílbility, etc. Various light-regulated alternative splicing events in early seedling development were discussed in an excellent review (Kathare & Huq, 2021). It was reported that chloroplasts, upon receiving light signal, can start a retrograde signaling to regulate AS of numerous light-regulated genes in nucleus; and the regulation is likely by controlling the transcriptional elongation rate (Hartmann et al., 2016; Petrillo et al., 2014). Genome-wide analysis of phytochrome-dependent AS revealed many splicing factor genes and photosynthesis- and plastid-related genes were enriched, suggesting that AS plays a big role in chloroplast development in response to light (Shikata et al., 2014). However, details on molecular components of these regulation are still in the early stage.

In AS regulation, splicing factors play a critical role as they can enhance or repress the usage of a splicing site, leading to differential splicing. Several RNA splicing factors were reported to be involved in early seedling development, such as SERINE/ARGININE-RICH PROTEINS (SR34b, RS31, RS41, etc.) that showed a diverse splicing pattern in

seedlings of different ages (Palusa et al., 2007). Another example is MERISTEM DEFECTIVE (MDF), a homolog of the human SQUAMOUS CELL CARCINOMA ANTIGEN RECOGNIZED BY T CELLS 1 (SART1), which is necessary for the interaction between spliceosome components. MDF is critical for AS of numerous genes in developmental and signaling, especially in cell division at early development (de Luxán-Hernández et al., 2023). Besides splicing factors, several genes related to morphogenesis during seedlings establishment were reported to be targets of AS regulation. One example is COP1, the master regulator of morphogenesis during germination and seedling development (Kim et al., 2017). The AS isoform of *COP1*, named *COP1b*, lacks 60 aa of the WD domain, and therefore exerts a negative effect on COP1 functionality (Zhou et al., 1998). In *Arabidopsis*, *HY5 HOMOLOG* (*HYH*) presents four splicing variants, all containing the bZIP domain for transcriptional activation but the COP1-interacting domain is varied, suggesting different *HYH* isoforms having different stability. Besides, all AS-derived *HYH* isoforms can compensate function of HY5 in hypocotyl elongation in the *HY5* knock-out mutant (Li et al., 2017a).

### 1.2.3.  Gene regulation in response to oxidative stress

Plants constantly produce reactive oxygen species (ROS) as products of metabolism. Though basal ROS level is necessary for a normal development, the excess of ROS, so called oxidative stress, can damage cellular molecules via diverse processes, such as protein oxidation, lipid peroxidation, enzyme inactivation, chlorophyll degradation, and destruction of nucleic acids. Therefore, plants have developed a complex signaling network with a profound set of ROS-responsive genes, which are required for repairing ROS-caused cell damage and also detoxifying ROS. These outcomes heavily depend on the generated ROS type, the subcellular origin, and the exposure time. In the present work, the focus is ROS produced in chloroplasts, mostly due to interaction between chlorophyll and high light intensity (Figure 4). With the experimental set-up in this work, excess light energy absorbed by chlorophyl molecules in photosystems induces ROS productions, including singlet oxygen $^1O_2$ and hydrogen peroxide ($H_2O_2$). These ROS likely activate antagonizing but harmonizing retrograde signalings from chloroplasts to nucleus and modulate gene expression.

The accumulation of ROS upon stress can lead to specific changes in the transcriptome.  The signature and common genes of different oxidative stress origin are addressed in two excellent works (Willems et al., 2016; Rosenwasser et al., 2013), in which ROS-specific genes were classified into ROS wheel and Rosmeter. In general, response to oxidative stress involves activation of multiple gene families, including genes encoding

components of antioxidant systems, either enzymatic-proteins, e.g., superoxide dismutase (SOD), peroxidase (POD), glutathione peroxidase (GPX), etc., or non-enzymatic proteins, e.g., flavonoid, glutathione (GSH), etc.; genes encoding signaling components such as $Ca^{2+}$ dependent protein and receptor-like kinases, and genes necessary for repair or replacement of cellular damage, such as chloroplast-associated genes (He et al., 2018; Willems et al., 2016). The regulation of expression of these genes depends mainly on various TF families that interact with oxidative-stress-responsive *cis*-elements on promoter regions. In *Arabidopsis*, the tryptophan (W)–arginine (R)–lysine (K)–tyrosine (Y) (WRKY) family and APETALA2/ethylene response factor (AP2/ERF) are the keys TFs, which recognize W-box (TTGAC/T) and the GCC box, respectively (Petrov et al., 2012). Besides, other TF families are also regulated in a wide range of oxidative stress, such as MYELOBLASTOSIS (MYB), NAM/ATAF/CUC (NAC), HEATSHOCK FACTOR (HSF), or ZINC FINGER PROTEIN (ZAT).
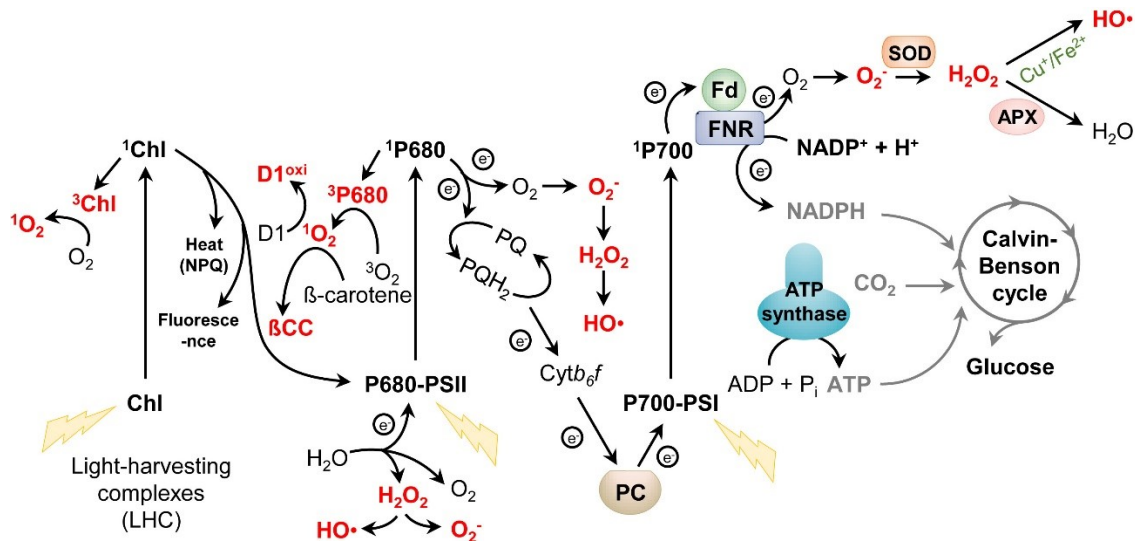


**Figure 4. Generation of reactive oxygen species (ROS) in chloroplast.** Upon excess light energy absorption, chlorophyll (Chl) in light-harvesting complexes (LHC) and P680 Chl in the reaction center of photosystem II (PSII RC) convert to the single-state $^1$Chl and $^1$P680, and spontaneously decay into triplet excited state $^3$Chl and $^3$P680 with an increased lifespan, respectively. These excited forms then transfer the energy to ground-state oxygen $O_2$, producing singlet oxygen $^1O_2$. Meanwhile, $H_2O_2$ can be generated by oxidation at the electron donor side of PSII. $^1$P680 at PSII and ferredoxin (Fd)-NADP$^+$ oxidoreductase (FNR) at PSI can transfer electron to $O_2$ to produce $O_2^-$. $O_2^-$ is further dismutased to $H_2O_2$ either simultaneously or via enzymatically reactions with SUPEROXIDE DISMUTASES (SODs). $H_2O_2$ also can convert to $O_2^-$ and HO•. Adapted from Li & Kim (2022).

Meanwhile, alternative splicing regulation in response to oxidative stress has not been studied well in *Arabidopsis*. However, there is emerging evidence showing that AS is also one key regulatory step in response to disturbing environmental stimuli. The role of AS in various abiotic stress responses has been reported (Calixto et al., 2018; Laloum et al., 2018; Cruz et al., 2014; Palusa et al., 2007). Most of AS events in response to abiotic

stress occur in genes with regulatory functions, including RNA metabolism such as mRNA synthesis and splicing, DNA repair, and phosphorylation (Martín et al., 2021). Splicing factors themselves are key regulators of AS in response to environmental stressors, as several splicing factors showed a fast and early response to changes in environment at transcriptional level as well as post-transcriptional level (Butt et al., 2022; Calixto et al., 2018; Palusa et al., 2007). Many SR splicing factors were shown to play function in stress response, such as SR34b in response to cadmium (Zhang et al., 2014), two SC35-like genes (SCL) in ABA response (Cruz et al., 2014), SR45a in salt stress (Li et al., 2021), and *Vitis vinifera* SR34, SR45, and SR30 in response to high temperature (Jiang et al., 2017a). Other splicing factors, for instance, U5 snRNP-associated STABILIZED1 (STA1) is involved in cold tolerance (Lee et al., 2006) and SKIP functions in AS of salt-responsive genes (Feng et al., 2015).

Furthermore, in response to abiotic stress, the ratio of different variants of stress-responsive genes is altered to optimize the functionlity of these genes under different conditions, since generally different proteins from AS-derived variants often have similar interaction target but most likely exert a different activity. One example is *CBL-INTERACTING PROTEIN KINASES 3* (*CIPK3*), which produced several AS variants in response to ABA treatment or drought stress (Liu et al., 2013). Among them, *CIPK3.1* and *CIPK3.4* were relatively more abundance than others in response to the treatment. Besides, CIPK3.1 has higher afinity to ABR1, an interactor of CIPK3, suggesting different abundance of AS variants can affect the selection of interacting partners in downstream pathway in regulation of ABA and other stress response (Sanyal et al., 2017). A key regulator in response to heat stress is *HEAT SHOCK TRANSCRIPTION FACTOR A2 (HSFA2)* producing an AS variant *HsfA2-III* in *Arabidopsis* seedling under heat stress. This new variant encodes a truncated peptide lacking the C-terminal activation domain, leading to a loss of transactivation function of HsfA2-III (Liu et al., 2013), and thereby affects plant response to heat stress.

# AIM OF DISSERTATION

Although function of AtWHIRLY1 at mature developmental stages has been well studied (Krupinska et al., 2022; Taylor et al., 2022), little is known about its function at the early stage of development in *Arabidopsis*, such as in young seedlings. Additionally, the capacity to bind to different types of nucleic acid as well as the dual localization between chloroplast and nucleus suggest that AtWHIRLY1 function is more diverse and can be transferable between cellular compartments. Therefore, this dissertation focuses on analyzing young seedlings of CRISPR/Cas9-based *WHIRLY1* knock-out mutants as a reverse genetic approach. By performing transcriptomic analysis at this early stage between the mutant and the WT, the overall aim is to gain insights into function of AtWHIRLY1 in nuclear gene regulation. Specifically, I aim to answer following questions:

1. **Does loss-of-function of WHIRLY1 cause an aberrant phenotype in young seedlings under control growth condition and under oxidative stress?** To answer the question, several physiological parameters are recorded in 5-day-old seedlings, including the two *WHIRLY1* knock-out mutants and the WT.

2. **How does loss-of-function of WHIRLY1 affect genome-wide gene expression and alternative splicing in young seedlings under control growth condition?** This question is addressed by performing comparative RNA-seq analysis of 5-day-old *WHIRLY1* knock-out and WT seedlings. Expressed genes in two lines are compared to find genes which are deregulated due to missing of WHIRLY1. Besides, the alternative splicing pattern of nuclear genes is investigated. Additionally, metabolite profile of *WHIRLY1* knock-out seedlings is further analyzed, focusing on glucosinolates.

3. **How do young seedlings respond to oxidative stress?** For this question, transcriptome of seedlings under oxidative stress is compared to that one in control condition. Hence, differentially expressed genes and differentially alternative spliced genes, which are related to stress responses, are identified and compared between the two genetic backgrounds.

4. **How does loss-of-function of *WHIRLY1* affect gene expression and alternative splicing in young seedlings under oxidative stress?** By comparing directly transcriptomes of the *WHIRLY1* knock-out and the WT seedlings grown under stress conditions, genes which might be regulated by WHIRLY1 under oxidative stress are revealed, both in term of gene expression and alternative splicing regulation.

## 2. RESULT

The early phase after germination is important and critical for plant development. It involves development of the young seedlings, including establishment of photosynthesis in the emerging cotyledons and growth of the root system to take up needed nutrients. Additionally, it is a sensitive phase where plants must be protected against various biotic and abiotic stresses. This stage also witnesses a massive reprogramming of gene expression at multiple levels. In the present work, the involvement of AtWHIRLY1 was examined on post-germination early seedling development of *A. thaliana* under control conditions, which is described in chapter 2.1. On the other hand, how AtWHIRLY1 functions in an abiotic stress response, namely oxidative stress, in this early stage of development was also investigated in chapter 2.2. Approaches used in the present work include (1) phenotype characterization, (2) transcriptome analysis, including gene expression level and alternative splicing, and (3) metabolite measurement.

In order to tackle these forementioned topics, I used a reverse genetic approach with two CRISPR/Cas9 *WHIRLY1* knock-out lines (*why1-4* and *why1-5*), which were created in the working group of Prof. Götz Hensel (Heinrich-Heine-University Düsseldorf, Germany). These two lines contain frame-shift mutations at the beginning of coding sequence, leading to altered encoded proteins (Figure 5A). These mutants were verified using PCR-coupled CAPS analysis (Figure 5B), showing they are homologous of the mutated alleles. Besides, the presence of Cas9 coding sequence in these mutants was examined by PCR with specific primers, showing both lines are Cas9-free (data not shown). The knock-out of *WHIRLY1* by CRISPR/Cas9 resulted to substantially reduction of *WHIRLY1* transcript level compared to the WT (Figure 5C). Additionally, expression level of two other *AtWHIRLIES* were similar as the WT level.
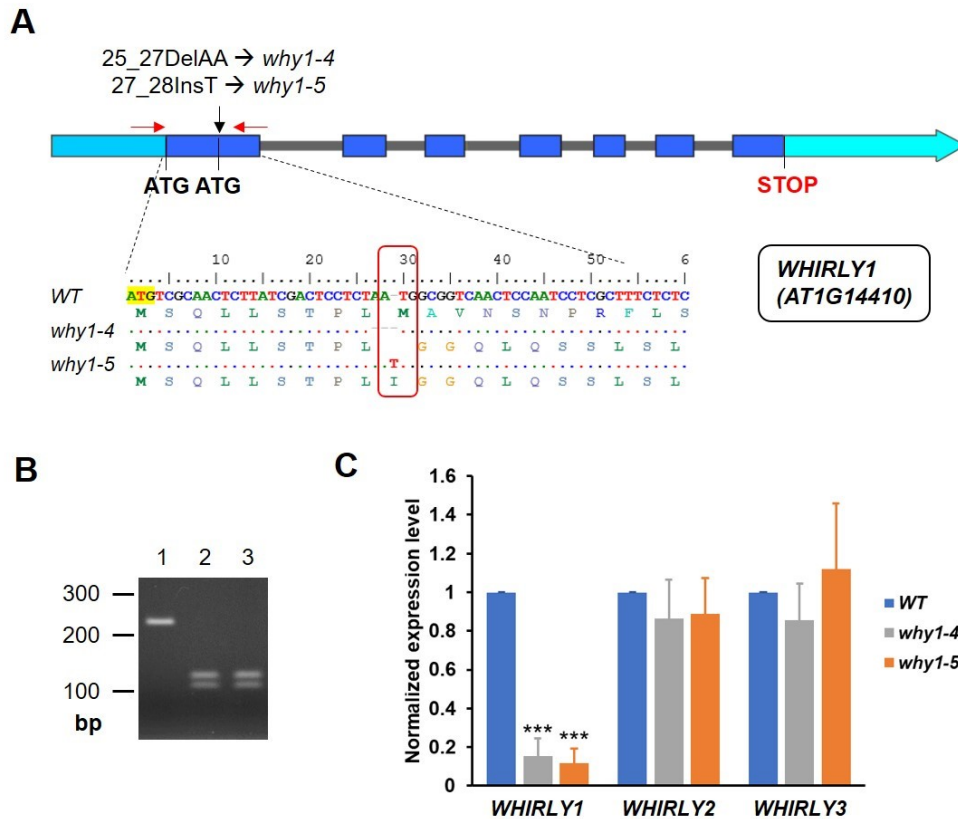
**Figure 5. Genotyping of CRISPR/Cas9 mutants of *AtWHIRLY1*. (A**) Scheme of two *WHIRLY1* knock-out mutants in the present work. (**B**) Two mutants were verified using cleavage amplified polymorphism site (CAPS) approach. WT (lane 1) resulted in a non-cut fragment (231 bp), while the homologous mutants *why1-4* (lane 2) and *why1-5* (lane 3) showed two bands of 126 bp and 105 bp. (**C**) Normalized expression level of three *WHIRLY* genes in two knock-out mutants by qRT-PCR compared to the WT. Bar chart represents average of mean and error bars indicate standard error of four independent biological replicates. Asterisks denote statistically significant level of Student's t-test between qRT-PCR-based normalized expression level of genes in *WHIRLY1* mutants and the WT, (*) p-value < 0.05, (**) p-value < 0.01, (***) p-value < 0.001.

## 2.1. Function of AtWHIRLY1 in post-germination early seedling development under control conditions

### 2.1.1. Phenotype of *AtWHIRLY1* knock-out mutants at early developmental stage under control conditions

At five days after germination (DAG), plants were pictured and their development was recorded via different parameters, including hypocotyl length, primary root length, fresh weight, maximum PSII efficiency, and chlorophyll content. Morphologically, the two *AtWHIRLY1* knock-out lines did not exhibit any noticeable phenotype compared to the WT. All seedlings from different genetic backgrounds displayed fully developed green cotyledons and a primary root, without visible true leaves at this stage (Figure 6A). The primary root length average was around 1.5 cm, and the hypocotyl length was approximately 0.3 cm under this growth conditions (Figure 6B-C). The three lines had also similar fresh weight (Figure 6D).
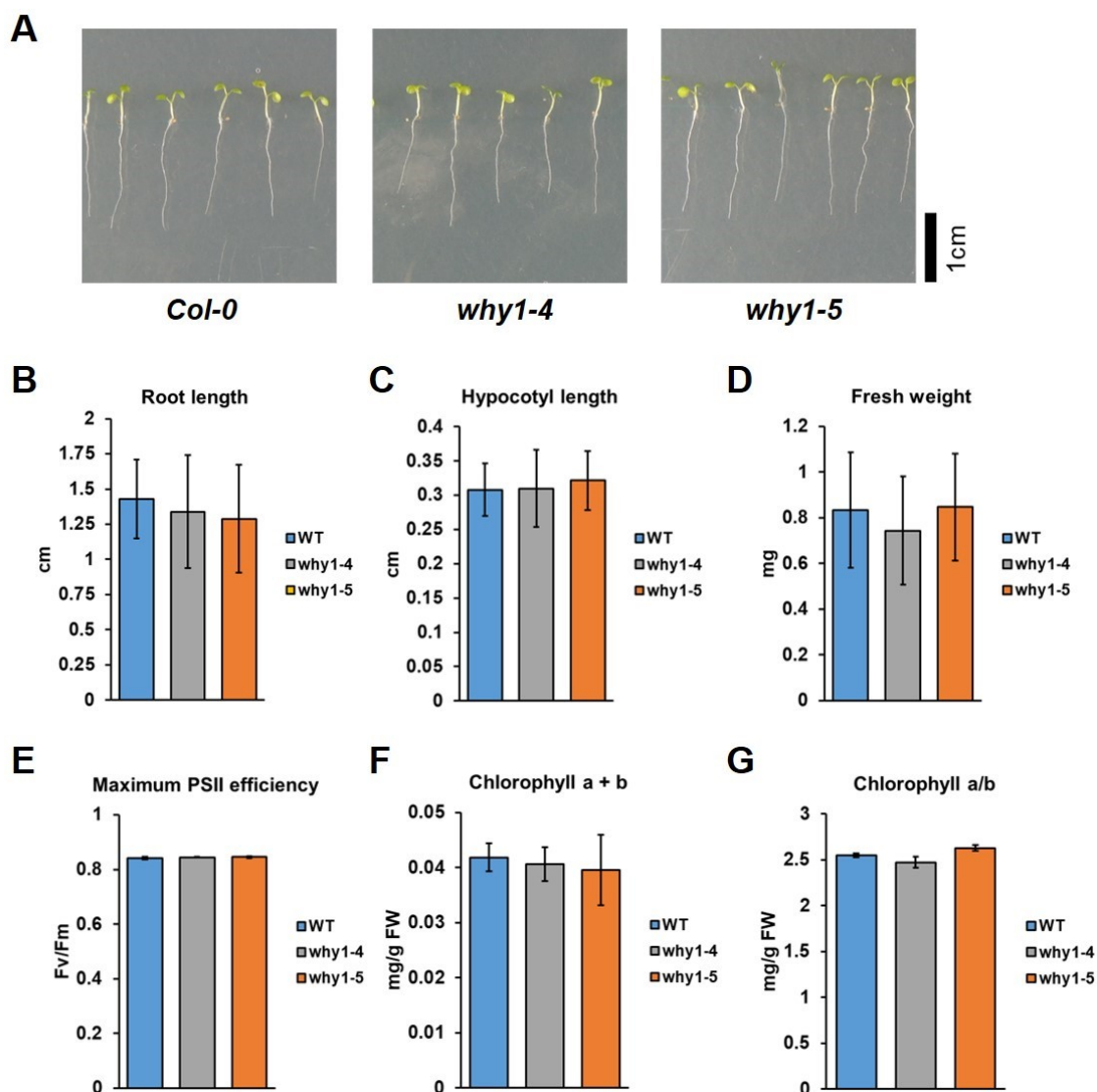
**Figure 6. Phenotype of *WHIRLY1* knock-out mutants at early stage of development.** (**A**) Picture of 5-day-old WT and two CRISPR/Cas9 *WHIRLY1* knock-out mutants, *why1-4* and *why1-5*. (**B**) Root length, (**C**) Hypocotyl length, (**D**) Whole seedlings fresh weight, (**E**) Maximum PSII efficiency (Fv/Fm), (**F**) Chlorophyll a, and (**G**) Chlorophyll b content of 5-day-old WT and *WHIRLY1* mutants. The bar charts show the average value of 4 biological replicate means. Error bars represent standard deviation of 4 biological replicates, each containing at least 20 seedlings.

Similarly, the photosynthetic performance of the knock-out lines, *why1-4* and *why1-5*, examined by measurements of maximum PSII efficiency and the chlorophyll content, showed similar results compared to the WT (Figure 6E-G). These parameters indicate that the loss-of-function of *WHIRLY1* appears to not significantly influence the vegetative growth and the development of photosynthetic capacity during early development in *A. thaliana*, at least under the growth condtions in the present work. It is worth noting that the results shown in Figure 6 were obtained with seedlings grown under continuous low light intensity (15 µE/m²s), which was the growth condition used for the oxidative stress experiment, which will be discussed later. However, under normal growth conditions (100

μE/m$^2$s of light intensity, 16h:8h of light-dark cycle), loss of WHIRLY1 function also did not show any effect on seedling development (data not shown).

Similar to my results, previous studies in dicots have shown that single or double mutants of *WHIRLIES* did not show apparent phenotypic changes but display abnormalities in subcellular structure or gene regulation (Golin et al., 2020; Zhuang et al., 2019; Maréchal et al., 2009). Therefore, to gain further insights about WHIRLY1 function at early seedling development, transcriptome of the *WHIRLY1* knock-out mutant was investigated to clarify the role of WHIRLY1 in reprogramming of gene expression. Due to both CRISPR/Cas9-based *WHIRLY1* knock-out lines have frameshift mutated alleles and also exhibited similar phenotypes, the line *why1-5* (hereafter mentioned as *why1*) was selected for RNA-sequencing (RNA-seq) approach.

### 2.1.2. Differentially expressed genes in the *WHIRLY1* knock-out mutant compared to the WT at control condition

The RNA-seq analysis revealed that 5 days after germination, both the *WHIRLY1* knock-out mutant, *why1-5*, and the WT expressed nearly 15 thousand genes (with an expression level above the threshold, i.e., fpkm value > 1). To identify the differences in gene expression between two lines, differential gene expression analysis was performed using the DESeq2 program. As the result, 73 upregulated differentially expressed gene (DEGs) and 42 downregulated DEGs were identified in the knock-out mutant *why1-5* compared to the WT (as p-value < 0.01; |log2FC| > 1), indicating that WHIRLY1 rather specifically regulates target genes as both a positive and a negative regulator. The volcano plot, displaying distribution of all DEGs between the *WHIRLY1* knock-out mutant and the WT, is shown in Figure 7A.

Gene Ontology (GO) enrichment analysis was performed to gain insights into the functional processes associated with these DEGs. However, the analysis failed to give any significant GO terms for upregulated genes in the mutant *why1-5*. In contrast, a small set of 42 significantly downregulated DEGs due to loss of *WHIRLY1* function showed a significant enrichment in GO terms related to sulfur-containing compound biosynthesis, specifically glucosinolates with a very high enrichment fold (Figure 7B).
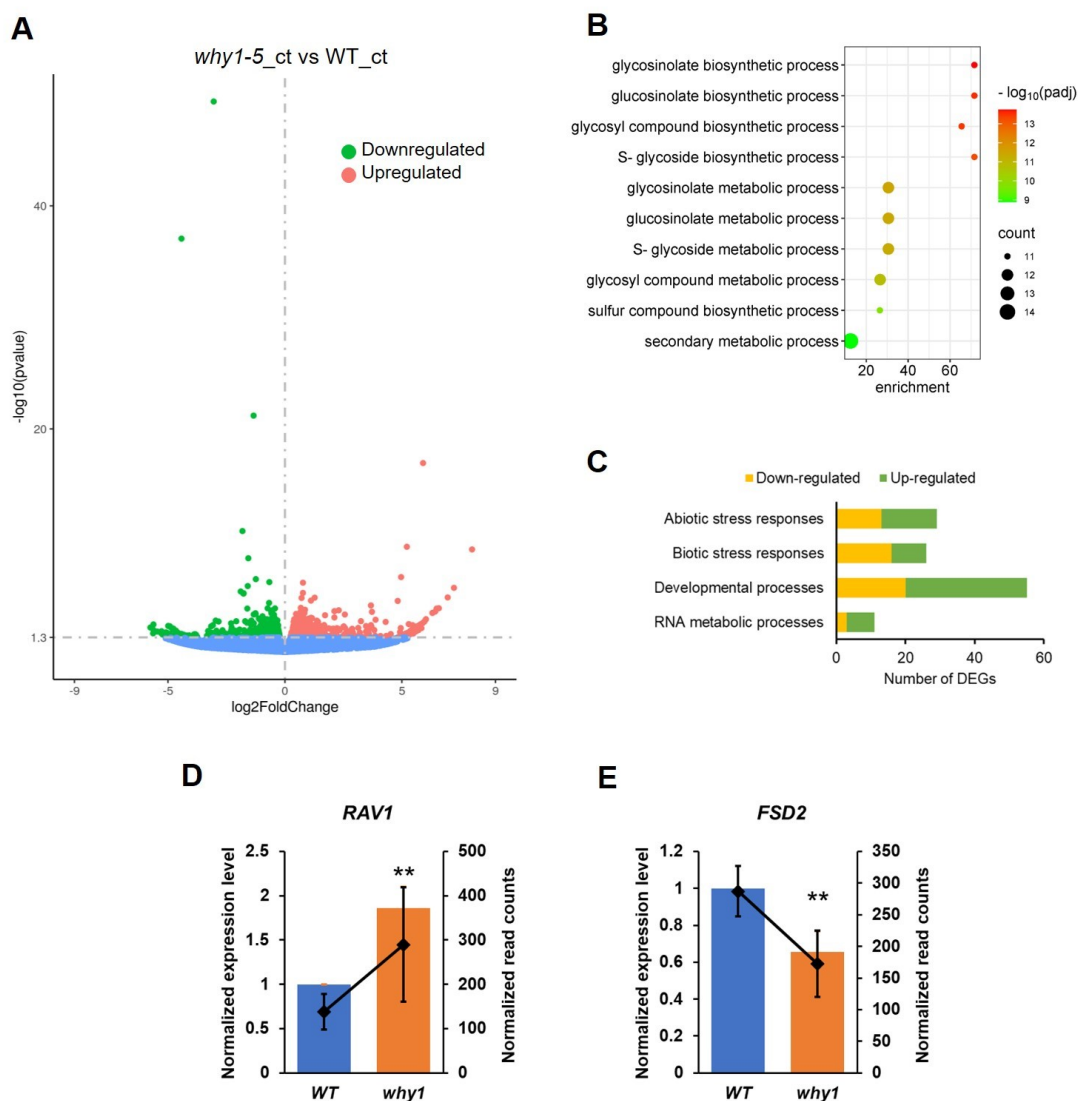
**Figure 7. Differentially expressed genes (DEGs) analysis between the knock-out mutant *why1-5* and the WT**. (**A**) Volcano plot showing distribution of DEGs between the mutant *why1-5* compared to the WT. (**B**) GO enrichment analysis of downregulated DEGs. Top 10 GO terms according to p-adj value are shown (**C**) Distribution of DEGs according to their biological functions. Validation of the expression level of *RAV1* (**D**) and *FSD2* (**E**) in seedlings by qRT-PCR. Normalized expression level by qRT-PCR (bar charts on left vertical axis) and normalized read counts by RNA-seq (line chart on the right vertical axis), both show the average of mean and error bars indicate standard deviation of three independent biological replicates. Asterisks denote statistically significant level of Student's t-test between qRT-PCR-based normalized expression level of selected genes in the mutant *why1-5* and WT, (*) p-value < 0.05, (**) p-value < 0.01, (***) p-value < 0.001.

Upon manual search in databases, such as The *Arabidopsis* Information Resource (TAIR), (www.*Arabidopsis*.org) and UniProt, (https://www.uniprot.org/), it was found that around 25% of DEGs are related to abiotic stress response, 21% to biotic stress response, 47.8% to developmental process, and nearly 10% to RNA metabolism (as summarized in Figure 7C). This indicates that the knock-out of *WHIRLY1* specifically affects the reprogramming of stress-responses and developmental processes during post-germinative developmental stage. Further investigation of these DEGs (Appendix Table

S1) revealed that many of them are involved in responding to abiotic environmental cues such as heat, drought, oxidative stress, etc. Search in databases also indicated that these genes are often related to abscisic acid (ABA) signaling, which plays a crucial role in abiotic stress responses. Meanwhile, a similar number of DEGs was found to be involved in signaling pathways associated with biotic stress, annotated as disease resistance, defense response to other organisms, immune system response, disease resistance protein, etc. These genes are often connected to the signaling pathways of jasmonic acid (JA), salicylic acid (SA), or ethylene (ET). A third class of DEGs is linked to various developmental processes, including amino acid biosynthesis, transportation, cell cycle, tissue development, and DNA and RNA metabolism. Many of these DEGs are known to be involved in auxin and cytokinin signaling pathways. Noteworthy, around 10% of the DEGs are involved in ribosomal function and RNA metabolism. Among these genes, 8 out of 11 were induced in the knock-out *WHIRLY1* mutant, and they function by binding to mRNA and rRNA.

Among the DEGs affected by the knock-out of *WHIRLY1*, several examples of multifunctional factors involved in upstream cross-talk between abiotic stress, biotic stress, and development pathways were identified (Appendix Table S1). An example is RAV1 (*AT1G13260*), which functions as a hub in ABA, ethylene, and cytokinin-mediated signaling pathways (Mandal et al., 2023; Feng et al., 2014; Kagaya et al., 1999). In *A. thaliana*, RAV1 was characterized as a positive regulator of leaf senescence (Woo et al., 2010) and a negative regulator of root meristem size (Mandal et al., 2023). RAV1 is also involved in plant defense against the bacterial pathogen *P. syringea* (Sohn et al., 2006)*,* and necrotrophic fungi such as *Rhizoctonia solani* and *Botrytis cinerea*. Missing of WHIRLY1 as an RNA-seq result induced its expression level, which was validated by qRT-PCR (Figure 7D), suggesting WHIRLY1 may act as a negative regulator toward this gene. On the other hand, knock-out of *WHIRLY1* in the seedling resulted in suppression of an Fe superoxide dismutase encoding gene, *FSD2* (Figure 7E). FSD2 is important for ROS detoxification in plastid, and it was shown that the loss of *FSD2* expression resulted in a lower photosynthetic activity and an overproduction of superoxide. Furthermore, FSD2 is necessary for protection of PSII from photodamage (Gallie & Chen, 2019). WHIRLY1 seems to positively regulate *FSD2* level at the early stage of development, perhaps connected to the WHIRLY1's function in plastid biogenesis (Qiu et al., 2022; Melonek et al., 2010; Prikryl et al., 2008).

### 2.1.3. Loss-of-function of *AtWHIRLY1* affected RNA splicing in seedlings

It was reported that WHIRLY1 in monocots participate in plastid RNA splicing (Melonek et al., 2010; Prikryl et al., 2008). Therefore, changes in alternative splicing (AS) pattern of nuclear genes was investigated in the loss-of-function of *AtWHIRLY1* seedlings. Indeed, knock-out of *WHIRLY1* caused major changes in splicing events in *A. thaliana*. In total, 1367 differentially alternative splicing events (DASEs) ($|\Delta PSI| > 0.15$, p-value $< 0.01$) were identified in 1043 genes, which can be classified into five categories: exon skipping (SE), retained introns (RI), alternative 3′ splice sites (A3SS), alternative 5′ splice sites (A5SS), and mutually exclusive exons (MXE) (Figure 8A). Intron retention is the most recorded AS type with 770 significantly DASEs, in which 353 increased and 417 decreased the percent spliced in (PSI) in the mutant *why1-5* compared to the WT. Next are A3SS and A5SS with 259 and 184 events in total, respectively. Meanwhile, only 2 events of increasing mutually exclusive exons were differentially AS in the *why1* line compared to the WT (Figure 8B). Interestingly, only 4 differentially alternative spliced genes (DASGs) were also differentially expressed, indicating that the missing of WHIRLY1 affected alternative splicing independently from regulating gene transcription level.

GO enrichment analysis of DASGs revealed that knock-out of *WHIRLY1* affected alternative splicing of various gene groups that are involved in different regulatory pathways (Figure 8C). For instance, many GO terms related to response to abiotic stimuli, including radiation, were overrepresented. These functional groups are connected to "DNA repair", suggesting function of these genes on DNA damage response due to abiotic stresses. Besides, many DNA and RNA-related GO terms (depicted diamond and triangle in Figure 8C) were enriched among DASGs between the knock-out line *why1*-5 and the WT, including aforementioned DNA repair and other nucleic acid metabolic processes, such as transcription and RNA processing. Furthermore, DASGs also overrepresent with regulational functions (purple letters), including post-transcriptional regulation of gene expression and also RNA splicing. These are equivalent to the molecular function GO term of "catalytic activity, acting on DNA". Remarkably, RNA processing-related functional groups were overrepresented with high enrichment levels. These results explain why a high ratio of DASGs encoding proteins located in "nucleus" and "nuclear protein-containing complex". Interestingly, many RNA splicing factors and associated proteins were found to be spliced themselves differentially in the *WHIRLY1* knock-out mutant compared to the WT (Appendix Table S2). Moreover, a high number of DASGs are involved in "phosphorylation" biological process, thereby have kinase or transferase molecular function.

.
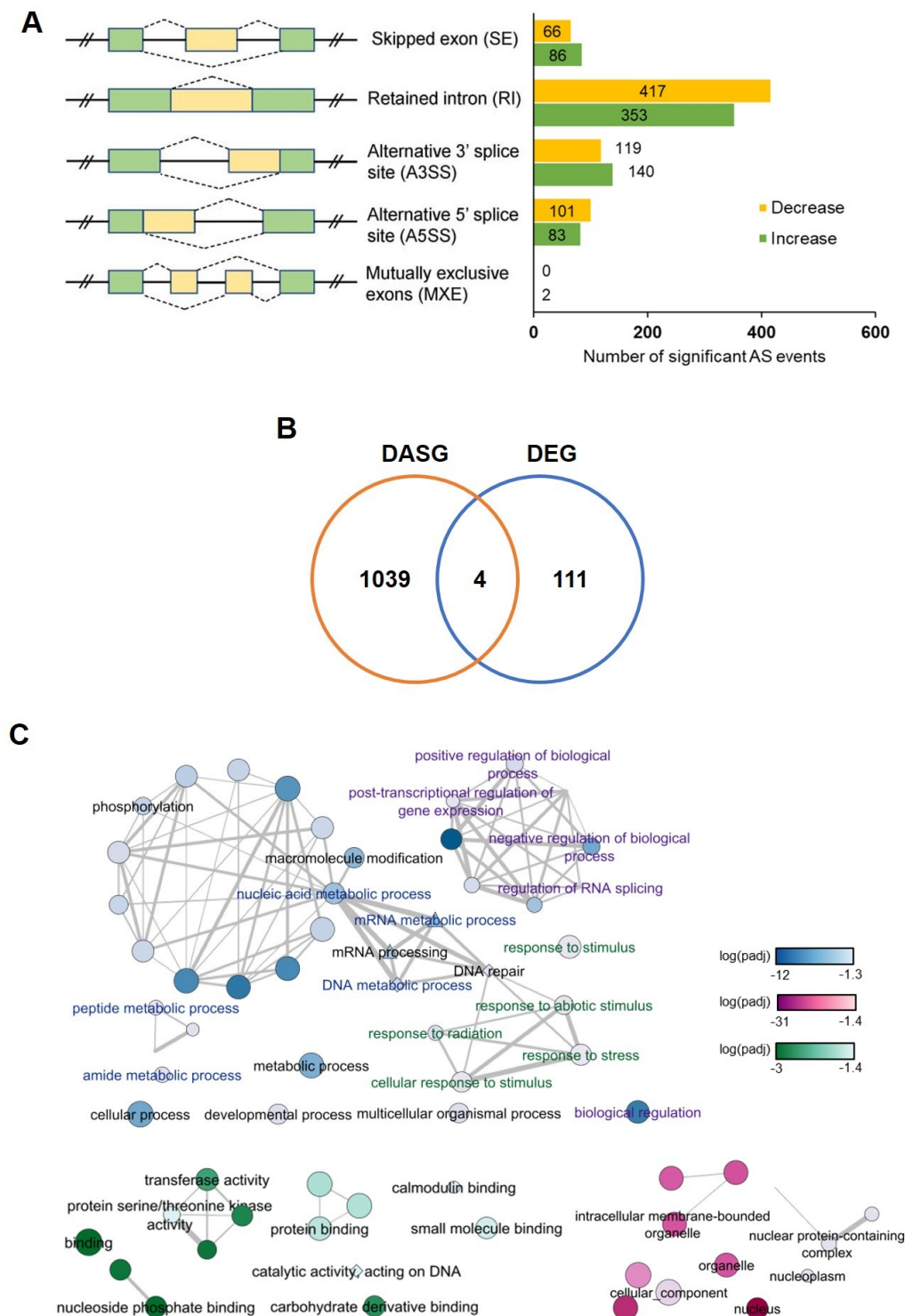
**Figure 8. Alternative splicing in the *WHIRLY1* knock-out mutant.** (**A**) Distribution of AS events that increased or decreased PSI in the knock-out seedlings *why1-5* compared to the WT. These types include exon skipping (SE), alternative 5′ splice site (A5SS), alternative 3′ splice site (A3SS), mutually exclusive exons (MXE), and retained intron (RI). The total number of significant DAS events, with only the reads span splicing junctions were taken into account. (**B**) Venn diagram represents overlap between differentially expressed genes (DEGs) and differentially alternative spliced genes (DASGs). (**C**) GO enrichment analysis of DASGs (p-adj < 0.05). Each node indicates a GO term, and related GO terms are connected by lines. Different colors of node illustrate GO categories (blue, biological process; green, molecular function; pink, cellular compartment) and different color shades show log(p-adj) of enriched GO term. Size of node reflects size of GO term in the whole genome.

Selected DASEs due to loss-of-function of *WHIRLY1* were validated by RT-PCR using primers flanking the splicing sites (Figure 9). SR34b, an RNA splicing factor (Cruz et al., 2014), increased retained intron isoform in the line *why1-5* compared to the WT (Figure 9A). The intron-inclusion variant introduces a premature stop codon at the C-terminal, resulting to a 19-aa-shorter peptide than the intron-exclusion isoform. More importantly, the intron adds an addition of 3'UTR segment, which might impact the mRNA localization, mRNA stability, and translation (Mayr, 2019). Another example is BPL2, an RNA-binding protein and a binding partner of ACD11 in negatively regulation of ROS-mediated defense response (Li et al., 2019). The loss of WHIRLY1 resulted in an increase of the A3SS variant, producing a protein with one additional glutamine at the 116 position, right after the RNA-binding domain (Figure 9B). This extra glutamine residue can interrupt the alpha-helix loop that might influence the interaction of BPL2 with ACD11 in stress response (Li et al., 2019). Besides, two other DNA bands with bigger size were observed (Figure 9B), they are either novel alternative splicing variants of *BPL2* or unspecific PCR products.

In the line *why1-5*, *FSD2* showed reduction in expression level compared to the WT (Figure 7E), and also exhibited difference in AS pattern (Figure 9C). Here, only the consitutive splicing isoform was lower in the *WHIRLY1* loss-of-function mutant, while the retained intron isoforms were similar, leading to overall a reduced expression level of *FSD2* in the *why1-5* seedlings. It also means that the ratio of RI variants among all *FSD2* transcripts was higher in the *WHIRLY1* knock-out mutant. The retained intron *FSD2* sequences have an early stop codon leading to shorten the C-terminal of the protein, which contains two α-helices as part of the conserved C-terminal domain of SOD family (InterPro ID: IPR036314). Furthermore, addition of introns leads to alteration of 3'UTR sequence. The increase of RI *FSD2* isoforms in the cell may together negatively influence the overall activity of FSD2 in the plastid, providing an extra regulatory mechanism besides the suppression of *FSD2* transcription level. Altogether, these results indicate that WHIRLY1, at least in post-germination early developmental phase, plays major roles in splicing activities.
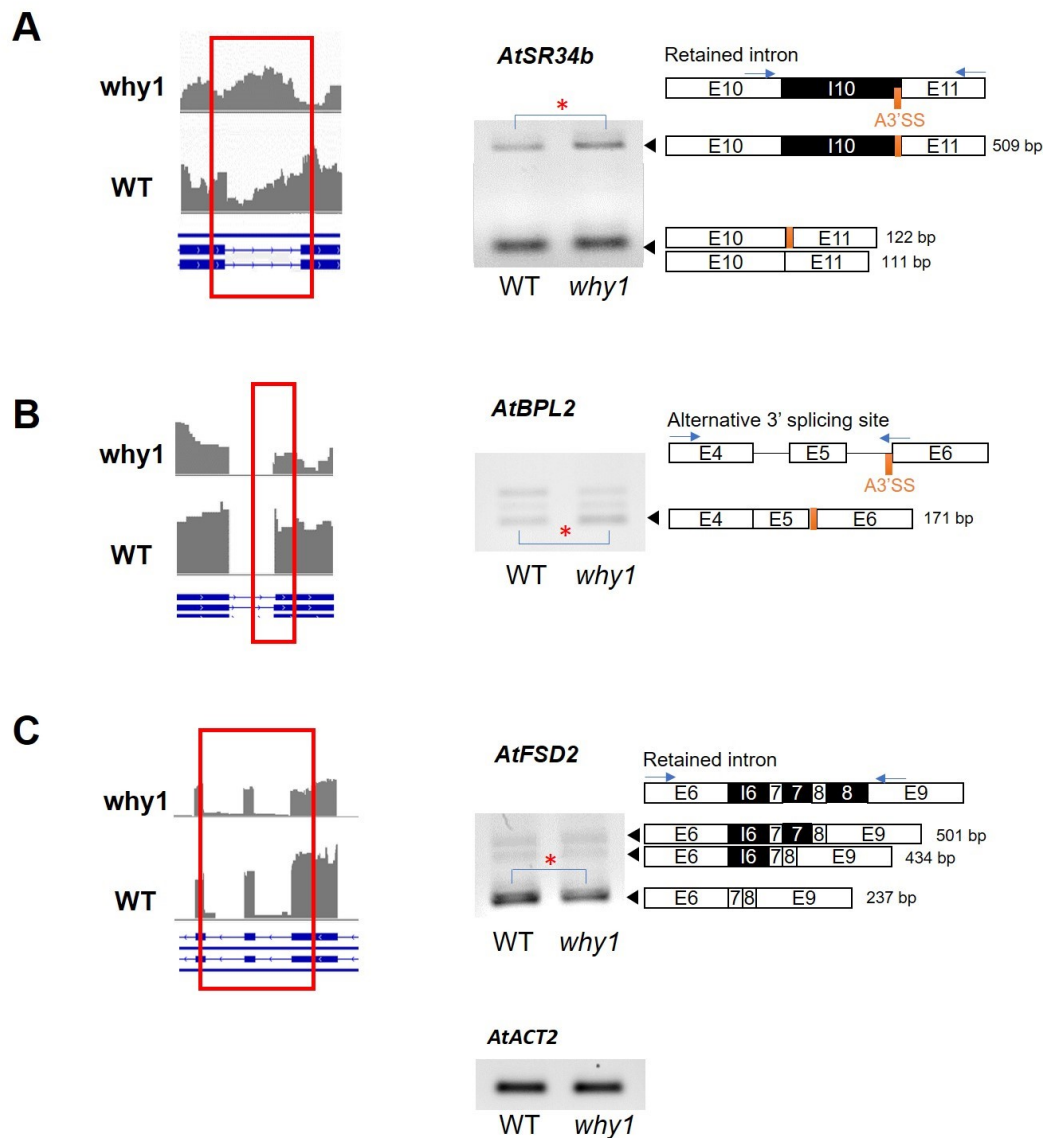
**Figure 9. RT-PCR validation of selected DASEs between the *why1-5* and the WT seedlings.** These include (**A**) Retained intron of *AtSR34b*, (**B**) Alternative 3' splicing site of *AtBPL2*, and (**C**) Retained intron of *AtFSD2*. The distribution of mapped reads on intron and exon region shown by sashimi plots on the left, and scheme of gene's exon-intron structure (E: exon, I: intron, with order numbers) and different splicing products are depicted on the right of the gel picture. Red asterisks denote bands showing differences in abundance of mRNA variants between the *why1-5* and the WT. Flanking primers are indicated by blue arrows. Alternative 3`splice site is shown in orange. *AtACT2* was used as the loading control.

### 2.1.4. Loss-of-function of *AtWHIRLY1* affected expression of glucosinolate biosynthetic genes

A striking result of RNA-seq analysis was that 10 subtaintially downregulated DEGs are involved in glucosinolates (GSLs) metabolism (Appendix Table S1). GSLs are secondary metabolites which are composed of a thioglucose, a sulfate group, and variable side chain derived from sulfur-containing amino acid (Figure 10A). Based on precursor amino acids, GSLs are categorized into two main groups, aliphatic or indole glucosinolates (aGSLs or iGSLs, respectively). As illustrated in Figure 10A, aGSL biosynthesis involves

three main stages: side-chain elongation occuring in plastid, and core-structure formation and side-chain modification, which both are located in the cytosol.
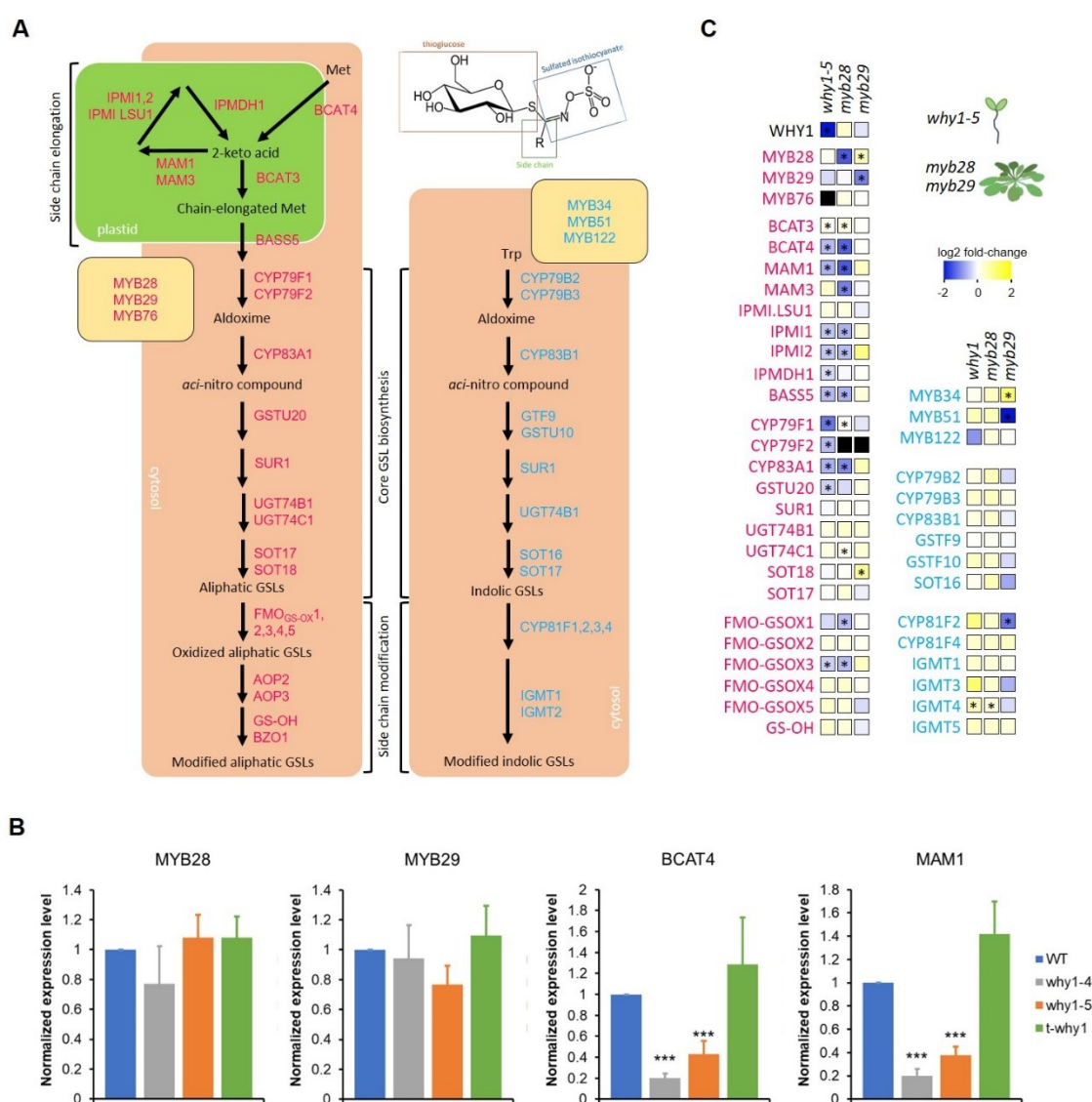


**Figure 10. Loss-of-function of WHIRLY1 affected glucosinolate biosynthetic genes**. (**A**) Scheme of a typical GSL with a thioglucose, a sulfated isothiocyanate, and a side-chain R which is derived from amino acid. Scheme of two major GSL biosynthetic pathways, aGSL and iGSL (adapted from Augustine & Bisht, 2017). (**B**) Normalized expression level of selected GSL-related genes by qRT-PCR. Bar charts depict the average of mean and error bars show standard deviation of four independent biological replicates. Asterisks indicate statistically significant level of Student's t-test between the *WHIRLY1* mutants and the WT; (*) p-value < 0.05, (**) p-value < 0.01, (***) p-value < 0.001. (**C**) Heatmap shows changes in transcript levels of GSL biosynthetic genes due to knock-out of *WHIRLY1* (measured in the present work) and due to knock-out of *MYB28* and *MYB29* (Sønderby et al., 2010). The log2FC is indicated by color bar, and asterisks (*) represent statistically significant difference as result of RNA-seq/microarray.

According to the RNA-seq data, loss-of-function of *WHIRLY1* specifically affected expression level of several genes encoding enzymes in early steps of Methionine (Met)-derived aGSL biosynthesis, such as *BRANCHED-CHAIN AMINOTRANSFERASE 4* (*BCAT4*), *METHYLTHIOALKYLMALATE SYNTHASE 1* (*MAM1*), *ISOPROPYLMALATE*

*ISOMERASE 1* (*IPMI1*), and *BILE ACID:SODIUM SYMPORTER 5* (*BASS5*). The down regulation of these genes in the *WHIRLY1* knock-out line *why1-5* was validated by qRT-PCR, showing a high correlation between RNA-seq and qRT-PCR results (Figure 10B; Appendix Figure S2). Expression level of many aGSL biosynthetic genes in the *why1* seedlings was substantially decreased to about a third of the WT level (Appendix Figure S2). Meanwhile, the expression of most genes involved in iGSL biosynthesis was not affected in the *WHIRLY1* knock-out mutant (Appendix Figure S2). Still, *SUPERROOT1* (*SUR1*), a shared gene between aGSL and iGSL pathways, was slightly reduced due to missing of WHIRLY1. Since aGSL biosynthesis is directly controlled by MYB28, MYB29, and MYB76 transcription factors while iGSL process is controlled by MYB34, MYB51, and MYB122 transcription factors (Mérillon & Ramawat, 2017), their expression level were also analyzed in the knock-out mutant *why1-5*. The qRT-PCR revealed that the expression level of all mentioned MYBs was similar between the *WHIRLY1* knock-out mutants and the WT (Figure 10B; Appendix Figure S2), with exceptions that *MYB76* and *MYB122* transcripts cannot be detected in this early stage. These findings suggest that WHIRLY1 does not act directly on the known upstream regulators of aGSL biosynthesis, which are *MYB28* and *MYB29*, but it rather regulates expression of aGSL biosynthetic genes via another mechanism.

To verify the involvement of WHIRLY1 in aGSL metabolism, the expression levels of selected genes were investigated in other *WHIRLY1* mutants, including another CRISPR/Cas9 knock-out line *why1-4* and a T-DNA insertion line *t-why1* (SALK_023713), which has been used in most of up-to-date studies about AtWHIRLY1. The other CRISPR/Cas9 knock-out line showed a similar downregulation of aGSL biosynthestic genes as the line *why1-5*, confirming the influence of WHIRLY1 on aGSL biosynthesis (Figure 10B). Interestingly, aGSL gene expression levels in *t-why1* line were not different from the WT at the seedling stage. Recently, a transcriptomics analysis of 6-hour imbiled seeds between the double knock-out mutant of the T-DNA insertion *WHIRLY1* and TILLING *WHIRLY3*, did not show any differences in expression of aGSL biosynthetic genes at imbibition stage compared to the WT (Taylor et al., 2023). These findings suggest that the deficiency of *WHIRLY1* in the *t-why1* mutant might not be the same as in the CRISPR/Cas9-based knock-out mutants.

The expression changes of other GSL-related genes in the *WHIRLY1* knock-out seedlings were compared with available RNA-array data of mature *MYB28* and *MYB29* knock-out plants, denoted lines *myb28* and *myb29*, respectively (Sønderby et al., 2010). Interestingly, a similar expression pattern of GSL genes between two lines *myb28* and *why1-5* was observed but not between *myb29* and *why1-5* (Figure 10C). Same genes related to side-chain elongation were suppressed due to missing of either WHIRLY1 or

MYB28 but not MYB29. For instance, the expression level of *BCAT4*, which encodes the main enzyme initiating the side-chain elongation process, was significantly reduced by more than 3-fold and 16-fold in the *why1-5* and *myb28*, respectively. However, *BCAT4* transcript level stayed the same in the *myb29* compared to the WT. The same trend was observed with other aGSL-related genes, such as *MAM1*, *IPMI1,* and *BASS5*. In the contrary, genes encoding enzymes in iGSL pathway were slightly but no significantly upregulated in the knock-out lines of *MYB28* and *WHIRLY1*. Whereas, there were significant increasements of selected genes, *MYB34* and *CYP81F2,* in the *MYB29* knock-out mutant.

### 2.1.5. Influence of loss-of-function of *AtWHIRLY1* to glucosinolate content

The strict and drastic suppression of genes encoding enzymes of early steps of aGSL biosynthesis, suggested a decrease in the concentration of these metabolites. Therefore, whole seedlings grown in the same conditions were collected to quantify GSLs (according to method described in Chapter 4.2.11). As the result, GSL profile of the knock-out mutant *why1-5* was in accordance with the transcriptomics data (Figure 11). Due to the missing of AtWHIRLY1, total GSL concentration in 5-day-old seedlings decreased significantly compared to the WT. The lower GSL concentration in the mutant was a result of the significant reduction of aGSL but not of iGSL (Figure 11A, Appendix Table 2.1.5).

To have a better insight about the involvement of WHIRLY1 in GSL biosynthesis, individual GSL concentrations were measured (Figure 11B, Appendix Table S3). The GSL quantification showed that at this developmental stage, 4-methylthiobutyl (4MTB) was the most abundant GSL, as around 1.81 ug/mg FW was recorded in the WT. On the other hand, 3-methylthiopropyl (3MTP), 3-methylsufinylpropyl (3MSOP), and 6-methylsulfinylhexyl (6MSOH) were presented with trace amounts (< 0.01 ug/mg FW). Additionally, the oxidized forms of aGSL (methylsulfinyl-) had lower concentration than the methylthio-precursor with the same side-chain length. Around 1-10% of *de novo* synthesized aGSLs were oxidized into the next chemical forms. Similarly, indol-3-ylmethylglucosinolate (I3M) as the first iGSL produced in the pathway presented with a higher amount compared to 4-methoxy-3-indolylmethyl (4MOI3M), a modified iGSL (Appendix Table S3).

As indicated in Figure 11B, individual aGSL concentrations were reduced in the line *why1-5* compared to the WT. Interestingly, short-chain aGSLs decreased more substantially in the *WHIRLY1* knock-out line, especially methylthio-GSLs (4MTB and 5MTP) with a strong statistical support. Indeed, WT contained 4MTB with the amount nearly double than the mutant, 1.81 ug/mg FW to 1.05 ug/mg FW, respectively. Moreover,

4MSOB and 5MTP concentration in the *why1* were nearly two third of that in the WT (Figure 11B). Meanwhile, long-chain aGSLs were reduced only mildly by less than 20% due to the missing of AtWHIRLY1 (Figure 11B). On the other hand, a small increase in iGSL content was measured in the knock-out mutant *why1-5*, but this was not significant (Appendix Table S3).
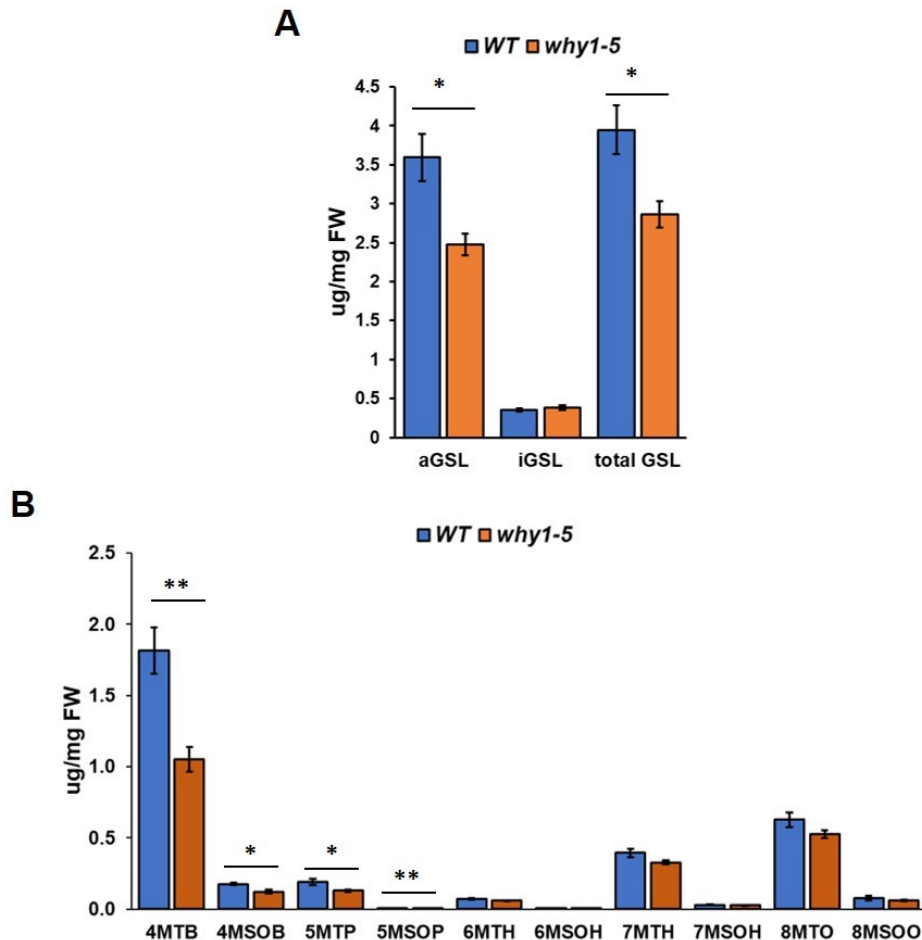


**Figure 11. Glucosinolate contents in the *WHIRLY1* knock-out and the WT seedlings**. (**A**) The concentration of aGSL, iGSL, and total GSL, and (**B**) individual aGSL compounds in the 5-day-old seedlings. Bar charts show the average of mean and error bars show standard deviation of three independent biological replicates. Asterisks indicate statistically significant level of Student's t-test between the *WHIRLY1* knock-out mutants and WT; (*) p-value < 0.05, (**) p-value < 0.01, (***) p-value < 0.001. More details about GSL content were presented in Appendix Table S3.

Taken together, our transcriptomic data showed that all genes encoding enzymes taking part in first steps of aGSL biosynthesis were significantly repressed in the *WHIRLY1* knock-out line. The reduction of these enzymes level was coincided with the reduction in aGSL quantity in the mutant *why1-5*. In contrary, iGSL genes were not influenced significantly due to loss of *WHIRLY1*'s function. It might be the reason why iGSL content was similar between two lines.

## 2.2. Function of AtWHIRLY1 in young seedlings in response to oxidative stress

In order to investigate the involvement of AtWHIRLY1 in seedlings responding to oxidative stress, an abiotic stress experiment was performed. In this experimental set-up, the oxidative stress was generated by transferring seedlings to cold and high ligh conditions after a dark adaptation as described in the Chapter 4.2.1. Under these conditions, excess light energy is absorbed but cannot be utilized for photosynthesis, causing formation of reactive oxygen species (ROS) (Kim et al., 2012). The molecular responses to the ROS burst in both the WT and the *AtWHIRLY1* knock-out mutant were assessed using transcriptomic analyses at two different time points: early response to stress (one hour under stress, 1h) and later response to stress (five hours under stress, 5h). Considering the prominent ROS involved during the stress, singlet oxygen ($^1O_2$) has a short lifespan and elicits a fast and provoke response, hence the 1-hour stress exposure could provide insights into the response to $^1O_2$. Prolonged stress exposure, such as five hours, would lead to higher production of hydrogen peroxide ($H_2O_2$). The transcriptome data obtained from this experiment will be discussed shortly and appears to reflect these expectations. Additionally, by investigated stress-related transcriptome of both lines, function of AtWHIRLY1 in oxidative stress response will be elucidated, focusing on gene regulation at transcriptional and post-transcriptional level.

### 2.2.1. Phenotype of the *WHIRLY1* knock-out mutant during oxidative stress

Five-day-old seedlings, after the first 5 hours under oxidative stress, showed no changes in term of phenotype in both the *WHIRLY1* knock-out mutant and the WT. In both genotypes, stress-related accumulation of purple anthocyanin can be seen after prolonged stress treatment (Figure 12A). In the meantime, seedlings which were continuously being grown for 4 more days at control conditions showed light green cotyledons and started having visible true leaves (Figure 12A). In term of photosynthetic capacity, the maximum PSII efficiency was reduced mildly from around 0.8 to 0.6 after 4 days under the stress. However, there was no difference between the *WHIRLY1* knock-out mutant and the WT (Figure 12B).

**A**



**B**
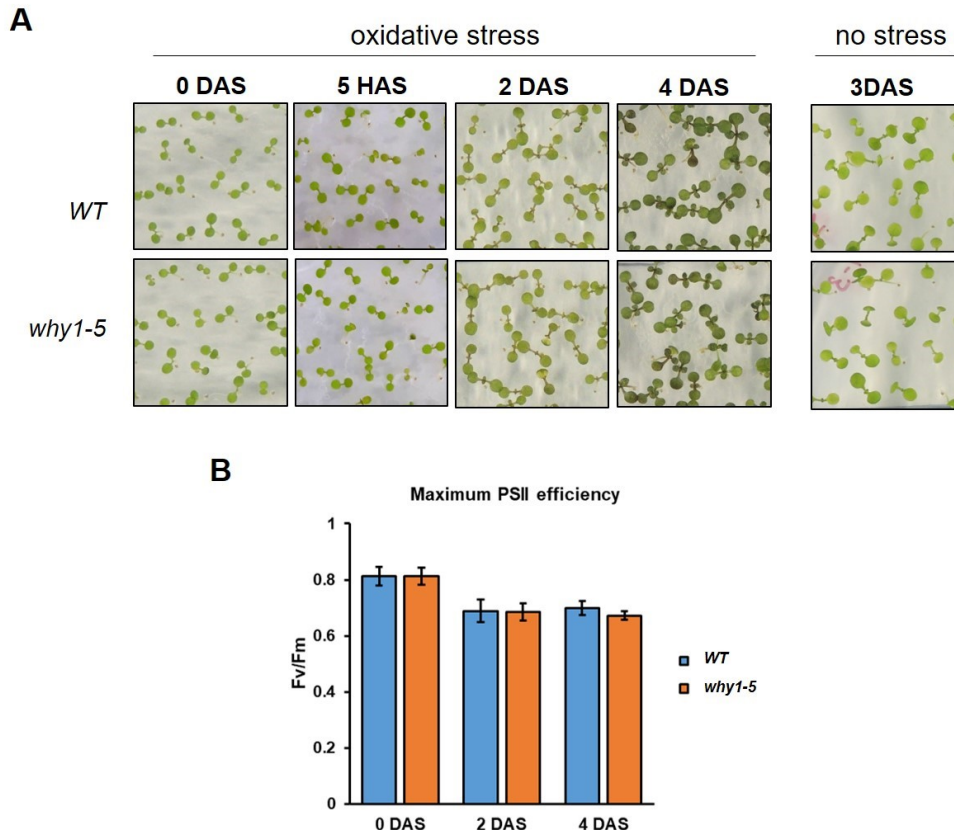


**Figure 12. Phenotype of the WT and the *WHIRLY1* knock-out seedlings during oxidative stress**. (**A**) Pictures of the WT and the *WHIRLY1* knock-out *why1-5* seedlings before stress, 5 hours, 2 days, and 4 days under oxidative stress conditions; also photos of 9-day-old seedlings grown continuously under control conditions (**B**) Maximum PSII efficiency (Fv/Fm) of two lines under oxidative stress. The bar charts show the average value of 4 biological replicate means. Error bars show standard deviation of 4 biological replicates, each contain at least 20 seedlings.

### 2.2.2. The expression pattern of *AtWHIRLIES* during oxidative stress

The expression level of the three *A. thaliana WHIRLY* genes were analyzed, showing that all significantly increased their expression level by nearly 2-fold after 5 hours under this abiotic stress (Figure 13), whereas one hour after onset of stress, transcript levels were not yet increased. The clear oxidative-stress-responsive induction of all three genes suggests a role for WHIRLIES in the response of seedlings to oxidative stress.
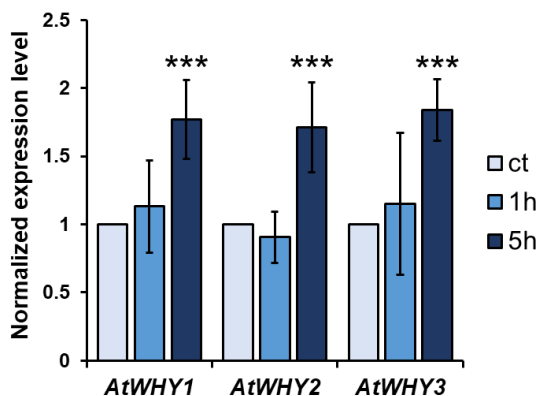


**Figure 13. Expression level of three *AtWHIRLIES* in the WT seedlings during oxidative stress.** Bar charts show the average of mean and error bars show standard deviation of four independent biological replicates. Asterisks indicate statistical significant level of Student's t-test between the stress samples and the control sample; (*) p-value < 0.05, (**) p-value < 0.01, (***) p-value < 0.001.

### 2.2.3.  Loss-of-function of *WHIRLY1* changed seedlings response to oxidative stress at transcription level

To gain further insights into the function of AtWHIRLY1 in the response of young seedlings to oxidative stress, stress-responsive transcriptomes of the *WHIRLY1* knock-out mutant, *why1-5*, and the WT were analyzed to identify potential downstream target genes of AtWHIRLY1. Besides differentially expressed genes (DEG), effects of loss-of-function of *WHIRLY1* on post-transcriptional alternative splicing events was also investigated.

#### *2.2.3.1.   Overview of oxidative-stress-responsive expressing transcriptomes*

The transcriptomes of the WT and the *why1-5* seedlings subjected to oxidative stress for 1 hour (WT_1h/why1_1h) and 5 hours (WT_5h/why1_5h) were compared to the control samples (WT_ct/why1_ct), i.e., five-day-old seedlings without stress. This comparison identified a high number of genes associated to this kind of oxidative stress in both genetic backgrounds. Specifically, 1149 and 1195 significantly DEGs were identified in the early stress sample of the WT and the *WHIRLY1* knock-out line, respectively (Figure 14A). Furthermore, prolonged oxidative stress resulted in approximately doubling of the number of DEGs. Interestingly, a slightly larger number of genes was induced than repressed in response to both early and late stage of the stress (Figure 14A). Besides, the *WHIRLY1* knock-out line had some more ROS-responsive DEGs compared to the WT (Figure 14A), suggesting that the mutant *why1-5* was more sensitive or exerted a stronger response to this unfavorable condition compared to the WT.
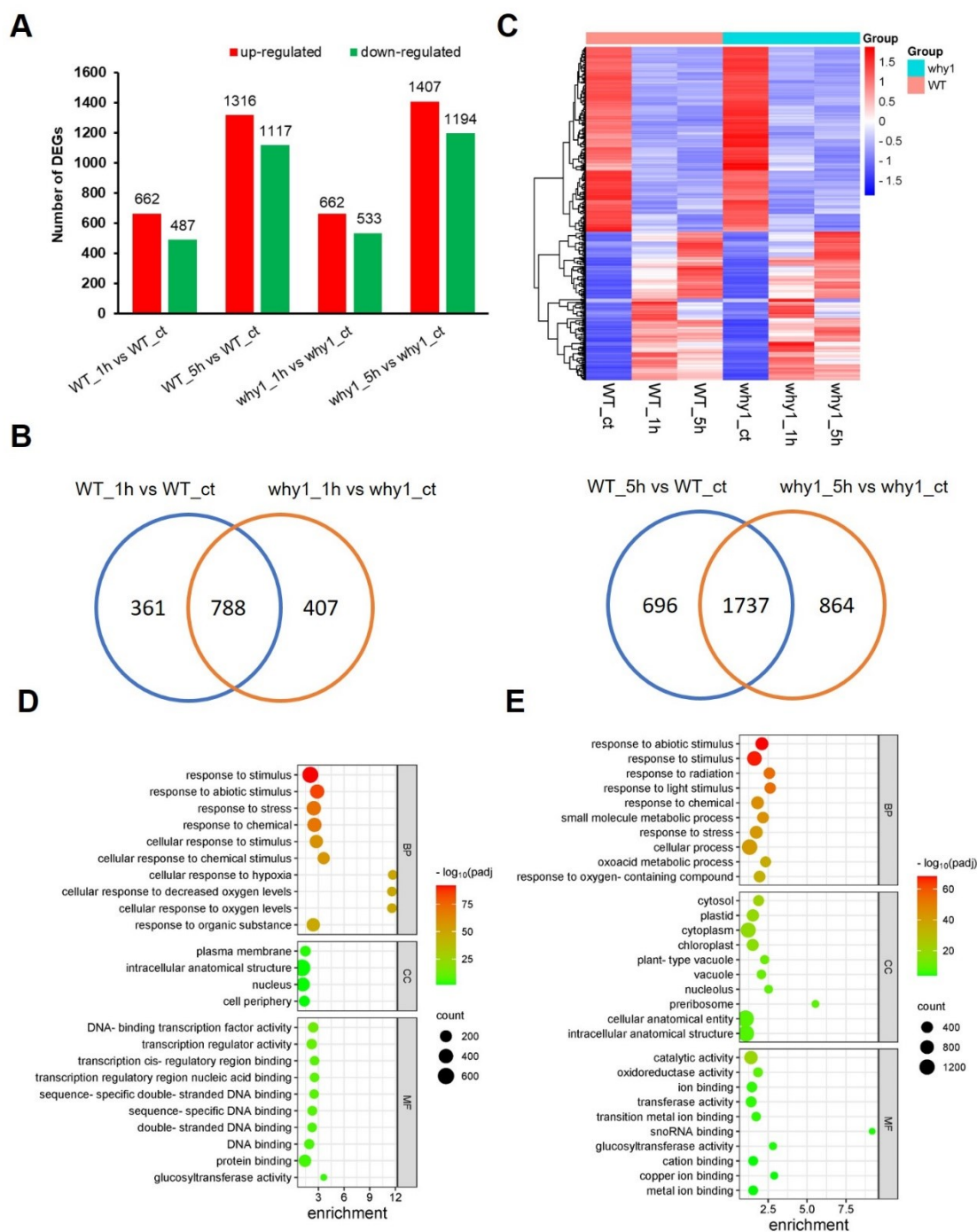
**Figure 14. Overview of transcriptomes of the *WHIRLY1* knock-out mutant and the WT seedlings under oxidative stress**. (**A**) Bar chart shows number of upregulated and downregulated genes between samples under oxidative stress and the control condition. (**B**) Venn diagram indicates overlapping stress-related DEGs between the *WHIRLY1* knock-out mutant and the WT at the early and late stages of stress. (**C**) Heatmap shows changes in transcripts level (based on fpkm values) of oxidative-stress-associated genes in both genetic backgrounds upon oxidative stress. Bubble charts show top 10 GO terms (according to p-value) enriched among oxidative-stress-associated genes after 1 hour (**D**) and 5 hours (**E**). More detail and interactive GO terms are illustrated in Appendix Figure S4, S5.

To investigate genes regulated differently due to the missing of WHIRLY1 in response to oxidative stress, ROS-responsive DEGs identified in two genetic backgrounds, *why1-5*

and WT, were compared. The comparison revealed a high number of common genes (788 after 1 hour and 1737 after 5 hours of stress), accounting for nearly two-thirds of ROS-responsive genes in each line, which are considered in the present work as oxidative-stress-associated DEGs (Figure 14B). Noticeably, a similar pattern of gene regulation upon oxidative stress was observed between the *WHIRLY1* knock-out line and the WT, as illustrated in the heatmap of around 400 oxidative-stress-associated genes in both early and late stages of the stress (Figure 14C). This is in line with the overall pattern observed in volcano plots depicting all ROS-responsive genes in each genotype (Appendix Figure S3).

Comparing ROS-responsive DEGs in the present work with an available dataset, named "the ROS Wheel" (Willems et al., 2016), where core ROS-responsive genes are categorized into different clusters, allowed a better understanding of these genes. In general, ROS-responsive DEGs identified in this work cover 17-26% of the already published core ROS wheel, in which more ROS-related genes were found at the later stage of stress treatment (Figure 15).
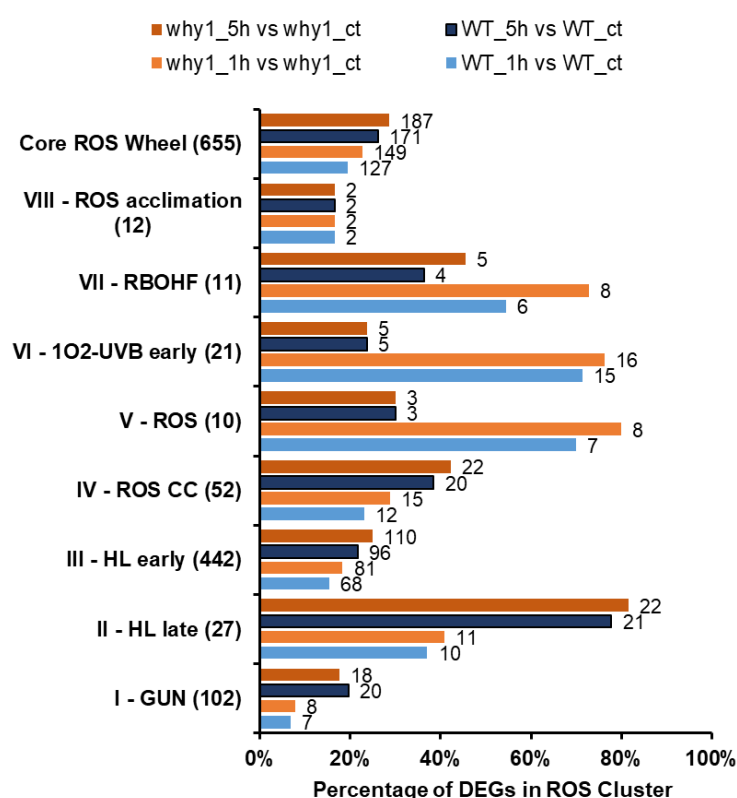


**Figure 15. Distribution of oxidative-stress-responsive DEGs in the core-ROS Wheel.** Bar chart shows the percentage of DEGs in different clusters defined in Willems et al. (2016). The number in brackets () presents number of genes in each cluster while number above the bar indicates number of actual DEGs in each comparison from the transcriptomic analysis.

In most of comparisons between stress and control samples, the *why1-5* line had more DEGs being classified into ROS Wheel than the WT, in line with the higher number of DEGs which were recorded in the *WHIRLY1* knock-out line compared to the WT (Figure 14A). Among 8 clusters in the ROS Wheel, the highest number of ROS-responsive DEGs

from this work belongs to the cluster III, i.e., early high-light (HL early). Fewer numbers of DEGs were classified into others, particularly only two genes in cluster VIII (ROS acclimation) were indentified in this experiment. Interestingly, prolonged oxidative stress led to an increased number of ROS-related DEGs in cluster I to IV, but an opposite tendency was observed in cluster V to VII (Figure 15). These later clusters include genes in all ROS-induced stresses, especially singlet oxigen signaling ($^1O_2$) (cluster VI), which is prominent at the beginning of the oxidative stress (as nearly 80% of cluster core genes were found in the present data of 1-hour-stress samples).

### 2.2.3.2.    Multiple functions of oxidative-stress-responsive differentially expressed genes

Functional analysis of oxidative-stress-associated DEGs revealed several overrepresented GO terms. A summary is shown in Figure 14D&E and a more detailed analysis can be found in the Appendix Figure S4 and S5. In general, GO terms, such as response to multiple stimuli, transcription regulation activity, and biosynthetic pathways, were highly enriched to maintain plant performance under oxidative stress (Figure 14D,E; Appendix Figure S4 & S5), which is in accordance to many other studies (Huang et al., 2019a; Willems et al., 2016; Rosenwasser et al., 2013).

Upon oxidative stress, many genes which are downstream of the *gun* retrograde signaling (Cluster I, Figure 15) were differentially expressed. Many nuclear DEGs in this category encode proteins in chloroplasts, such as components of photosystems I and II, e.g., *PHOTOSYSTEM I REACTION CENTER SUBUNIT II-2* (*PSAD2*) and *CHLOROPHYLL A-B BINDING PROTEINS* (*LHCB2.4* and *LHCB4.2*); plastid functional proteins, e.g., *FRUCTOSE-BISPHOSPHATE ALDOLASE 1* and *CHLOROPLASTIC* (*FBA1*) in carbon fixation, and *GLUTAMYL-TRNA REDUCTASE 1* (*HEMA1*) in early steps of chlorophyll biosynthesis. Clearly, the reprogramming of nuclear gene expression is directed downstream of a ROS burst in the chloroplast, implying the action of a fast retrograde signaling in this oxidative stress response. The retrograde signaling led to downregulation of photosystems and photosynthetic related genes in both WT and the knock-out mutant *why1-5*, especially under prolonged stress (Appendix Table S4).

As expected, majority of ROS-responsive DEGs were well characterized as genes responding to high light stress, both to early and late responses (Core Cluster II and III, Figure 15) (Willems et al., 2016 and references within). These genes were reported to be differentially expressed when plants were exposed to a light intensity generally 10-time higher than the previous conditions. In the present work, an 18-time increased light intensity (270 µE/m$^2$s) compared to the previous growth light (15 µE/m$^2$s) was used to induce ROS production in the cell. It was observed that a number of DEGs in this work

which are related to HL clusters increased over time. Among them, several well-known high-light marker genes were significantly induced. For example, transcription factors (TFs), e.g., *ZINC-FINGER PROTENS* (ZAT10 and ZAT12), *APETALA2/ETHYLENE RESPONSE FACTORS* (*ERF5* and *ERF6*), and *ELONGATED HYPOCOTYL 5* (*HY5*), which play major roles in light signaling and responses, elevated their expression level already after one hour. Moreover, two *EARLY LIGHT-INDUCED PROTEIN* genes, *ELIP1* and *ELIP2*, were upregulated substantially with a high fold-change in both lines (Appendix Table S4). Additionally, genes encoding several receptor kinases, which act as common regulators between abiotic and biotic stresses, were differentially expressed due to oxidative stress, e.g., *RECEPTOR KINASE LIKE 1* (*RKL1*) and *CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE 22* (*CRK22*).

Noticeably, after prolonged oxidative stress, TFs and signaling-related proteins no longer play prominent roles, but ROS-scavengers and maintainance proteins are massively produced, which is in line with the functional analysis of DEGs (Figure 14D,E; Appendix Figure S4 & S5). For instance, genes related to antioxidant biosynthesis, such as *CHALCONE SYNTHASE FAMILY PROTEIN* (*CHS*) and BETA-CAROTENE HYDROXYLASE 2 (BETA-OHASE 2), and detoxification, such as *GLUTATHIONE S-TRANSFERASE* (*GSTUs*) were regulated in response to this abiotic stress (Appendix Table S4). Besides, it is obvious that under oxidative stress, proteins are intensively misfolded, leading to upregulation of more than 20 genes encoding heat-shock or other chaperone proteins, e.g., *HEAT SHOCK 70 KDA PROTEINS* (*HSP70-2, HSP70-4*, etc.), and *CHAPERONE PROTEIN CLPB3* (Appendix Table S4). In addition, many glucosinolate (GSL)-related genes, including those being repressed in the *WHIRLY1* knock-out mutant at the control condition, belong to high-light-responsive clusters, demonstrating the role of GSLs in abiotic stress response.

Furthermore, it appears that at early stress (1 hour), preferentially $^1O_2$ signaling is induced, and during later stress (5 hours) hydrogen peroxide ($H_2O_2$) signaling is favoured. In the present work, 80% of core cluster VI, a collection of genes retrived from UV-B exposure and the *flu* mutant with short exposure (Willems et al., 2016 and references within), are found among 1-hour oxidative-stress-responsive genes, highlighting that $^1O_2$ signaling is an important factor in unfourable environmental response. Moreover, extending oxidative stress to 5 hours led to a reduced number of DEGs belong to this cluster, suggesting $^1O_2$ signaling did not play a critical role anymore. It was reported that $^1O_2$ signaling can lead to program cell death in a long exposure (Kim et al., 2012). Indeed, ROS-responsive DEGs belonged to cluster VI are highly induced in senescent tissue, such as *WRKY TRANSCRIPTION FACTORS* (*WRKY33, WRKY40,* and *WRKY53*), which

indicates a connection between $^1O_2$ signaling and programmed cell death (Appendix Table S4).

### 2.2.3.3. *WHIRLY1-mediated gene expression during oxidative stress*

To investigate genes whose expression is specifically up- or down-regulated in the *WHIRLY1* knock-out line under oxidative stress, the transcriptomes of the mutant *why1-5* and the WT at each stress condition were compared. Similar as in the control condition, a relatively small number of deregulated genes was identified when *WHIRLY1* function was lost compared to the WT at early and late stages of stress, with in total 104 and 80 genes, respectively (Figure 16A).

Interestingly, only 12 genes (8 down, 5 up) were differentially expressed in both control and stress conditions (Figure 16B), among which four genes are involved in aliphatic glucosinolate (aGSL) biosynthesis (*BCAT4, MAM1, CYP83A1,* and *CYP79F1*). As indicated in the heatmap created from RNA-seq data (Figure 16C) and also by qRT-PCR analysis (Figure 16D), the expression level of aGSL-related genes were lower in the knock-out of *WHIRLY1* compared to the WT in all conditions, though, upon oxidative stress, these genes were all induced in both lines. A two-way ANOVA test was performed and showed that the difference in expression level of aGSL genes not only was affected by different genetic backgrounds, but also by the stress (Appendix Table S5). However, these two factors did not interact. In other words, the suppression of aGSL gene expression due to knock-out of *WHIRLY1* was not intefered by stress condition factors. The same trend was observed for house-keeping-like function genes *AK3* (*AT3G02020*), *MTO1 RESPONDING FACTOR* (*MTO1*), and *lncRNA AT3G06355* (Figure 16C). In contrary, only a few non-characterized protein-encoding genes had a higher expression level in the knock-out mutant *why1-5* in all conditions compared to the WT, including *AT3G05945*, *AT1G75945*, and *AT2G11405* (Figure 16C).
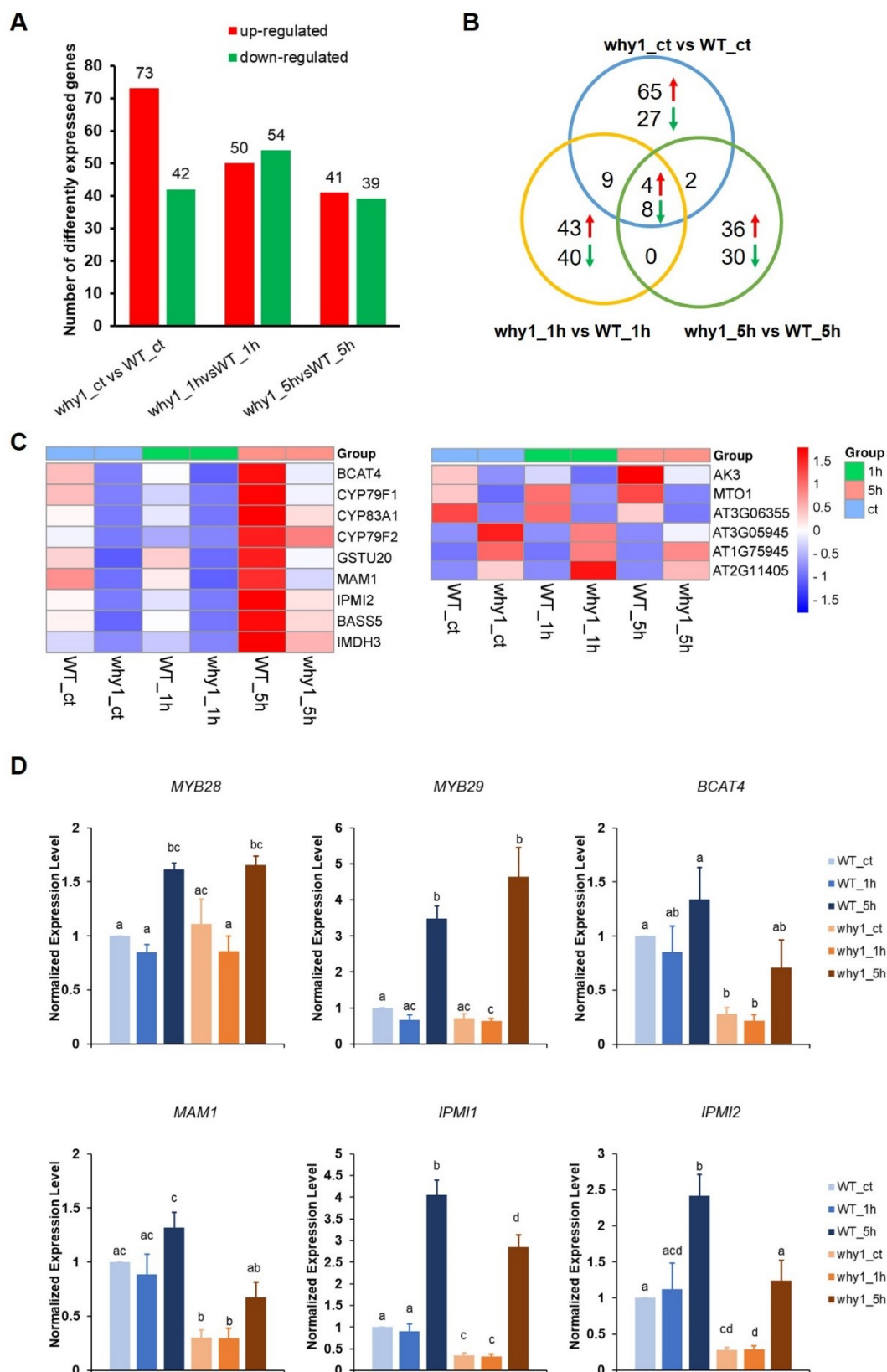
**Figure 16. Genes regulated differentially under oxidative stress between the *WHIRLY1* knock-out and the WT seedlings**. (**A**) Bar chart shows number of upregulated and downregulated genes in the *why1-5* compared to the WT under different conditions. (**B**) Venn diagram shows overlapping differentially expressed genes between two lines under different conditions. Red and green arrows represent upregulated and downregulated genes in both lines,

respectively (**C**) Heatmap indicates changes in transcripts level (based on fpkm values) in both genetic backgrounds upon oxidative stress. (**D**) Expression level of aGSL-related genes upon oxidative stress by qRT-PCR. The bar charts show average of mean and error bars show standard error of four independent biological replicates. Letters indicate significant differences between samples by post hoc test.

Besides genes, due to knock-out of *WHIRLY1* consistently deregulated regardless of condition, such as aGSL biosynthetic genes, others genes were differentially regulated in a condition-dependent manner. Some genes only deregulated under oxidative stress in the *WHIRLY1* knock-out mutant are shown in Figure 17A. For instance, some genes involved in biosynthetic processes, e.g. *SUCROSE SYNTHASE 4* (*SUS4*), *PHOSPHORYLCHOLINE CYTIDYLYLTRANSFERASE 2* (*ATCCT2*), and *IQ-DOMAIN 11* (*IQD11*), were induced by a high fold-change only in the WT at the early stage of stress, followed by a decrease in their expression level under prolonged stress, whereas the expression levels of these genes were gradually reduced overtime in the *WHIRLY1* knock-out mutant. Additionally, after 5 hours under stress condition, stress-induced genes, e.g., *CHLOROPHYLLASE* 1 (*CLH1*), *STRICTOSIDINE SYNTHASE-LIKE 7* (*SSL7*), and *CELLULOSE SYNTHASE-LIKE A01* (*CSLA01*), were strongly upregulated only in the WT but maintained their expression level in the knock-out line *why1-5*. Whereas, *ETHYLENE-RESPONSIVE TRANSCRIPTION FACTOR 54* (*ERF054*), *COLD AND ABA INDUCIBLE PROTEIN* (*KIN1*), and *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN A2 & H17* (*PCMP-A2/H17*) substaintially increased their expression level only in the knock-out line (Figure 17A).

Interestingly, several genes in response to oxidative stress seem to be earlier regulated in the *WHIRLY1* knock-out mutant than in the WT seedlings (Figure 17A). Specifically, expression levels of *SQUALENE EPOXIDASE 2* (*SQE2*), *RHO-RELATED PROTEIN FROM PLANTS 6* (*ARAC3*), and *CALCIUM-BINDING PROTEIN CML20* (*CML20*) were already reduced in the *why1-5* line one hour after the stress, while in the WT, the downregulation of these genes was observed later, i.e., after 5 hours. An opposite trend, meaning the knock-out mutant showed a higher gene expression level than the WT after 1 hour, was observed for *ELF4-LIKE 2* (*EFL2*), and *JACALIN-RELATED LECTIN* (*JAL10*), etc., followed by the upregulation in both genetic backgrounds after a prolonged stress. These results suggest that loss-of-function of *WHIRLY1* somehow affected the timing and the extent of transcription changes in response to this applied oxidative stress.
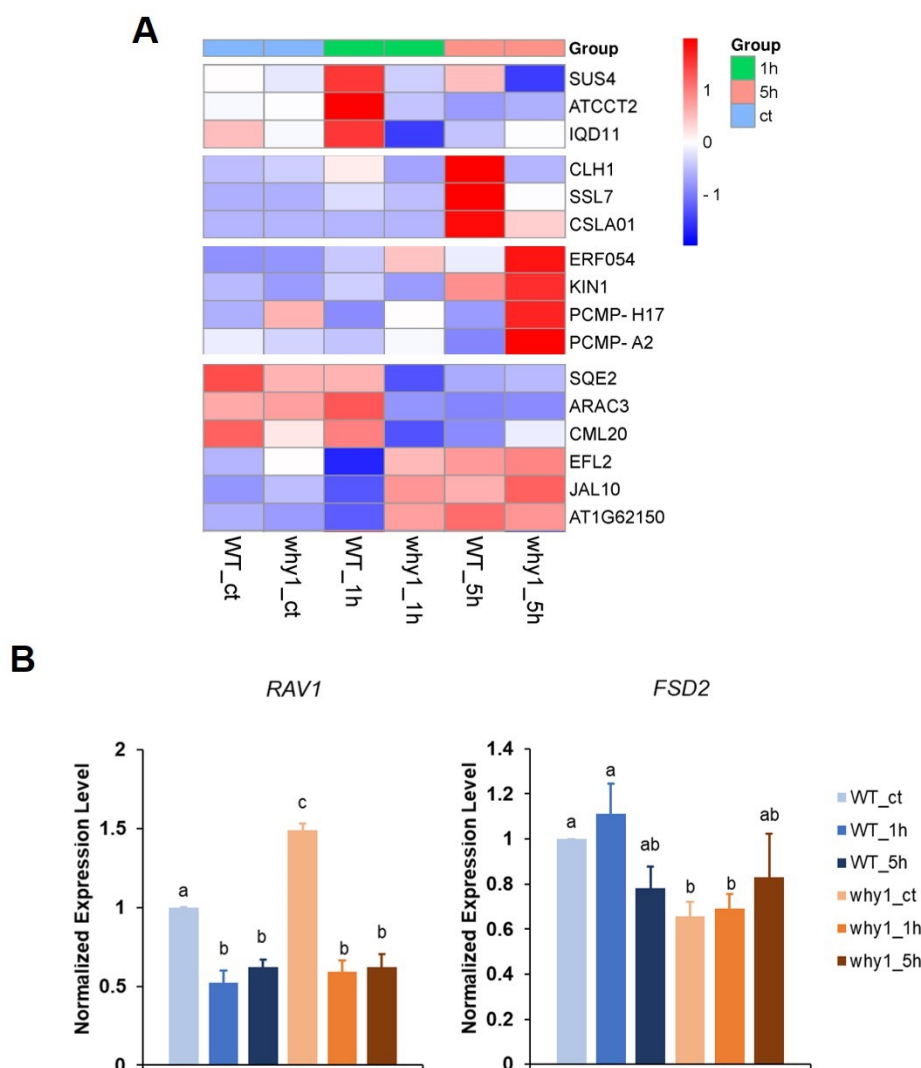
**Figure 17. Genes regulated differentially under oxidative stress between the *WHIRLY1* knock-out and the WT seedlings.** (**A**) Heatmap indicates changes in transcripts level (based on fpkm values) in both genetic backgrounds upon oxidative stress. (**B**) Expression level of *RAV1* and *FSD2* upon oxidative stress by qRT-PCR. The bar charts show average of mean and error bars show standard error of four independent biological replicates. Letters indicate significant differences between samples.

Interestingly, a number of deregulated genes in control condition was no longer differentially expressed when seedlings were exposed to stress. For instance, *RAV1*, which was up-reguated in the *why1-5* line in the control condition, lowered its expression level in both genetic backgrounds upon oxidative stress, and reached to the same level after 5 hours of stress. Meanwhile, *FSD2*, which was downregulated due to missing of WHIRLY1, was gradually adjusted to a similar level between the two lines after prolonged stress. The expression patterns of *RAV1* and *FSD2* were validated by qRT-PCR (Figure 17B). These results suggest that influence of knock-out of *WHIRLY1* on gene expression is depend on the growth conditions, in this context, stress versus non-stress conditions.

2.2.4. Loss-of-function of *WHIRLY1* changed seedling response to oxidative stress at post-transcription level

### 2.2.4.1. *Overview of oxidative-stress-associated alternative splicing*

To gain a better understanding about how seedlings respond to oxidative stress, regulation of alternative splicing (AS) of nuclear genes was investigated. The changes in the percent spliced in (PSI) between stress samples and the control sample were analyzed. As the result, a high number of significant differentially alternative splicing events (DASEs) ($|\Delta PSI| > 0.15$, p-value < 0.01) was identified in the WT and the *why1-5* seedling in response to oxidative stress (Figure 18A). Furthermore, WT appeared to have some more DASEs and differentially alternative spliced genes (DASGs) compared to the *WHIRLY1* knock-out mutant (Figure 18A). Moreover, prolonged oxidative stress somewhat led to an increased number of DASEs in the mutant while reduced AS activity in the WT seedlings (Figure 18A). Interestingly, ROS seems to exert generally a negative effect on AS, since more AS events showed reduction in inclusion of alternative sequences in both genetic backgrounds. Besides, many genes exhibited multiple splicing variants upon this stress, for example, 1514 DASEs observed after 1 hour in the WT happened in only 1115 genes.

Although all five AS types were detected, most of the oxidative-stress-responsive (ROS-responsive) DASEs were retained introns, which is consistent with previous studies (Martín et al., 2021a and references within) (Figure 18A). Besides, oxidative stress decreased the splice in of exon and mildly reduced the inclusion of alternative 5' and 3' alternative sites, whereas it exerted a nonbias effect on intron retention (Figure 18A). These findings suggest that the oxidative stress strongly influences alternative splicing activity of multi-exonic genes, in which the dynamics of AS isoforms and AS types are critical for seedling performance under the stress.

DASGs, which responded to oxidative stress in the present work, were further investigated if they were also differentially expressed genes (DEGs). Surprisingly, only a small ratio of the DASGs showed differences in their expression level in both lines (Figure 18B). Furthermore, an increased number of overlapping genes after 5 hours is likely due to a doubled number of DEGs under prolonged stress (Figure 18B). These results indicate that a majority of DASGs were only regulated by AS upon oxidative stress, highlighting that AS serves as an own regulatory mechanism to cope with stress independently from gene expression regulation.

**A**



**B**
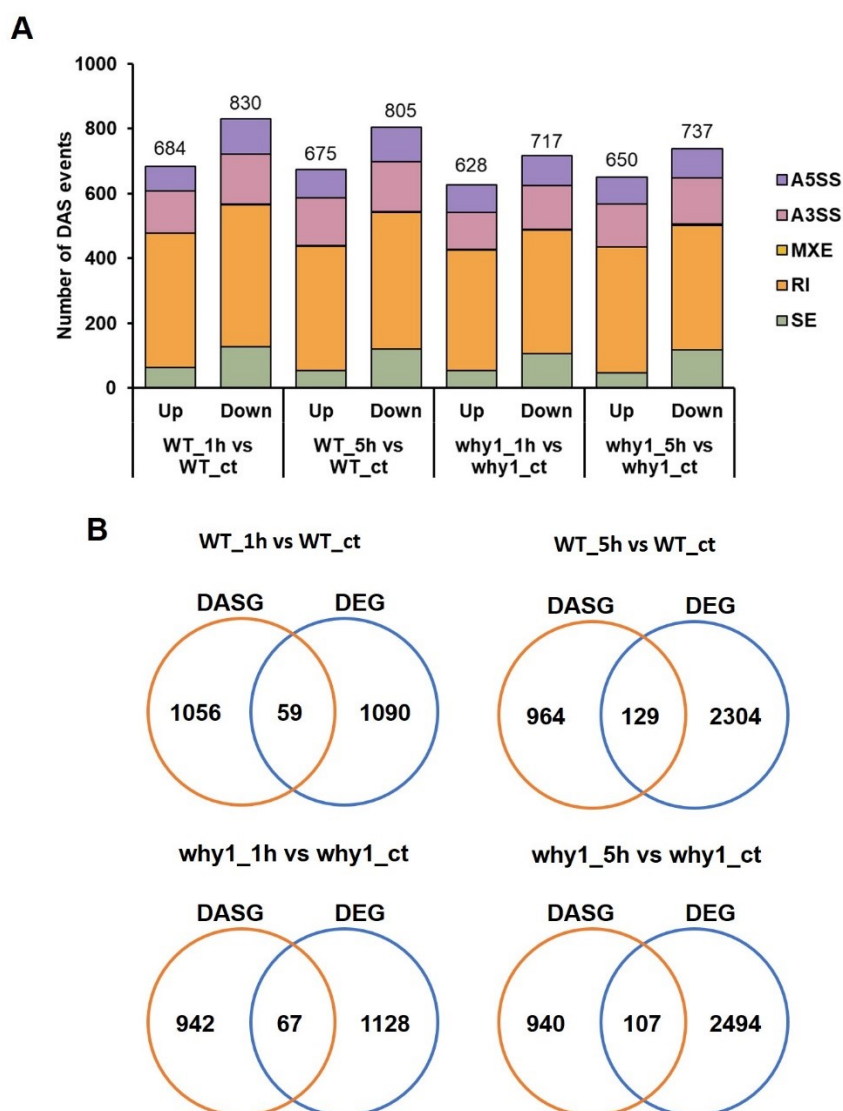


**Figure 18. Alternative splicing in response to oxidative stress**. (**A**) Bar chart shows distribution of up and down inclusion-spliced-in events due to oxidative stress. Numbers indicate the total number of AS events in each bar. (**B**) Venn diagrams depict overlap between DASGs and DEGs in response to oxidative stress.

To investigate the effect of knock-out of *WHIRLY1* on regulation of AS activity upon oxidative stress, ROS-responsive DASEs occured in the knock-out line *why1-5* were compared to those in the WT. As the result, only a small ratio of DASE was overlapping between the two genetic backgrounds, indicating a substantial influence of loss of *WHIRLY1* function in AS regulation in response to oxidative stress (Figure 19A). Indeed, 460 DASEs of 355 genes and 470 DASEs of 338 genes (Figure 19A) were observed in both lines at the early and later stress, respectively. However, unlike overlapping DEGs which exhibited a similar trend (Figure 14B), common ROS-responsive DASEs showed diverse regulation patterns between the two lines (as indicated in heatmaps in Figure 19B and in scatter plots in Appendix Figure S6). Clustering of DASEs segregated them into different coregulated modules (Figures 19B).
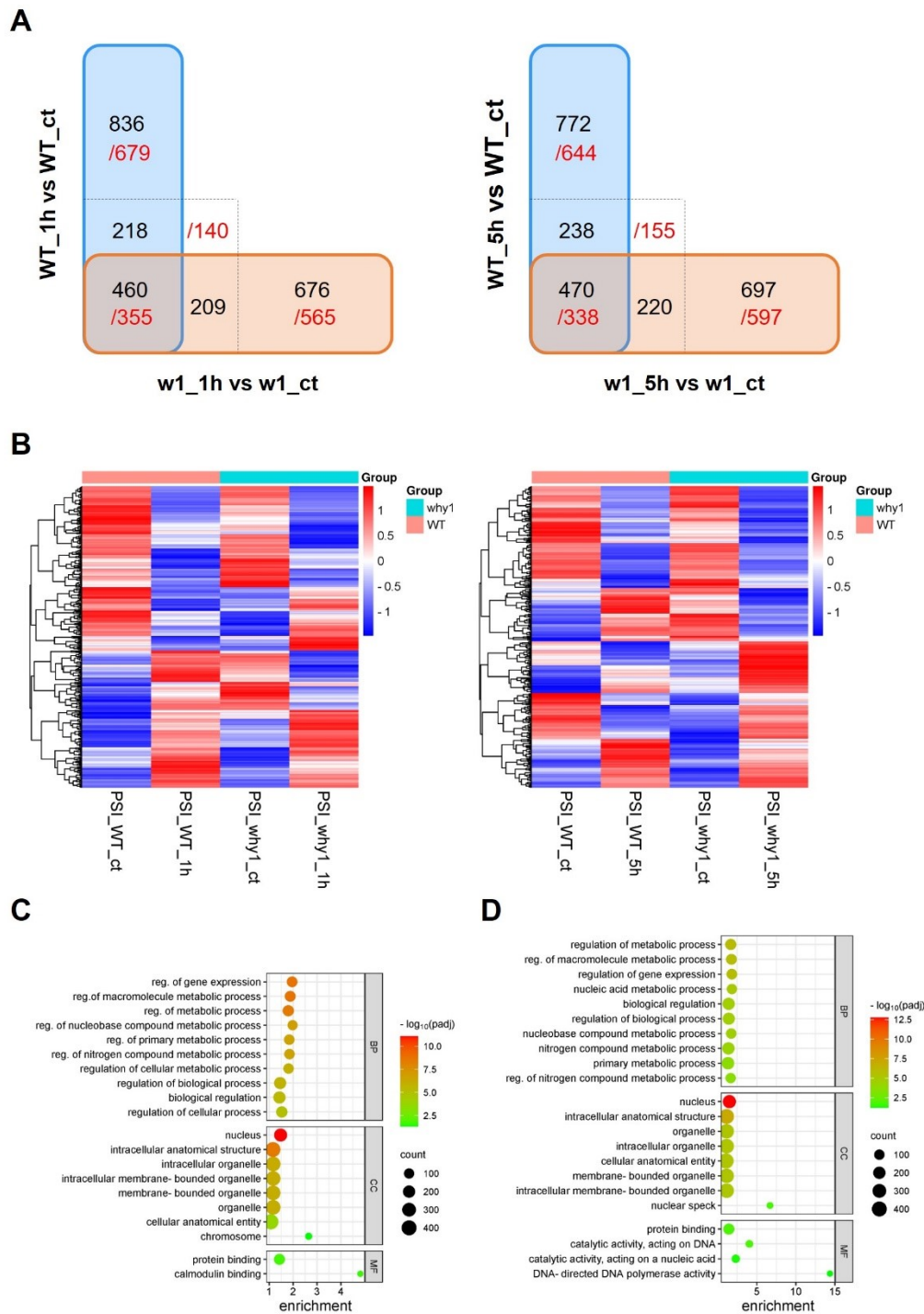
**Figure 19. Oxidative-stress-associated DAS events of the *WHIRLY1* knock-out mutant and the WT**. (**A**) Diagrams show overlapping DAS events between the *WHIRLY1* knock-out mutant and the WT after 1 hour and 5 hours under oxidative stress. Unique DASEs can occur in unique DASGs of each genotype or in overlapping DASGs in both lines (indicated by non-filled dashed-border boxes). Black numbers indicate number of DASEs, while red numbers present number of DASGs. (**B**) Heatmaps illustrate changes in percent spliced in (PSI) of DASG variants in both genetic backgrounds under oxidative stress. Bubble charts show top 10 GO terms (according to p-adj value) enriched among oxidative-stress-associated alternative splicing genes after 1 hour (**C**) and 5 hours (**D**). More details and interactive GO terms are illustrated in Appendix Figure S7 and S8.

Noticeably, a high number of unique ROS-responsive DASEs were identified in each line (Figure 19A). WT appeared to have more unique DASEs than the *WHIRLY1* knock-out seedlings, for instance, 1054 to 885 after 1 hour of stress, respectively. Interestingly, several unique DASEs in each genotype actually occurred in the same gene. For examples, after 5 hour under oxidative stress, 238 and 220 unique DASEs in the WT and in the *WHIRLY1* knock-out mutant, respectively, were observed in a total of 155 common DASGs (Figure 19A). Clearly, although sharing DASGs in response to oxidative stress, the WT and the *why1-5* mutant attuned their AS pattern differently, especially of genes generally have multiple AS isoforms. These findings suggest that the determination of which genes and their variants, and how enriched they are in ROS response, were clearly affected by loss-of-function of *WHIRLY1*.

### 2.2.4.2. *Multiple functions of oxidative-stress-associated differentially alternative spliced genes*

In the present work, all common genes between the *WHIRLY1* knock-out mutant and the WT, which exhibited changes in splicing pattern in response to oxidative stress, were considered as oxidative-stress-associated DASGs (Appendix Table S6). The functional analysis of oxidative-stress-associated DASGs revealed several enriched biological process (BP) terms, in which top GOs related to nucleic acids metabolic processes and regulation of these processes (Figure 19C-D). The overrepresented cellular compartment (CC) term, nucleus, further supports the highly enrichment of these RNA/DNA regulators (Figure 19C-D). A more detailed analysis can be found in the Appendix Figure S7 and S8.

Regardless oxidative stress exposing time, it is consistent that many DASEs occurred in genes with DNA and RNA-related functions, such as RNA processing/metabolism, DNA repair, and regulation of nucleic acid biosynthesis, which is similar as in response to other stresses (Martin et al., 2021). Specifically, RNA processing, including mRNA splicing and regulation of mRNA splicing/metabolic process were overrepresented (Appendix Figure S7 & S8). Indeed, it was reported that genes involved in mRNA processing and related functions were significantly enriched among DASEs regardless of conditions, due to the fact that these RNA processing factors generally regulated AS of themselve as a feedback loop (Kashkan et al., 2022), especially in response to stress. Several RNA splicing factors, such as *ARGININE/SERINE-RICH SPLICING FACTORS* (*SR45, SR30,* and *SR45A*), which were reported to undergo AS when plants are exposed to salt stress or cold (Calixto et al., 2018), were shown to be differentially spliced upon oxidative stress in the present work.

Besides, a high number of DASGs is involved in DNA metabolic process, especially in DNA repair after 5 hour under stress (Appendix Figure S7 & S8). DNA repair appears to be a signature functional group undergoing AS upon abiotic stress in plant (Martín et al., 2021). Under oxidative stress, the over production of ROSs can damage DNA in several compartments, which requires DNA repair mechanisms to be excecuted quickly and efficiently. Obviously, AS of several DNA repair genes can protect seedlings from harmful ROSs, such as ones have translesion-DNA-synthesis-related function (Santiago et al., 2009), e.g. *DNA POLYMERASE ETA* (*POLH*), *DNA POLYMERASE KAPPA* (*POLK*), and *DNA REPAIR ENZYME REV1*; and ones are involved in DNA recombinant and repair (Cecchini et al., 2022; Li et al., 2017b), e.g., *DNA REPAIR PROTEIN RAD51 HOMOLOG 4* (*RAD51D*) and *DNA REPAIR PROTEIN XRCC3 HOMOLOG* (*XRCC3*). All mentioned genes exibited AS in response to the oxidative stress applied in the preset work in both the *WHIRLY1* knock-out mutant and the WT seedlings (Appendix Table S6).

Noticeably, GO term regulation of gene expression was highly enriched among oxidative-stress-associated DASGs. This functional group of DASGs contains several transcription factor (TF) families, such as WRKYs, MYBs, and bHLHs and also general TFs and mediators. These genes are supposed to be controlled by both transcriptional and post-transcriptional regulation (Martín et al., 2021). In the present work, differentially spliced TFs can be considered as novel oxidative stress regulators. For example, *MYB59* has three AS isoforms and all showed reduction in response to drought in both leaves and roots (Fasani et al., 2019), in which the retained intron variant *MYB59.2* was reduced with a lower fold-change compared to constitutive one *MYB59.1*. In oxidative stress in the present work, *MYB59* also was reduced in expression level and the abundance of RI variant was increased in both lines. Other TFs, such as, MYC70, which negatively regulates cold-responsive genes (Ohta et al., 2018), and bHLH104 functioning in iron homeostasis and ABA signaling (Min et al., 2019), changed AS variant preference in response to oxidative stress, though they maintained expression level through out the experiment (Appendix Table S6).

Besides RNA and DNA processing related genes, many genes related to stress responses were differentially spliced upon oxidative stress, for instance, ABA-related genes, such as *ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR 2* (*ABF2*) and *ABRE BINDING FACTOR 4* (*ABF4*); ROS-scarvenging genes, such as *FE SUPEROXIDE DISMUTASE 2* (*FSD2*); light-responsive genes, such as *FAR1-RELATED SEQUENCE* genes (*FRS1, FRS2*, etc.). Interestingly, many genes encoding kinases enzymes, such as *MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASES* (*MAPKKK3* and *MAPKKK10*), or related to phosphorylation activity, such as *PHOSPHATIDYL INOSITOL MONOPHOSPHATE 5 KINASES* (*PIP5K7* and *PIP5K9*)

were also differentially spliced. Further investigation about biological relevance of these AS events and especially the activity or the downstream target of each isoform from these TFs/regulators could shed the light into function of alternative splicing in response to oxidative stress.

### 2.2.4.3.    WHIRLY1-mediated alternative splicing during oxidative stress

To investigate genes which might be regulated post-transcriptionally by WHIRLY1 under oxidative stress, the AS patterns of the knock-out mutant and the WT under each oxidative stress condition were compared. The comparison revealed 1421 events in 1068 genes and 1273 events in 968 genes that were differentially splicing between the two lines at the early and late stage of stress, respectively (Figure 20A). Moreover, unlike at the control condition, where clearly more decreased retained intron (RI) events were observed than increased, there was no bias in changes of intron retention activity at stress conditions (Figure 20A).

Additionally, a small number of overlapping DASEs due to knock-out of *WHIRLY1* (124 events in 112 DASGs) occurred at all three investigated conditions (Figure 20B). These genes were spliced differently consistently throughout the experiment, indicating that WHIRLY1 might modulate the AS of these genes regardless of condition (Figure 20B). Furthermore, knock-out of *WHIRLY1* exerted diverse effects on these DASEs depending on conditions, as indicated in the heatmap plotting the ∆PSI value of 124 overlapping DASEs (Figure 20C). For examples, *AT1G17520*, which functions in nucleosome assembly, *CYCLIN T1-3* encoding a protein in cyclin/CDK positive transcription elongation factor complex, and a lession-bypass *DNA POLYMERASE ZETA* (*REV3*) showed consistent trends in AS due to missing of WHIRLY1. Meanwhile, a *SERINE/ARGININE-RICH SPLICING FACTOR RS31A* (*RS31A*) had less A5SS variants in the *WHIRLY1* knock-out mutant at the control condition, whereas under prolonged oxidative stress, there were more AS variant of *RS31A* found in the mutant compared to the WT (Figure 20C).
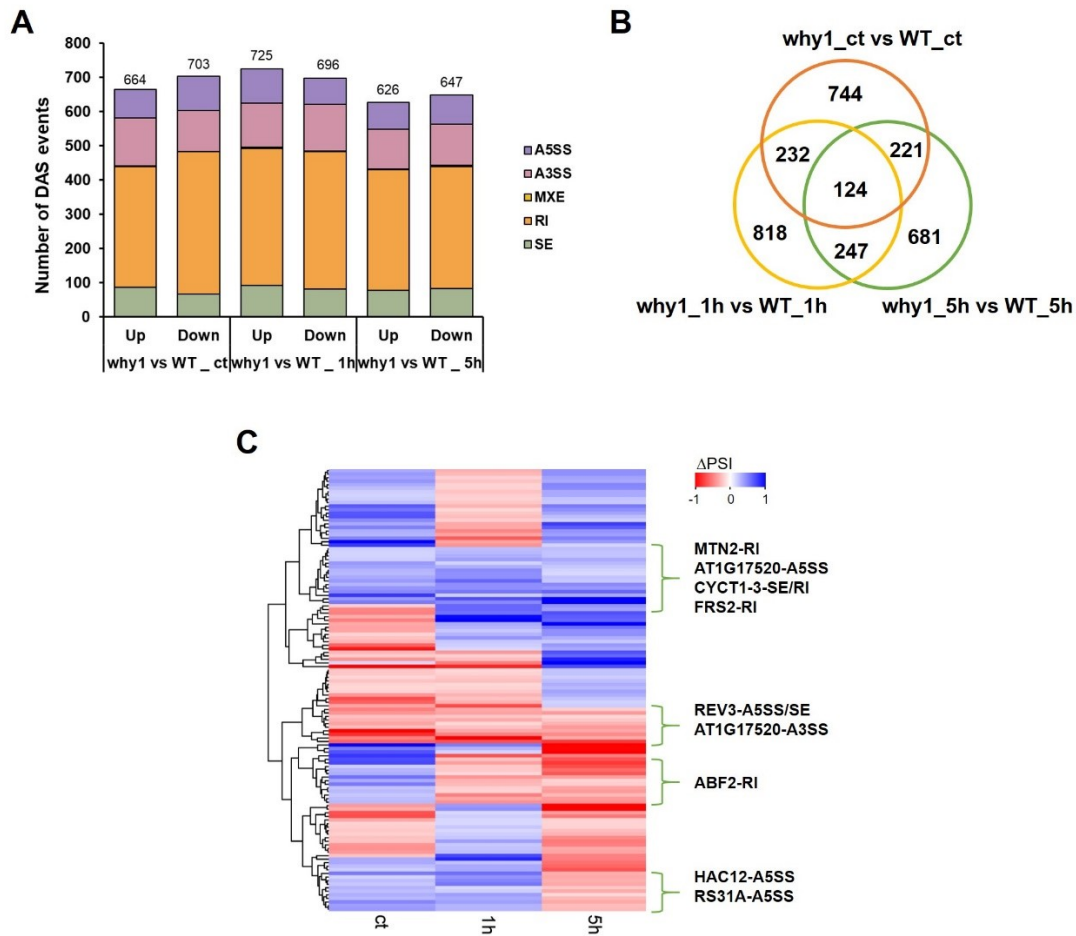
**Figure 20. Alternative splicing events regulated differentially under oxidative stress between the *WHIRLY1* knock-out mutant and the WT**. (**A**) Bar chart shows number of up- and down-inclusion alternative spliced events due to oxidative stress in the why1 line compared to the WT under different conditions. (**B**) Venn diagram presents overlapping DASEs between two lines under different conditions. (**C**) Heatmap presents overlapping DASEs due to missing of WHIRLY1 in all three conditions, plotting using ΔPSI value from 124 events. Examples of genes in some clusters are shown on the right.

Furthermore, upon oxidative stress, 818 and 681 DASEs between the *why1*-5 mutant and the WT were identified only after 1 hour and 5 hours but not in the control condition (Figure 20B), implying AS of these genes was affected directly due to missing of WHIRLY1 only under stress conditions. GO enrichment analysis of these specific DASGs showed that knock-out of *WHIRLY1* affected several processes as in the control condition, such as RNA processing, DNA repair, chromatin remodeling, gene expression, and phosphorylation (data not shown). For example, the abundance of RI variants of the two RNA splicing factors, *SR34a* and *SR45*, was higher in the *WHIRLY1* knock-out mutant only at 1 hour under oxidative stress. Whereas, missing of WHIRLY1 after the prolonged stress increased the intron and A3SS inclusion activity of *SR34, SR45A*, and *RNA-BINDING PROTEIN 25 (RBM25)* in the *why1-5* mutant compared to the WT. Meanwhile, DNA repair genes, such as *DNA MISMATCH REPAIR PROTEIN* (*MUTS*) and *UV REPAIR*

*DEFICIENT 7* (*ERCC1*) had lower inclusion level of intron due to loss-of-function of *WHIRLY1* only under stress conditions. *HEAT SHOCK TRANSCRIPTION FACTOR A2* (*HSFA2*), which was intensively DAS due to high temperature (Liu et al., 2013), decreased the exon-skipped variant in the *WHIRLY1* knock-out mutant only after 1 hour but not in the WT. Altogether, these findings suggest the significant involvement of WHIRLY1 in AS of nuclear genes under oxidative stress, which affected a wide-range of genes and diverse functional groups, implying the whole transcriptome-scale modulation by the single ssDNA binding protein, WHIRLY1.
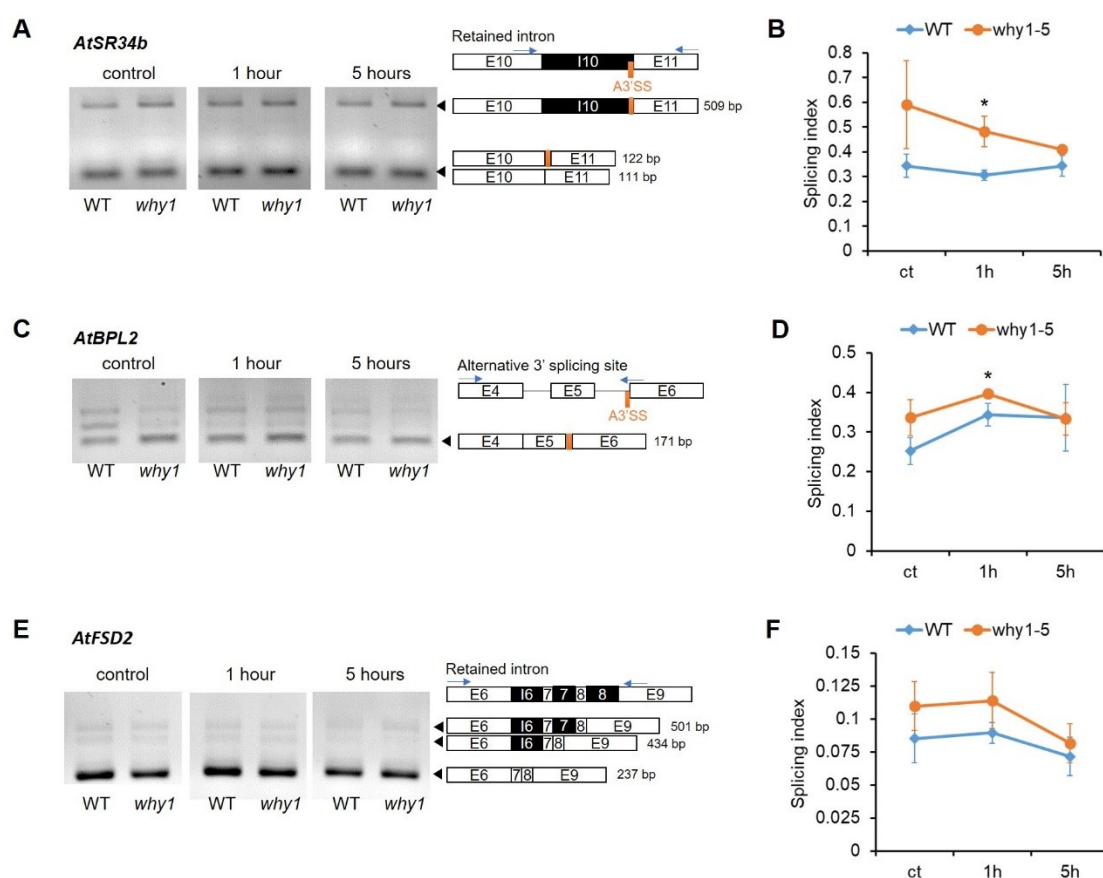


**Figure 21. RT-PCR validation of alternative splicing events under oxidative stress**, including (**A,B**) Retained intron of *AtSR34b*, (**C,D**) Alternative 3' splicing site *AtBPL2*, and (**E,F**) Retained intron of *AtFSD2*. Scheme of gene's exon-intron structure (E: exon, I: intron, with order numbers) and different splicing products are depicted on the right of the gel picture. Flanking primers are indicated by blue arrows. Alternative 3`splice site is shown in orange. Line charts illustrate splicing index, calculated as the ratio between alternative spliced mRNA(s) to constitutive spliced mRNA. Marks show the average value of three biological replicates and error bars indicate standard deviation.

DASEs observed in the control condition (Figure 9), including *SR34b*, *BPL2*, and *FSD2*, were validated to check if they are still differentially spliced between two genetic backgrounds at the stress conditions (Figure 21). Specifically, though *SR34b* had higher intron retaining level in the *why1-5* mutant compared to the WT at the control condition,

the difference between the two lines, during oxidative stress, was not substantial anymore (Figure 21A-B). It seems that the RI isoform of *SR34b* is necessary to be kept at optimum level for an efficient stress response, that the ratio between RI variant and the consitutive spliced variant in the *why1-5* mutant was reduced (Figure 21B). On the other hand, the A3SS isoform of *BPL2* substaintially increased after 1 hour before gradually decreasing in both genetic backgrounds, and eventually showed the same ratio in both lines (Figure 21C-D), suggesting a dynamics in alternative splicing activity of the *BPL2* pre-mRNA in oxidative stress response. Meanwhile, *FSD2* was altered spliced due to missing of *WHIRLY1* at the control condition and also under the oxidative stress. $H_2O_2$ accumulated in plastid requires an efficient detoxification mechanism, which likely needs a sufficient level of functional FSD2, which is encoded by the constitutive spliced variant. Clearly, the non-functional isoform of FSD2 needs to be replaced by a full-length functional peptide, thereby retained intron isoforms showed a reduced ratio among transcripts (Figure 21E-F). These results show that the AS-related function of WHIRLY1 on many genes can be changed due to the stress, suggesting another factor(s) might intefere or complement the loss-of-function of *WHIRLY1*.

Altogether, these results indicated that WHIRLY1 is involved profoundly in alternative splicing of nuclear genes upon oxidative stress. Because the *WHIRLY1* knock-out mutant already exhibited the AS pattern as similar as when seedlings are under abiotic stress, the DAS events and DASGs recorded in the mutant upon the oxidative stress were significantly different from that of the WT seedlings, especially in term of differences in preferable DAS variants. Besides many consistent DAS events due to the loss-of-function of *WHIRLY1* throughout the experiment, many DAS events appear to exhibit different patterns between the stress and non-stress conditions, suggesting the dynamics of WHIRLY1 and other AS regulators in the process.

# 3. DISCUSSION

WHIRLY1 has been described as a multifaceted transcription factor that modulates and integrates various plant development and stress responses (Krupinska et al., 2022; Taylor et al., 2022). AtWHIRLY1 function is well documented at later stages of development. In this study, I investigated the function of AtWHIRLY1 at an early post-germination stage, i.e., in 5-day-old seedlings. Young seedlings are very sensitive to environmental cues during this important phase of development, and dramatic changes in their habitat could end plant's life at this early stage. Therefore, the growing seedling has to be able to quickly react to deterioration of surroundings to survive. On the other hand, responses to these stressors have to be balanced with growth processes to ensure ongoing growth to a mature robust plant. This cross-talk between abiotic and biotic stress responses and developmental processes has to be orchestered by molecular nodes, acting in response to environmental cues on different pathways, and WHIRLY1 is one possible candidate.

## 3.1. Involvement of AtWHIRLY1 at the early post-germination developmental stage

### 3.1.1. In early post-germination stage, WHIRLY1 has no impact on morphological parameters and photosynthetic capacity

In the present study, the function of WHIRLY1 in this early stage was investigated by a reverse genetic approach, employing two CRISPR/Cas9 knock-out lines. Loss-of-function of *WHIRLY1* did not result in an obvious phenotype during this early post-germination phase, with similar hypocotyl length, fresh weight, maximum PSII efficiency, and pigment content in all lines (Figure 6), suggesting that knock-out of *WHIRLY1* does not cause severe effects to young seedlings at macro-morphological level. Similar results with a lack of an obvious phenotype due to loss of *WHIRLY1* function have already been described at different developmental stages of *Arabidopsis* (Maréchal et al., 2009) and tomato (Zhuang et al., 2019, 2020). On the other hand, WHIRLY-deficient monocots usually appear a severe phenotype. Barley *HvWHIRLY1* knock-down mutant showed delayed formation of green tissue as well as photosynthetic malfunction (Krupinska et al., 2019), while missing of WHIRLY1 in maize or in rice caused abnormal embryos and albino plants (Qiu et al., 2022; Prikryl et al., 2008). However, lacking of an apperant phenotype cannot rule out that WHIRLY1 plays roles in this critical stage. In the present work, it was shown from transcriptomic analysis that loss of *WHIRLY1* function affected nuclear gene regulation with a striking impact on the expression pattern of

aliphatic glucosinolate (GSL) biosynthetic genes. And though GSLs have significant role in plant defense and growth, it was shown before that aGSL-deficiency *myb*-mutants also did not show any obvious phenotype compared to the WT (Sønderby et al., 2007, 2010).

As shown in the present work, WHIRLY1 seems not to be necessary for seedling formation, at least under the standardized conditions used in this experiment. One explaination could be that its function can be complemented, preferably by the other two AtWHIRLIES, AtWHIRLY2 and AtWHIRLY3. It was observed before that the expression level of *WHIRLY3* was enhanced at shoots of *WHIRLY2* knock-out line at 2- to 4-true-leaf stage, while a similar expression of *WHIRLY3* as in the WT was recorded in roots (Golin et al., 2020). In this paper, it is discussed that WHIRLY3 was imported into mitochondria in the *ko-why2* to take over WHIRLY2 function. Therefore, in a similar case, AtWHIRLY2 or AtWHIRLY3 might compensate AtWHIRLY1 function in the *WHIRLY1* knock-out mutant in this study. Indeed during germination, all three *WHIRLY* genes in *Arabidopsis* increase their expression level and the most dramatical change was observed with *AtWHIRLY1* (Appendix Figure S9). In fact, AtWHIRLY1 and AtWHIRLY3 share 77% similarity in protein sequence (Desveaux et al., 2004), both naturally are targeted to chloroplasts (Golin et al., 2020; Krause et al., 2005) and nucleus (Xiong et al., 2009), making AtWHIRLY3 a promising replacement for WHIRLY1. Additionally, WHIRLY2 also could act as a potential functional homologue of WHIRLY1 as it contains the conserved Whirly domain and it could be found in all three compartments under specific conditions (Huang et al., 2020). Besides, it was shown that the presence of 10% functional HvWHIRLY1 in the knock-down RNAi W1-1 mutant was enough to maintain plastid wildtype-like nucleoids (Krupinska et al., 2014b). Thus, it could be hypothesized that part of already-made AtWHIRLY2/3 in plant cells could take over WHIRLY1 function in this case. This has to be clarified in future experiments.

Nevertheless, it is worth to note that besides the lack of an apparent phenotype under control conditions, loss-of-function of *WHIRLY1* obviously affects expression of a very specific set of genes (as in detail described in next chapters). It means that other WHIRLY proteins could compensate several WHIRLY1 functions due to their homologies in protein structure and localization, but AtWHIRLY1 still exerts unique functions which cannot be compensated by other WHIRLIES. In this work, we investigate these unique functions of WHIRLY1 by comparing transcriptomes of the knock-out line *why1-5* with the WT.

### 3.1.2. AtWHIRLY1 regulates expression of specific nuclear genes involved in multiple regulatory pathways in young seedlings

Besides the lack of an obvious phenotype, knock-out of *WHIRLY1* caused distinct alterations in gene expression at control condition, indicating specific functions of

WHIRLY1 in controlling expression of nuclear genes at seedlings establishment stage. Though a small number of DEGs were recorded, loss-of-function of *WHIRLY1* particularly affected genes involved in various developmental processes, as well as abiotic and biotic stress responses. Interestingly, no previous target genes, e.g., *AtPR1* or *AtWRKY53*, was differentially expressed in the *why1-5* seedlings. On one hand, the expression pattern of these targets is mostly in senescence tissue or in response to pathogen attack or ABA accumulation (Miao et al., 2013; Desveaux et al., 2002). Hence, without any pertubations, the *WHIRLY1* knock-out mutant did not show any differeces in *AtPR1* and *AtWRKY53* expression level compared to the WT.

Nevertheless, DEGs identified at the control condition from the present work propose new target genes of WHIRLY1 at this early stage, and shed new light into WHIRLY1's function. Loss-of-function of *WHIRLY1* in post-germination stage affected especially genes involved in three functional classes, abiotic stress response, biotic stress response, and a high number in regulation of (early) plant development. Genes downstream of WHIRLY1 are involved in intrinsic processes, such as transporters of various compounds, different metabolite biosynthesis, and tissue growth and development. Moreover, genes related to DNA and RNA metabolism also were deregulated due to knock-out of *WHIRLY1* in this early stage. Besides, target genes of WHIRLY1 also take part in responses to harmful changes in abiotic environment as well as from pathogens and herbivores attack. One of the most striking effects of *WHIRLY1* knock-out was a strict suppression of aGSL-related genes, which play roles in plant defense against pathogens (Figure 10). This result and a possible function of WHIRLY1 in GSL metabolism will be discussed in a separate chapter (Chapter 3.3).

Quite often, the genes, which affected by loss of *WHIRLY1* function, are connected not only to one of these regulatory pathways, but rather are involved in both abiotic and biotic stress responses and development (Figure 7D-E; Appendix Table S1). One example is *RAV1*, which is a multi-functional gene that is involved in various hormone signaling and development, as well as in plant defenses (Chandan et al., 2023; Mandal et al., 2023; Feng et al., 2014; Fu et al., 2014; Woo et al., 2010). Another example for such a multi-functional gene affected by knock-out of *WHIRLY1* is *FSD2*, which encodes a dismutase that has a house-keeping-like function and critical in the redox homeostasis in chloroplasts. Further experiments should be done to study the biological function of these connections.

### 3.1.3. AtWHIRLY1 regulates RNA alternative splicing of nuclear genes in young seedlings

Function of WHIRLIES related to plastidial RNA alternative splicing (AS) was reported before in monocots, including barley (Melonek et al., 2010), maize (Prikryl et al., 2008), and rice (Qiu et al., 2022). In these plants, WHIRLY1 interacts with plastidial splicing factors and mediates splicing process of RNA containing group II introns. As WHIRLY1 is dually targeted in both plastid and nucleus (Krause et al., 2005), it is reasonable that nucleus-located WHIRLY1 could also have a similar function in RNA splicing. In the present work, it is observed for the first time the involvement of AtWHIRLY1 in nuclear gene splicing. Loss-of-function of *AtWHIRLY1* resulted in a profound change in AS pattern at the seedling stage, as more than a thousand genes were differentially spliced compared to the WT (Figure 8A). It has been increasingly evident that AS plays significant roles in seedling establishment (Kathare & Huq, 2021; Narsai et al., 2017; Shikata et al., 2014). These changes in AS pattern due to knock-out of *WHIRLY1* suggest a novel role of WHIRLY1 in *Arabidopsis* young seedling establishment on another layer of regulation, i.e., post-transcription level.

Many DASGs in the *WHIRLY1* knock-out line are themselves involved in the RNA-processing/-splicing process, including serine/arginine-rich (SR) proteins and other splicing regulators, which are essential for AS in development and stress responses (Appendix Table S2). Indeed, *SR34b* and *BPL2* were shown to be differentially spliced due to knock-out of *WHIRLY1* in the seedling stage (Figure 9). Cruz et al. (2014) examined 20 genes in the SR family under ABA treatment and found that the plant-specific SC35-like (SCL) subfamily was distinctively responsive to ABA and some *SR* genes were affected due to the disruption of ABA pathway. As ABA is a key hormone in embryo maturation and germination, and also seedling establishment (Yadukrishnan & Datta, 2021; Wang et al., 2011), these RNA-splicing factors (SR34b, SCL33, SR45a, etc.) are promising post-transcriptional regulators of other ABA-targeted genes during early development and also stress responses. WHIRLY1 was reported to be involved in ABA-mediated processes, such as germination in *A. thaliana* and leaf senescence in barley (Manh et al., 2023; Karpinska et al., 2022; Kucharewicz et al., 2017). Therefore, it is possible that WHIRLY1 acts as a mediator between ABA signaling and the alternative splicing of *SR* genes. Furthermore, WHIRLY1 could together with these AS regulators bind to the same pre-mRNAs and modulate their splicing, including the RNA splicing factors themselves.

Not only genes involved in RNA metabolism but also genes with house-keeping-like functions were differentially spliced in the *WHIRLY1* knock-out seedlings. One of example shown in the present work is *FSD2* which was deregulated by loss of *WHIRLY1* function

at both transcriptional and post-transcriptional level. *FSD2* expression level was significantly reduced in the *WHIRLY1* knock-out mutant (Figure 7E). It was the consequence of the reduction in the constitutive splicing isoform, and in the meantime the increase in RI isoforms, which are unstable and might not function (Figure 9). Nevertheless, these findings need to be verified by further genetic and molecular studies, and it will be necessary to identify the biological relevance of these WHIRLY1-mediated post-transcriptional regulations.

### 3.1.4. Alternative splicing in the *WHIRLY1* knock-out seedlings mimics an abiotic stress condition

GO enrichment analysis of differentially alternative spliced genes (DASGs) between the *WHIRLY1* knock-out and the WT at the control condition (showing effect of loss of *WHIRLY1* function) showed that many of these genes have DNA and RNA-related functions (Figure 8C; Appendix Table S2). Strikingly, similar functional groups were found to be enriched among DASGs between oxidative-stress-exposed and the control-grown WT samples (showing effect of oxidative stress). For instance, in both cases, genes involved in regulation of post-transcriptional events, especially mRNA splicing via spliceosome, were enriched. Interestingly, "DNA repair" GO term was enriched with 27 genes differentially spliced in the *WHIRLY1* knock-out line (Figure 8C). These DNA repair genes are quite specific in response to various abiotic stresses, such as drought and salt (Martín et al., 2021 and references within), or in the present work, oxidative stress, which will be discussed later in Chapter 3.2.2.

A comparison between these two DASG sets was performed, and suprisingly, more than half of DASGs due to the knock-out of *WHIRLY1* in control condition are also found in ROS-responsive DASGs in the WT at both the early and late stages of oxidative stress (Figure 22). Interestingly, loss-of-function of *WHIRLY1* in the control condition caused changes in the AS pattern in a comparable manner as the oxidative stress affecting the WT seedlings (Figure 22A,B).
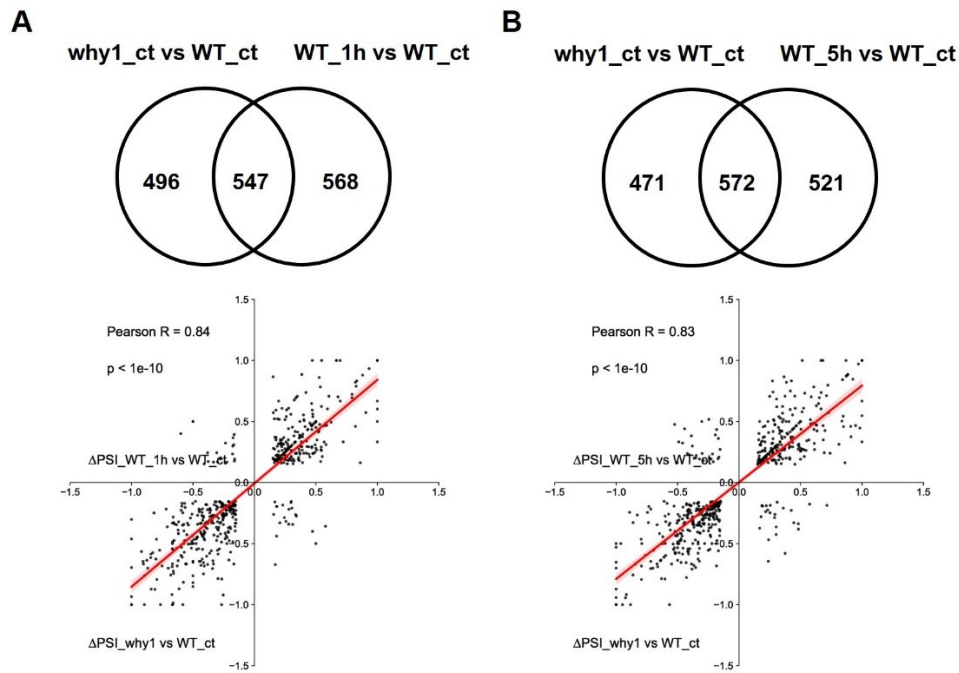
**Figure 22. Relation between WHIRLY1-mediated AS in control condition and oxidative-stress-associated AS events in the WT.** Venn diagrams show overlapping DASGs while scatterplots showing correlation between ΔPSI values of all overlapping DASEs in which one is due to loss of *WHIRLY1* function at the control condition and one is due to oxidative stress in WT seedlings at early (**A**) and late (**B**) stages. The Pearson correlation was calculated automatically. p-values correspond to binomial tests for quadrants I and III vs the total.

It was reported that knock-out mutants of some core spliceosome components also led to an altered AS pattern, mimicking the AS pattern found under abiotic stresses (Martín et al., 2021 and references within). Loss-of-function of *GEMIN2* and *LSM8*, which are critical for spliceosomal assembly, showed an increase in exon- and intron-retained sequences, it is similar as core abiotic stress DASEs which were identified in the study of Martín and colleagues (2021). Meanwhile, *luc7* and *rbm25* mutants, which affected the 5' splice site regconition, only showed AS correlation in exon skipping (ES) pattern with AS pattern in abiotic stresses. The specificity of these mutants on ES pattern is likely due to they misread exon definition. These observations also imply that changes in different AS types, might be related to the specific function of each RNA processing/splicing factors. In the *WHIRLY1* knock-out mutant, there was no bias on any AS types that reproduced the changes by oxidative stress, suggesting WHIRLY1, if functioning in AS process, might not participate in any specific step, such as recognition of splicing cites but might act rather more generally, such as in spliceosomal assembly and stability. Nevertheless, another more indirect explaination of these results cannot totally be excluded, saying that *WHIRLY1* knock-out seedlings might just misread environmental cues, and thereby sense and reprogram the transcriptome as under a abiotic stress condition, at least at post-transcriptional level.

### 3.2.     Response of seedlings toward oxidative stress includes different layers of gene regulation.

#### 3.2.1.  Oxidative stress causes massive reprogramming of gene expression

Reactive oxygen species (ROS) are constantly produced in plants through metabolism as by-products of incomplete oxygen reductions. Excessive levels of ROS, or oxidative stress, due to many reasons (high light stress, malfunction of ROS-scavenger enzymes, blockage of electron chain chemically, etc.) cause cellular stress and damage and consequently lead to cell death. To maintain ROS homeostasis, firstly different downstream signaling pathways are triggered, leading to reprogramming of the transcriptome to detoxificate ROS and repair ROS-caused cell damages. The oxidative-stress-reponsive transcriptomes in the present work investigated the early (1h) and late (5h) responses to a ROS burst in chloroplasts. By exposing seedlings to a condition with a substantially increasing light intensity and low temperature, the absorbed excess energy, which cannot all be used in photosynthesis, resulted in an immense production of ROSs, including singlet oxygen ($^1O_2$) and hydrogen peroxide ($H_2O_2$) (Figure 4).

The transcriptome analysis in the present work reflected the responses of seedlings toward oxidative stress, which begin with signal perception and transduction followed by the reprogramming of many transcription factors. These actions consequently lead to genome-wide changes in gene expression and finally in metabolic and functional cellular restructuring. The applied oxidative stress caused differential gene expression of more than a thousand genes at the early stage and the number of DEGs was double at the later stage (Figure 14). GO analysis allowed to functional classification of these ROS-responsive DEGs, which is consistent to previous studies (Willems et al., 2016 and referrences within). The distribution of DEGs to the core-ROS Wheel also confirmed the molecular response of seedlings toward oxidative stress (Figure 15). Besides, many genes involved in not only abiotic responses, such as light, UV, and cold, but also in biotic responses, such as bacterium and fungus, were altered in their expression patterns. These results suggest a crosstalk between different regulatory networks of abiotic and biotic stress responses that were co-activated upon oxidative stress.

Interestingly, a transition in transcriptional program was observed in seedling responding to oxidative stress from short (1 hour) to prolonged (5 hours) exposure. At 1 hour, transcription factors (TF) play major roles for a quick reprogramming of the transcriptome, involving many well-known TF families such as WRKYs, ZATs, and ERFs. Besides, many ROS-responsive DEGs at the early stage have function in kinase/ phosphorylation activity, in line with the activation of multiple hormone signaling pathways,

highlighting the importance of stimuli perception and signaling at the beginning of stress response (Appendix Figure S4 & S5). On the other hand, after a long exposure to stress, the involvement of TFs and signaling-related proteins becomes less prominent, instead, DEGs are related to translation activity and repair/detoxification mechanisms are highly enriched. Furthermore, the biosynthesis of various compounds was also adapted to the high production of ROS in the cell. Apparently, genes encoding enzymes involved in metabolisms supporting cell in defense, such as glucosinolate, flavonoid, glutathione, and in particular cell wall components, were highly produced after a prolonged stress. Meanwhile, biosynthesis of chlorophyll was suppressed right after 1 hour of oxidative stress treatment. Altogether, these changes in metabolism indicate a shift in seedling metabolism from growth to defense, then to repair of the photosynthetic machinery and other cellular damages.

### 3.2.2. Oxidative stress causes significant changes in the alternative splicing of nuclear genes

Recently, alternative splicing (AS) in development and stress responses has been received increasing attention as it, besides the well-known transcriptional regulation, provides another regulatory layer, and thereby contributes to a more complex and flexible molecular landscape. Many studies have confirmed that abiotic stresses remarkably change AS in plants (Liu et al., 2022; Martín et al., 2021; Laloum et al., 2018). The kind of oxidative stress in the present work, also caused a global change in AS pattern in both the *WHIRLY1* knock-out mutant and the WT seedlings, with a comparable number of differentially alternative splicing events (DASEs) and differentially alternative spliced genes (DASGs) as was reported before in other stress conditions (Appendix Table S7). Besides, number of DASEs and DASGs after prolonged stress exposure were not substantially different from ones at the early stage, confirming AS is likely a fast stress response, which was shown in similar as other studies (Calixto et al., 2018; Feng et al., 2015).

The type of AS in response to oxidative stress is also important, since each AS type can lead to a distinct fate of the new transcript and thereby can influence protein abundance and functionality differently. Retained intron (RI) normally results in an instable mRNA or a variant with a premature stop codon, resulting in lower protein levels without changing transcript level (Kashkan et al., 2022). In this work, oxidative stress reduced RI activity in both genetic backgrounds (Figure 18A), which is different compared to other abiotic stresses, where more upregulated RI events were found (Martín et al., 2021). The reason could be that to quickly response to oxidative stress, the abundance of functional proteins, which mostly are involved in DNA and RNA processes and their regulations, has

to be increased. On the other hand, other types of AS, e.g., skipped exon (SE), alternative 5' spliced sites (A5'SS), and A3'SS events, were reduced their activity in response to the stress (Figure 18A), which is consistent with that reported for most of other abiotic stresses in Martín and colleagues' study (2021). These three AS types generally remove/add a few amino acids or a functional motif from/to the full-length protein, resulting in a different function or proactivity for the new protein isoform. Therefore, downregulation of these AS types suggests a restoration of original peptides in many genes during oxidative stress. Taken together, the oxidative stress in the present study led to a distinctive attribution from different AS types compared to other abiotic stresses. It also implies that in response to various stresses, plants flexibly use different AS modulations to respond quickly and efficiently.

Furthermore, it can be seen that the alternative splicing regulation also partially reflects the dynamics of oxidative stress response depending on exposing time. At the beginning of stress, many genes involved in signal transduction and phosphorylation were spliced differently. Additionally, AS changes were observed in genes related to the regulation of gene expression at pre- and transcriptional level, mostly via DNA- and RNA-related functions, e.g., epigenetic regulation of gene expression, regulation of transcription, and demethylation (Appendix Table S6). Prolonged stress led to DAS of genes prominently involved in post-transcriptional regulation and also an increasing number of DASGs in DNA damage response. This time-scale of stress-related reprogramming of AS correlates with that of gene expression at the transcriptional level (Figure 14; Appendix Figure S4 & S5), indicating a chronologically fine-tuned response to the oxidative stress, which is further discussed in the next chapter.

### 3.2.3. Transcriptional and post-transcriptional regulation work in concert in response to oxidative stress

Upon exposure to oxidative stress, plants activate various molecular mechanisms to cope with the damaging effects. It appears that the dynamic changes at both regulatory levels, gene expression and alternative splicing, play central roles in stress responses in *A. thaliana* (Martín et al., 2021; Jabre et al., 2019; Egawa et al., 2006). In the present work, transcriptional regulation controls expression level of almost three thousand genes in response to the applied oxidative stress (Figure 14). Additionally, alternative splicing (AS) generates different mRNA isoforms from about a thousand genes, subsequently leading to proteome diversity (Figure 18). By this, AS adds a higher complexity to the gene regulation in response to oxidative stress, which would have been underestimated if only gene transcription level was analyzed.

Interestingly, though thousand genes have been differentially expressed or spliced due to oxidative stress, the number of overlapping genes between the two mechanisms was quite small. In WT, 4.43% and 11.8% of oxidative-stress-responsive DASGs were under both transcriptional and post-transcriptional regulation after 1 hour and 5 hours of stress, respectively (Figure 18, Appendix Table S7). The data in the *WHIRLY1* knock-out mutant were slightly higher (6.63% and 12.08%, respectively) (Appendix Table S7). The low percentage of overlapping DEG/DASG in the present work is in agreement with that reported for other abiotic stresses (Appendix Table S7), in which roughly 11-14% of DASGs were also DEGs. However, depending on stress and sample type, the overlap between DEGs and DASGs obviously can change. For example, Huang et al. (2019b) reported that only 5 overlapping genes among 345 DEGs and 251 DASGs (2%) were recorded in *A. thaliana* seedlings in response to elevated $CO_2$ level. Whereas, under 4 days at 4°C condition, nearly one third of DASGs (795 of 2442) were also differentially expressed in rosettes of 5-week-old *A. thaliana* (Calixto et al., 2018), compared to only 13.72% DASGs, which were also DEGs in 9-day-old seedlings growing under 10°C cold stress for one day (Martín et al., 2021 and references within). These results imply that to different extents, plants flexibly use both transcription and alternative splicing layer to fine-tune regulatory pathways according to stress conditions and developmental stages.

Strikingly, DEGs and DASGs in response to oxidative stress have quite distinct functions, indicating that different pathways are regulated by them, offering various ways to protect young seedlings (Figure 14, Figure 19). Indeed, besides high-hierachy GO terms, e.g., response to light stimuli or regulation of cellular process, which are common for both DEGs and DASGs, there are overrepresented functional groups which were specifically regulated either via differential transcription or via alternative splicing. For instance, DASGs were highly enriched with GO terms functioning in mRNA processing, especially via splicing (Figure 19), while these were underrepresented in GO analysis of DEGs. Similarly, DNA repair related genes were only regulated by alternative splicing during the stress (Figure 19). Altogether, GO analysis of those DASGs revealed that gene products are functioning mostly inside nucleus, especially in chromatin structure and other nuclear protein-contaning complexes (Figure 19). These GO CC terms were actually underrepresented among DEGs. Meanwhile, most of DEGs are related to known physiological and molecular responses in oxidative stress, such as response to different stimuli, signal transduction, transcriptional activity, ROS scavengers, ribosome and translation related activities, and various metabolite biosynthetic pathways (Figure 14). However, most of aforementioned GO terms were not enriched among DASGs during the abiotic stress treatment.

During oxidative stress, plants orchestrate these two regulatory layers (and others), to optimize the responses precisely and effciently. It was suggested that genes prone to be under AS-regulation have a short promoter region with few TF binding sites, a long genomic sequence, a high number of  introns, and weak 5' and 3' splice sites (Martín et al., 2021). *Vice versa*, genes that likely to be regulated at transcriptional level have opposite properties. Though it seems that a preferable regulation mechanism for each gene is already written in its genomic sequence, there have to be master regulators that upstream modulate and integrate abiotic, biotic, and developmental response pathways. It can be interpreted that transcription and alternative splicing works gene-wise independently but genome-wise co-dependently.

### 3.3.    AtWHIRLY1 is a novel regulator of aliphatic glucosinolate metabolism at early development

#### 3.3.1.  AtWHIRLY1 influences aliphatic glucosinolate but not indole glucosinolate metabolism

Glucosinolates (GSLs), only found in Brassicaceae plants, are secondary metabolites that play a major defensive role against pathogens and herbivores, as well as in abiotic stress responses. GSLs are constitutively produced in almost all plant tissues throughout life cycle. Besides defensive role, GSLs and their products play also a role in plant growth and metabolism. GSLs biosynthesis is regulated precisely at many levels and involves diverse components (Mitreiter & Gigolashvili, 2021). In five-day-old seedlings, the absence of WHIRLY1 showed a strict influence on GSL metabolism at transcript and metabolic levels. Transcriptionally, the expression level of Methionine(Met)-derived aGSL biosynthetic genes was significantly suppressed in the knock-out line *why1-5*. Interestingly, not all genes in aGSL biosynthetic pathway were downregulated due to knock-out of *WHIRLY1*. In fact, loss-of-function of *WHIRLY1* resulted in a specific downregulation of almost all genes involved in first steps in aGSL biosynthesis, e.g., *BCAT4, MAM1, IPMI1, BASS5,* etc. (Figure 10B). Consequently, a clear reduction in aGSL concentration was observed in the knock-out *why1-5* compared to the WT seedlings (Figure 11). Though almost all measured aGSL compounds showed the effect, significant reduction was seen mostly with short-chain Met-derived aGSLs, i.e., the 4C and 5C compounds. On the other hand, the iGSL metabolism was not affected in the *WHIRLY1* knock-out mutant. The expression level of most genes in the iGSL pathway was similar as in the WT, including of two MYB transcription factors, *MYB34* and *MYB51*. The exceptions were *CYP83B1* and *SUR1*, shared genes between aGSL and iGSL biosynthetic pathways, showed a mild reduction in expression level due to loss-of-function of *WHIRLY1* (Appendix Figure S2). Therefore, iGSL concentration in the *why1-5* were similar as in the WT seedlings at the five-day-old stage (Figure 11A).

#### 3.3.2.  AtWHIRLY1 influences Met-derived GSL biosynthesis but not Leucine biosynthesis

It was shown that methionine (Met) side-chain elongation and leucine (Leu) biosynthetic pathways share an evolutionary origin (Chen et al., 2021). Indeed, the three-step elongation cycle in Met-drived aGSL pathway is chemically similar to late steps in Leu biosynthesis (Figure 23), and the catalytic enzymes are encoded by the same gene families. For example, BCAT3/4/6 catalyze the first reaction to direct the flux of Met into aGSL biosynthesis, among them, only BCAT3 is involved in Leu biosynthesis. In the

present study, only *BCAT4* expression level significantly dropped to nearly 30% of the WT level while *BCAT3/6* were not affected in the *WHIRLY1* knock-out mutant (Figure 23). Furthermore, the next step in aGSL biosynthesis requires an isopropylmalate synthase, which is encoded by a small family of four genes, *MAM1*, *MAM3*, *IPMS1*, and *IPMS2*, in which the last two genes do not participate in side chain elongation of aGSL but are solely involved in Leu biosynthesis. Here, only the downregulation of *MAM1* expression level was observed due to knock-out of WHIRLY1. Similarly, in the next two steps of the elongation cycle, the loss of *WHIRLY1* function showed effect on the genes encoding enzymes in the Met elongation pathway, but no paralogs in Leu biosynthesis were influenced (Figure 23). These findings imply that WHIRLY1 strictly acts as a regulator in aGSL biosynthesis but not in Leu biosynthesis.
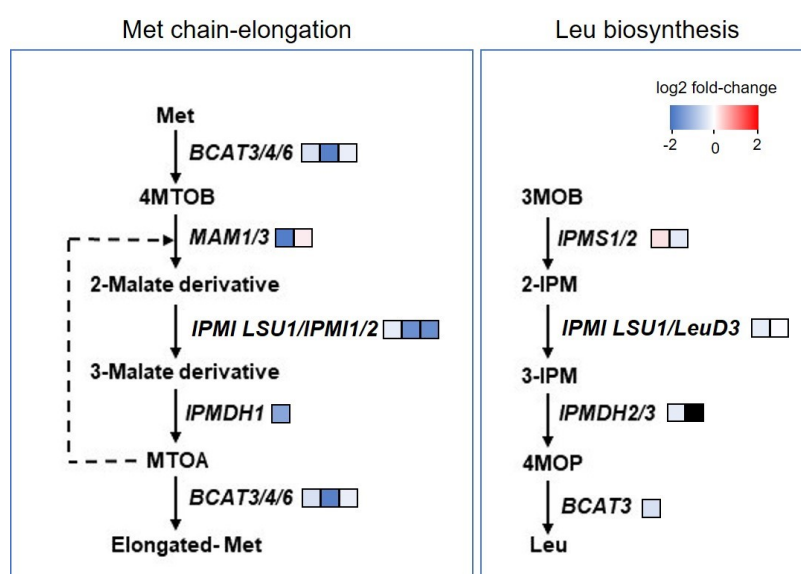


**Figure 23. Influence of knock-out of *WHIRLY1* on Met side-chain elongation (left panel) and Leu biosynthesis (right panel).** Color bars represent gene expression fold-change in the *why1-5* compared to the WT seedlings according to RNA-seq results. The scheme was adapted from (Chen et al., 2021).

### 3.3.3. AtWHIRLY1 influences aliphatic GSL accumulation during seed maturation

GSL profile of *A. thaliana* seeds and seedlings were intensively studied recently (Mitreiter & Gigolashvili, 2021; Meier et al., 2019). Due to their sulfur-rich chemical composition, GSLs are proposed to be, in addition to their defensive function, a sulfur storage, especially during early development. The turn-over of GSLs after germination plays an important role, as amino acids and sulfates are released to serve as building blocks for other molecules. It was reported that most of GSLs found at the young seedling stage actually have been already produced during embryo development and seed maturation (Mitreiter & Gigolashvili, 2021; Mérillon & Ramawat, 2017). Novel-synthesized GSLs only contribute a small amount in the metabolite contents post germination.

Therefore, GSLs measured in the young seedlings most likely reflects GSLs pre-synthesized and stored in the seeds. According to the microarray data of AtGenExpress Project number PRJNA96827 on GeneInvestigator database (Hruz et al., 2008), during embryo morphogenesis, aGSL biosynthetic genes are highly expressed before being suppressed at maturation steps (Appendix Figure S10). WHIRLY1 also maintains a high expression level during these first steps of embryogenesis (Appendix Figure S10). Probably, the missing of *WHIRLY1* also negatively impacts expression of aGSL biosynthetic genes in the embryo, leading to the reduction of GSLs in the *WHIRLY1* knock-out seedlings. This hypothesis needs to be further studied.

### 3.3.4. AtWHIRLY1 and AtMYB28 regulate aGSL genes in the same manner but possibly with different regulatory mechanism(s)

In the present work, knock-out of *WHIRLY1* suppressed expression of a specific set of GSL biosynthetic genes, which is similar to what reported in the knock-out line of *MYB28* (Figure 10C), a known major transcriptional regulator of aGSL biosynthesis (Sønderby et al., 2010). However, *MYB28* expression level in the *WHIRLY1* knock-out mutant was as similar as in the WT in the control condition (Figure 10B). Besides, upon oxidative stress, *MYB28* was upregulated, thereby leading to increased expression levels of other aGSL genes even in the *WHIRLY1* knock-out background (Figure 16C-D). Still, aGSL-related genes showed downregulation through out the experiment in the *WHIRLY1* knock-out mutant compared to the WT. Altogether, these results implicate that WHIRLY1 regulates aGSL gene expressions in a similar manner as MYB28, but at least in this early post-germination stage, WHIRLY1 functions via an independent pathway, which is not by affecting *MYB28* expression. Both regulatory pathways, WHIRLY1-mediated and MYB28-mediated, seem to function independently. Though upon oxidative stress, MYB28 was involved in upregulation of aGSL biosynthetic genes in both the WT and the knock-out of *WHIRLY1*. But the knock-out of *WHIRLY1* still affected expression level of aGSL-related genes, so that their expression levels were still much lower compared to the WT.

Interestingly, the influence of knock-out of *WHIRLY1* on GSL metabolism was not observed in the mature plants (Appendix Figure S11). This might indicate that at this sensitive early developmental stage, WHIRLY1 overruns aGSL regulation by MYB28 with another higher-order mechanism. There are many reports about function of WHIRLIES in epigenetic regulation. In barley, the *HvWHIRLY1* knock-down mutant showed changes in histone modification markers in some senescence-related genes (Janack et al., 2016). All three WHIRLIES in *A. thaliana* were found to be connected to epigenetic reprogramming of meristem tissue, both in root and shoot (McCoy et al., 2021). Recently, AtWHIRLY1 was shown to interact to the histone deacetylase AtHDA15 to modify histone markers of some

flowering-related genes (Huang et al., 2022). Perhaps, WHIRLY1 regulates GSL-related genes in this early stage of development via a similar mechanism. This has to be further investigated in the future.

Another possible scenario could be that WHIRLY1 affects *Arabidopsis* plastid development, similar as in monocots (Qiu et al., 2022; Prikryl et al., 2008), in this early development stage. This consequently, via retrograde signaling, could repress nuclear-encoded genes for plastidial proteins. In fact, most of significantly downregulated aGSL genes actually encode plastid-located enzymes in first steps of aGSL biosynthesis. As seedlings develop true leaves or leaves become mature, the effect of knock-out of *WHIRLY1* might be compensated by other WHIRLIES, in a similar manner as it was reported in *Arabidopsis WHIRLY2* knock-out mutant (Golin et al., 2020).

### 3.3.5. AtWHIRLY1 could be neo-functionalized as a novel regulator of aGSL biosynthesis

In the present work, knock-out of *WHIRLY1* strictly affected first steps of Met-derived aGSL metabolism but did not influence either iGSL or Leu biosynthesis. It was reported that Met-derived aGSLs are novel GSLs that appeared most recently thanks to whole genome duplication (WGD) events (Figure 24) (Barco & Clay, 2019). These events, named At-α and At-β, occurred after the separation of Caricaceae and Cleomaceae family, respectively. Gene duplication at At-β was the origin of novel iGSL and Met-derived aGSLs pathway. Whereas, the recent WGD event At-α, which happened nearly the emergence of the Brassicaceae family, gave rise to many side-chain modification genes for both iGSL and Met-derived aGSLs. Therefore, Met-derived aGSLs only are present in core families of Brassicales, including Brassicaceae (Mérillon & Ramawat, 2017). This evolutionary time-scale coincides with the appearance of an extra WHIRLY protein in these lineages (Figure 24).
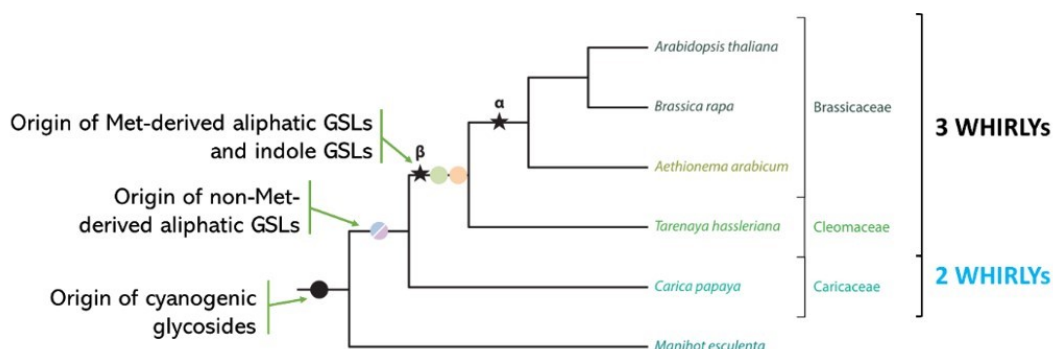


**Figure 24. Phylogeny and origin of GSLs.** The phylogenetic tree depicted major evolutionary events in GSL diversity uniquely in Brassicales. Circles represent origin of novel GSL groups and stars illustrate WGD events. Number of WHIRLIES in these families are shown on the right. The scheme was adapted from (Barco & Clay, 2019).

Gene duplication is the major drive for evolution as it provides material for natural selection. Duplicated genes could either maintain the ancestral function, or become silent (pseudogene), or develop a new function. Therefore, the co-occurance of the novel Met-derived aGSL biosynthesis and an additional *WHIRLY* locus in genome suggest that the new WHIRLY protein might obtain a new function and consequently was retained during evolution. Furthermore, synteny analysis showed that *AtWHIRLY1* and *AtWHIRLY3* share the same ancestral gene with monocot *WHIRLY1*. However, it was suggested that AtWHIRLY3 seems to be more close to monocot WHIRLY1 as they both harbor a putative copper binding motif (Krupinska et al., 2022). Additionally, as monocot WHIRLY1 is important and the missing of WHIRLY1 generally causes a severe developmental phenotype, AtWHIRLY3 also was argued to be indispensible as no *AtWHIRLY3* knock-out mutant was obtained to date (Krupinska et al., 2022). Altogether, it is possible that WHIRLY1 is the new WHIRLY copy arised after the WGD event At-β and likely serves a new function. Based on the present work, *WHIRLY1* is hypothesized to co-appear with other aGSL genes during evolutional course and neo-functionalized to be a regulator of these genes, at least at the post-germination seedling development stage.

## 3.4.    Function of AtWHIRLY1 in oxidative stress response

The functions of WHIRLIES, including AtWHIRLY1, in response to different abiotic and biotic stresses are thoroughly discussed in two recent reviews (Krupinska et al., 2022; Taylor et al., 2022). However, little is known about the involvement of WHIRLY1 in oxidative stress, especially in the early development stage. It was suggested that AtWHIRLY1 is involved in retrograde signaling, transmitting oxidative stress signal from chloroplast into the nucleus (Foyer et al., 2014). Besides, based on GeneInvestigator database (Hruz et al., 2008), upon pertubations that likely disturb ROS homeostatis, e.g., dark-light transition, increasing of light intensity, and reducing temperature, the transcript level of *AtWHIRLY1* is induced significantly. Under oxidative stress in the present work, the expression of *AtWHIRLY1* and other *WHIRLIES* was gradually upreguated and elevated its expression level by about two-fold after 5 hours (Figure 13). The mild upregulation of a transcriptional regulator in response to stress can still exert a profound impact on transcriptome as it can induce a regulatory cascade influencing a wide range of responsive genes. Indeed, analyzing effects of loss-of-function of *WHIRLY1* on nuclear gene expressions and splicing events allows to clearly pinpoint the role of WHIRLY1 in a concerted regulation of stress response mechanisms.

In other species, WHIRLY1 also was reported to gradually be induced in response to stress. For example, in tomato, *SlWHIRLY1* doubled its transcript level after 3 hours under cold or in a treatment with 20mM $H_2O_2$ for 6 hours (Zhuang et al., 2019). Zhuang and colleagues also showed that the protein accumulated in both chloroplasts and nucleus upon the abiotic stress, with a noticeable increasing amount after 6 hours of stress. In the present work, missing of a specific antibody against AtWHIRLY1 makes it difficult to test oxidative-stress-related changes in the WHIRLY1 protein level in the cells as well as its localization in either chloroplasts or the nucleus. Still, analyzing stress-related changes in transcript levels of *WHIRLIES* give an indication of total changes in their protein levels. Nevetheless, findings in the present work are independent from knowing WHIRLY1's exact distribution between chloroplast and nucleus, which has to be clarified in future investigations.

### 3.4.1.  AtWHIRLY1 influences transcription rate both positively and negatively

The dynamics of WHIRLIES in regulating gene expression upon stress conditions was studied in overexpressing (OE) lines of other plants (Manh et al., 2023; Zhuang et al., 2019). In tomato, it was shown that an increased level of WHIRLY1 led to upregulation of many photosynthesis-related genes during chilling stress but not at control condition. However, there was also a number of photosynthesis-related genes, which were induced

in the OE lines already at control condition, which were then all repressed due to low temperature in both lines (Zhuang et al., 2019). Similar effects are shown for genes related to chilling response, which were all upregulated during stress in both genetic backgrounds, however these genes showed a diverse pattern of expression fold-change due to gain of function of *SlWHIRLY1*. Specifically, *SlAMY3-L* encoding α-amylase was upregulated in the *OE-SlWHY1* line while downregulated in WT upon cold; while *SlISA2*, which promotes starch synthesis, was repressed in both lines but with a higher fold in the mutant, leading to an accumulation of soluble sugars and a reduction of starch during chilling stress.

Recently, Manh et al. (2023) investigated the changes in transcriptome of an *HvWHIRLY1* over-expression line during drought-induced senescence. Many genes involved in the senescence process, biotic- and abiotic-stress responses, including ABA biosynthesis were only upregulated in WT but not in the OE line. Among them, ABA-biosynthesis-related genes showed changes in expression pattern due to gain of WHIRLY1's function. Specifically, *HvNCED1*, *CCD5*, and *CCD8*, which encode three enzymes in glucosidase-mediated and stress-activated ABA production, were induced only in the WT but not in the overexpressing line upon drought stress. Whereas, ZEP5 which can be involved in the first step of ABA synthesis exhibited downregulation due to the overexpression of *HvWHIRLY1*, and also was repressed upon drought-induced senescence in both genetic backgrounds. Meanwhile, several genes of later steps in ABA metabolism showed higher expression level in the OE line compared to the WT. Altogether, it appears that WHIRLY1, especially upon stress, specifically affect nuclear gene expression pattern and by this, contribute to the responsiveness and adaptability of the plant toward unfavorable conditions.

Similarly, in the present work, the oxidative stress-related reprogramming of gene transcription was in principle similar in the *WHIRLY1* knock-out and in the WT seedlings, as seen with volcano plot (Appendix Figure S3) and also the heatmap indicating DEGs (Figure 14C). In both lines, similar sets of genes were regulated, which have function in coping with oxidative stress, though there was a slightly difference in gene expression fold-change. Besides, upon oxidative stress, a number of DEGs in the knock-out mutant *why1-5* was higher than that in the WT. Despite of an in general similar transcriptomic reprograming in both lines, it is worth noting that many genes exhibited quite noticeable differences in term of expression pattern upon oxidative stress in the loss-of-function of *WHIRLY1* seedlings (Figure 16&17). In many cases, the missing of WHIRLY1 did not completely switch on/off target gene transcription but rather affected the time of onset and the magnitude of expression changes. Indeed, many oxidative-stress-associated DEGs were repressed/induced after prolonged stress, but their expression levels were already down-/up-regulated after 1 hour of stress in the mutant *why1-5* when compared to the WT,

such as *SQE2* and *JAL10* (Figure 17A). Meanwhile, other genes only showed differences in the expression fold-change due to loss of WHIRLY1's function in a specific investigated timepoint, such as *CLH1* and *SUS4* (Figure 17A).

This phenomenon is quite interesting as it suggests that WHIRLY1 might not act as a direct transcription factor governing expression of some specific genes but rather, as a ssDNA-binding protein, interacts with the main TF(s) and/or the transcriptional complex components, or is involved in a higher order of regulation, such as chromatin remodeling. In fact, it was observed that other ssDNA/RNA binding domain-containing transcriptional regulators have a similar effect on gene transcription. For instance, Sub1 is a yeast ssDNA binding protein and also a highly conserved transcriptional regulator in the eukaryote cell. The knock-out of *Sub1* in yeast affected sporulation in a way that the expression level of the related genes could not reach the optimum level, though their expression level were still increased during the process (Gupta et al., 2015). This is due to the fact that Sub1 can bind to the promoter region of its target genes, stablize the ssDNA, and physically interact with components of transcription preinnitial complex. This transcriptional coactivator modulates various transcription stages, from initiation to termination, in which positive or negative effects are based on its phosphorylation status (Garavís & Calvo, 2017). Perhaps, WHIRLY1 can function in a similar manner, as WHIRLY1 can be phosphorylated and two isoforms of WHIRLY1 were found in the nucleus (Zhuang et al., 2019; Ren et al., 2017). In *Arabidopsis*, the phosphorylated form of WHIRLY1 had higher binding affinity to *WRKY53* promoter, leading to a suppression of *WRKY53* expression at the early senescence stage (Ren et al., 2017). In tomato, the presence of two bands of SlWHIRLY1 in the nuclear fraction on immunoblot also suggests that WHIRLY1 exists in the nucleus in two phosphorylated states, and the dynamics in the two isoform ratio during cold and $H_2O_2$ suggests that the function of nuclear WHIRLY1 depends on its phosphorylation status (Zhuang et al., 2019).

### 3.4.2. AtWHIRLY1 affects alternative splicing of nuclear genes under oxidative stress

In plants, genes encoding for stress responsive proteins/regulators generally exhibit multiple AS isoforms. By modulating the ratio between different isoforms, specifically active versus non-active isoforms, AS can fine-tune functional protein level in response to unfavorable conditions (Laloum et al., 2018). Therefore, factors that alter these dynamics can influence the performance of plants in response to environmental stressors. In the present work, knock-out of *WHIRLY1* impacted the ratio of different AS variants of ROS-responsive genes substantially. Firstly, in the *WHIRLY1* knock-out mutant, there was a smaller number of differentially alternative splicing events (DASEs) and differentially

alternative spliced genes (DASGs) in response to oxidative stress compared to the WT (Figure 18). This result is opposite to the transcriptional regulation where the mutant appeared to have more DEGs (Figure 14). These findings imply that WHIRLY1 is involved in both regulatory levels. Secondly, the *why1-5* line and the WT exhibited a high ratio of unique DASEs and also DASGs in response to oxidative stress. Moreover, nearly a half of overlaping DASEs between the two lines were oppositely regulated (Figure 19B; Appendix Figure S6), meaning that the actual number of AS events differently regulated in the mutant is even higher than the ordinary number of DASEs. Furthermore, even in the same DASGs, different preferable variants were observed in each line (Figure 19A). Clearly, oxidative stress, together with knock-out of *WHIRLY1* significantly affected the the AS activity of nuclear genes. Consequently, many genes were spliced differently under oxidative stress between the two genetic backgrounds (Figure 20).

In the control condition, the missing of WHIRLY1 affected AS pattern of some genes in a similar way as when WT seedlings were exposed to oxidative stress (Figure 22, discussed in Chapter 3.1.3). That could be the reason why under stress conditions, the mutant did not change the overall AS pattern as similar as observed in the WT, since in the mutant, it was already pre-set for the stress condition. For instance, the alternative 3' splicing site (A3SS) variant of *BPL2*, which is needed for oxidative stress response, was already present in a higher ratio in the *WHIRLY1* knock-out compared to the WT before the stress (Figure 21A). Hence, many mRNA variants likely were mildly adjusted to the optimum ratio under stress in the *WHIRLY1* knock-out line (Figure 19B; Appendix Figure S6). Nevertheless, AS of several other genes was regulated similarly in both lines upon oxidative stress, such as *FSD2* and *SR34b* (Figure 21B-C).

Alternative splicing actually is regulated by the selection of splicing sites (SS), in which enhancers such as SR proteins increase the usage of SS, while repressors such as hnRNP proteins reduce it (Laloum et al., 2018). The dynamics of ratio between action of enhancers and repressors and also the local concentrations of these RNA splicing factors determines AS. Besides, spliceosome is formed *de-novo* on the splicing sites in a step-wise manner where in every round, the interaction between RNAs and proteins in the complex themselves or with other regulatory factors are critical for splicing site recognition, assembly, and efficiency. In the present work, knock-out of *WHIRLY1* did not significantly change expression level of these RNA splicing factors under oxidative stress. On the other hand, the missing of WHIRLY1 clearly affected AS of various splicing factors, such as *RS31A, SR34,* and *SR45a*. The deregulation of AS activity of these splicing factors can change their overal activity, leading to the different AS of other nuclear genes.

Additionally, it is possible that WHIRLY1 can bind to pre-mRNA at the splicing site flanking sequences. As RNA and ssDNA share similar chemical properties, WHIRLY1 can

bind to RNA likely with also a high affinity but little specificity. In plastids, WHIRLY1 was shown to affect splicing of intron group II and was suggested to be involved in intron folding (Melonek et al., 2010; Prikryl et al., 2008). Besides, it seems that WHIRLY1 is prone to bind to AT-rich sequences, such as EREs that overlap wtih W-box, telomere sequence, etc. Meanwhile, though the SS in *Arabidopsis* is also the conserved GT-AG motif, flanking sequences of SS are rich with A and T (Appendix Figure S12). Furthermore, 70 last nucleotides of *Arabidopsis* introns are prone to contain T followed by A (Appendix Figure S12). Hence, if WHIRLY1 binds to these sequences, it might change the interaction between pre-mRNA and other splicing factors, changing thereby the rate of spliceosome formation and execution. Moreover, as similar to plastid WHIRLY1 function, nucleus-located WHIRLY1 could affect the folding of intron to facilitate splicing. The hypothesis should be further verified.

### 3.4.3. AtWHIRLY2 and/or AtWHIRLY3 may complement AtWHIRLY1 loss-of-function under oxidative stress

It is observed that there was a rather small number of DEGs due to loss of *WHIRLY1* function under oxidative stress, even smaller than that under control condition. The slight effect of knock-out of *WHIRLY1* on gene expression level possibly because of other WHIRLIES complementing WHIRLY1's function after prolonged stress-treatment, as all three *WHIRLY* genes in *A. thaliana* were induced significantly after 5 hours (Figure 13). As discussed in chapter 3.1, the possibility that WHIRLIES can replace others in specific conditions is feasible. Here, WHIRLY3, as a homologous protein and also sharing the same localization, could be the most promising candidate to cover WHIRLY1 function in the nucleus. Additionally, the oxidative stress, which causes a ROS burst in chloroplasts, can affect the redox balance in the whole plant cell, including other compartments as mitochondria. Indeed, *AOX1A*, a marker of mitochondrial retrograde signaling was upregulated in my experiment after prolonged oxidative stress (Appendix Table S4). As WHIRLY2 is located in mitochondria, WHIRLY2 might move into the nucleus and in turn affects ROS-responsive genes.

Regardless WHIRLY1's mode of action in regulation of gene expression, its function is likely related to DNA-binding activity via the conserved ssDNA-binding domain Whirly. Up to date, all binding motifs of WHIRLIES were proven by EMSA or ChIP assay and all showed high affinity of binding, though these motifs share few similarity to each other (reviewed by Krupinska et al., 2022). Moreover, high through-put genome-wide scale analysis failed to identify a specific binding motif for WHIRLIES (O'Malley et al., 2016; Franco-Zorrilla et al., 2014). Therefore, it was suggested that WHIRLY1 binds to any ssDNA sequences (Cappadocia et al., 2010, 2012). Possibily, other WHIRLIES also do,

as they share a similar conserved ssDNA-binding domain (Figure 1, Appendix Figure S1). However, there was still a specific subset of deregulated genes in the loss-of-function of *WHIRLY1* under oxidative stress, suggesting that regulation of these DEGs could not be complemented at least completely by other WHIRLIES. The differences can be explained by different affinity between WHIRLIES and ssDNA as well as other proteins/TFs in transcription complex. This hypothesis needs to be studied further. In a similar manner, the function of WHIRLY1 in alternative splicing (AS) could also have been compensated by WHIRLY3 or WHIRLY2 in the knock-out mutant.

### 3.4.4. AtWHIRLY1 regulation of gene expression under oxidative stress might depend on its interaction with other TFs

Though in the present work only a small number of genes was differentially regulated due to the loss of *WHIRLY1* function, most of them were deregulated in a condition-dependent manner. For instance, *FSD2* and *RAV1*, which were differentially expressed at the control condition due to knock-out of *WHIRLY1*, under stress, especially prolonged oxidative stress, no longer showed a different expression level between the two genetic backgrounds. Meanwhile, *SUS4*, *IDQ11*, and *CHL1*, etc., were regulated differently in the *WHIRLY1* knock-out line only under stress but not under normal growth condition.

Gene expression is intricately regulated by not only one but many transcriptional regulators via multiple *cis*-elements on the promoter region and even within the genomic sequences. As discussed in chapter 3.4.1, it is possible that WHIRLY1 interacts to main TF(s) and the transcription preinitiation complex on the target promoter, instead of being a main TF itself. In fact, WHIRLY1 was found in the complex binding to the ERE motif, which is overlapping with W-boxes and TGACG elements, which are binding sites of WRKY and TGA TFs, respectively (Desveaux et al., 2005). Besides, WRKYs and TGA1 have been identified as interacting proteins of WHIRLIES (Liu et al., 2018; Trigg et al., 2017). Noticeably, the WRKY and TGA proteins prefer to bind to dsDNA, while WHIRLY1 has a substantially higher affinity to ssDNA. These clues actually suggest that perhaps, a TF initially binds to its binding site on an enhancer, which is located proximal from the core promoter, leading to assembly of the pre-initiation complex which contains general transcription factors, and recruits RNA Polymerase II at core promoter region of the target gene. This action causes meltings of the DNA duplex and hence activation of the transcription and in the meantime, exposing ssDNA for WHIRLY1 binding. In this circumstance, WHIRLY1 is not a direct prominent regulator of the target gene, but it likely is being recruited by interacting with the main TF(s), and thereby affects gene transcription. Therefore, the difference in regulation of nuclear genes in a condition-dependent manner

in the *WHIRLY1* knock-out seedlings perhaps is due to the flexible interaction between WHIRLY1 and other developmental-/stress-related main TFs. When WHIRLY1 interacts to the main TF that controls a target gene, missing of WHIRLY1 will affect its expression under such specific condition, given that for such genes, other WHIRLIES could not completely complement WHIRLY1 function.

Analysis of *cis*-elements on promoters of genes that were deregulated in the knock-out mutant at the control and oxidative stress conditions showed a bias in the distribution of putative TF binding sites (Appendix Figure S14). Genes whose expression levels were affected only in the control condition due to knock-out of *WHIRLY1* contain binding sites for TF families that are prominently involved in developmental regulation, including C2H2-type zinc fingers TFs (e.g., MGP and Adof1 that regulate DNA-template transcription), bZIP TFs (e.g., ABI5 which is the master regulator of ABA signaling and regulation), and bHLH TFs (e.g., PIF7 which is important for light response during early development). On the other hand, majority of DEGs between the knock-out *why1-5* and the WT seedlings under oxidative stress conditions, contain *cis*-elements of TF families that are involved in hormone regulation and stress response. For instance, 119 promoters (per 149 genes) contain the binding motif for WRKY28, one TF of the WRKY family which plays a critical role in stress responses. Besides, other TFs families MYB and NAC also have enriched binding sites on promoters of unique ROS-responsive DEGs. These findings suggest that WHIRLY1 potentially regulates target genes depending on the environmental and developmental conditions via interacting with these TFs.

Moreover, for aliphatic glucosinolate biosynthetic genes, which consistently had lower expression level in the *WHIRLY1* knock-out mutant (Chapter 3.3), their promoters also harbor different binding sites for both MYB TFs, such as MYB28 and MYB29, and many W-boxes and Inverted Repeat 2 (IR2) sequences, which are potential binding sites for WRKY and ERE TFs, distributing through out promoter sequences. Perhaps, WHIRLY1 modulates aGSL gene expression via modulating the interaction with/within these TFs. Hence, investigation of physically/biologically interactions between WHIRLY1 and other TFs in the transcriptional complex context, as well as the binding of WHIRLY1 to the common DNA binding sites can be useful.

## 3.5.    Conclusions

### 3.5.1.  AtWHIRLY1 as a central regulator at early seedling developmental stage

After germination, seedlings need to adjust quickly to the environment as well have to reprogram their metabolite-portfolio toward an autophototrophic organism. During this process, abiotic and biotic factors can intefere the seedling development, which requires a quick response. However, this response has to be balanced with the ongoing growth process. Based on results from the present work, it is proposed that WHIRLY1 plays an important role in balancing nuclear gene expression, including both regulatory levels, transcription and post-transcription. At control conditions, WHIRLY1 modulates gene expression in both positively and negatively manner, especially genes which play a role in cross-talk between multiple stress responses and development, e.g., *RAV1* and *FSD2*. Strikingly, in this work, AtWHIRLY1 was found for the first time to specifically regulate expression of genes involved in the first steps of aliphatic GSL metabolism. In addition, it for the first time shows that WHIRLY1 appears to significantly influence the alternative splicing of nuclear genes. Specifically, the knock-out mutant of *WHIRLY1*, *why1-5*, exhibited an AS pattern at control conditions as if the seedlings were under an abiotic stress.

Upon oxidative stress, seedlings respond by hamorizing multiple regulatory layers, including gene expression and alternative splicing regulation. WHIRLY1 is also involved in these molecular responses, in which knock-out of *WHIRLY1* exerts a higher impact on the post-transcriptionally network. In this work, the missing of WHIRLY1 altered the intensity and also the onset of transcription of oxidative-stress-related genes. Additionally, WHIRLY1 specifically affected the splicing site selections upon the oxidative stress, leading to significant differences in AS pattern between the WT and the *WHIRLY1* knock-out mutant.

Interestingly, it appears that the effect of loss of *WHIRLY1* function on nuclear gene expression depends on the environmental and developmental conditions. It is postulated that by interacting with different transcription factors, WHIRLY1 balances developmental and/or abiotic and biotic stress-related regulatory pathways. My findings suggest that WHIRLY1 under the control and stress condition acts on different levels, possibly via interacting with transcription factors, splicing factors, and also the epigenetic machinery, to orchestrate the whole transcriptome of the seedlings in this sensible growth phase.

### 3.5.2. Proposed working model

There is emerging evidence that epigenetic modifications, transcription rate, and alternative splicing events are mechanistically connected. They are orchestrated by the dynamic changes in chromatin landscape in response to stress or developmental stimuli (Jabre et al., 2019 and references within). WHIRLY1 has been reported to physically interact or to be biologically related to all of these regulatory complexes. Epigenetically speaking, all three WHIRLIES in *Arabidopsis* were reported to be involved in epigenetic reprogramming of meristem tissue, both in root and shoot (McCoy et al., 2021). Recently, AtWHIRLY1 was shown to physically interact to HDA15 and both modulate histone modification of some flowering-related genes (Huang et al., 2022). In barley, reduction of HvWHIRLY1 content affected euchromatin markers of some senescence-related and ABA biosynthetic genes (Manh et al., 2023; Janack et al., 2016). At transcription level, the influence of WHIRLY1 on nuclear gene expression has been well-documented. WHIRLY1 was shown to positively affect selected genes, which function in different aspects of plant development and defense (Yan et al., 2020; Zhuang et al., 2020; Liu et al., 2018; Huang et al., 2017, 2018; Janack et al., 2016; Comadira et al., 2015; Miao et al., 2013) . Besides, the altered concentration of WHIRLY1 also impacted transcription both in negative and positive way (Manh et al., 2023; Huang et al., 2022; Zhuang et al., 2019; and this work). Additionally, WHIRLY1 was proved to bind to multiple DNA sequences *in vitro* (Cappadocia et al., 2012), which are similar to the binding sites of other TFs. Perhaps by associating with the same binding sites, WHIRLY1 can interact with these TFs, such as the WRKY or TGA family (Liu et al., 2018; Trigg et al., 2017; Desveaux et al., 2005). Besides, due to the similarity in chemical properties between ssDNA and RNA, WHIRLY1 can also bind to RNA via its conserved ssDNA binding domain Whirly. Therefore, WHIRLY1 was reported to affect RNA splicing in plastids (Melonek et al., 2010; Prikryl et al., 2008). In this work, it is shown for the first time that WHIRLY1 affects splicing of pre-mRNA also in the nucleus.

Altogether, from previous knowledge and from results of the present work, I propose that WHIRLY1, based on its ability to bind ssDNA and also RNA, is a central component of regulatory hubs, where it, by interacting with transcription factors and other proteins in transcription complexes/splicing complexes/epigenetic regulatory complexes affecting chromatin structure, has an impact on multiple gene regulatory levels in the nucleus, acting in fine-tuning and balancing several regulatory pathways (Figure 26).
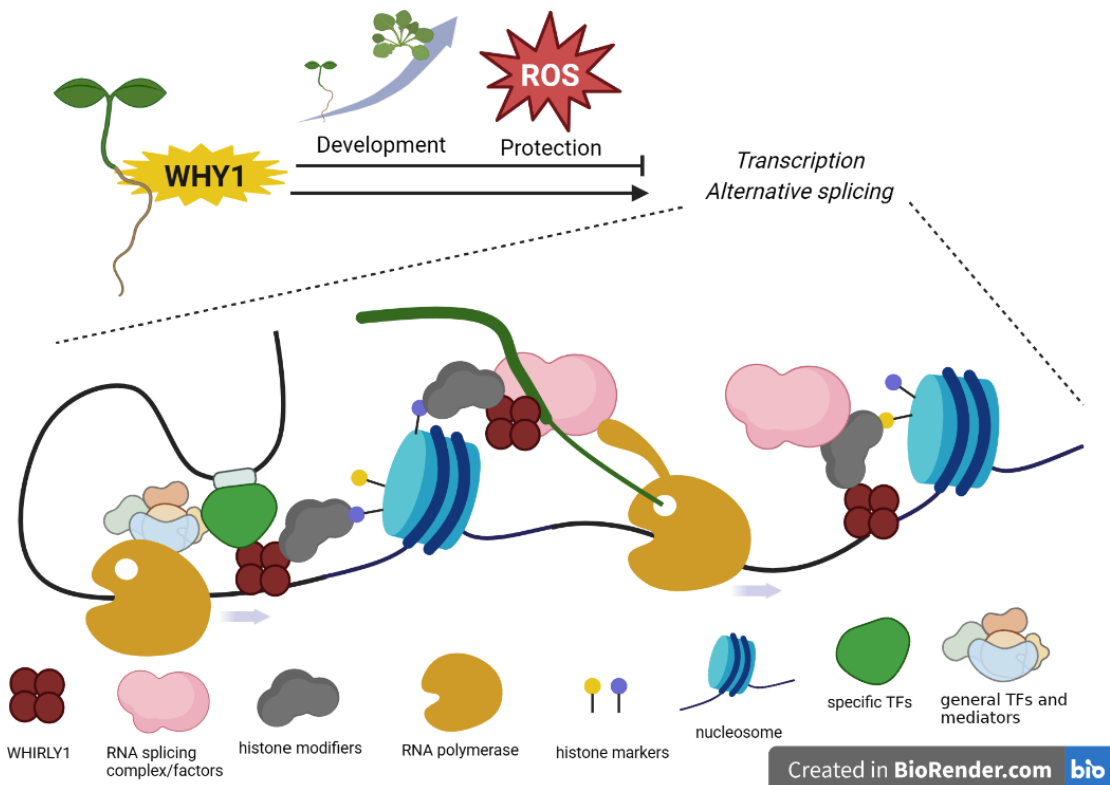
**Figure 26. Working model of AtWHIRLY1 as a central regulator of nuclear genes in early seedling development.** In young seedlings, WHIRLY1 regulates and fine-tunes nuclear gene expression and alternative splicing under both non-stress and stress conditions. Target genes of WHIRLY1 are diverse, including genes that are involved in development, stress response, and hormone signaling, especially aliphatic glucosinolate biosynthesis. Controlling of gene regulation by WHIRLY1 is concerted at multiple levels, and WHIRLY1 is placed at the center of a regulatory hub. WHIRLY1 can be recruited at the opening promoter region by interacting with the main transcription factor (TF) of a target gene. The main TF is either developmental- or stress-related, and therefore determines biological function of WHIRLY1. By binding to the melted DNA, WHIRLY1 helps to stabilize the transcriptional complex and thereby may influence transcriptional initiation. The binding of WHIRLY1 upstream of Pol II can also affect the elongation speed, which consequently influences the transcription rate. In the mean time, WHIRLY1 can bind to nascent pre-mRNA and might affect the interaction between pre-mRNA and the RNA splicing machinery, leading different splicing patterns. The binding of WHIRLY1 in the promoter region can also change histone markers as WHIRLY1 is able to interact with histone modification complex. Histone markers of a nearby nucleosome can also affect the recruitment of RNA splicing factors to nascent pre-mRNA. The nascent pre-mRNA molecule is presented in green while DNA is shown in blue, with exons are shown by thicker lines. The large subunit of RNA polymerase II (Pol II) is shown in yellow, transcribing from left to right as depicted by arrows.

# 4. MATERIALS AND METHODS

## 4.1.    Materials

### 4.1.1.  Chemicals, enzymes, and kits

All chemicals, enzymes, and kits used in the present work were qualified for molecular biotechnology and provided by well-known companies, including Carl Roth GmbH & Co. KG (Germany), AppliChem GmbH (Germany), Sigma-Aldrich GmbH, ThermoFisher Scientific (USA), Kapa Biosystems (USA), NEB (USA), etc.

### 4.1.2.  Primers

Primers were designed and selected using the NCBI Primer-BLAST tool and IDT OligoAnalyzer tool. All oligoes were synthesized by Eurofins Genomics (Germany) or IDT (Germany). Primer names and sequences are provided in the Appendix Table S8.

### 4.1.3.  Plant materials

The CRISPR/Cas9 knock out lines (denoted *why1-4,* and *why1-5*) of AtWHIRLY1 (At1g14410) were created by Prof. Götz Hensel research group at Heinrich-Heine-University Düsseldorf. Primary transformation (T1) plants were screened on Murashige and Skoog (MS) media supplemented with 33.3 µg/ml hygromycin B. Resistant seedlings were checked the presence of DNA encoding Cas9 endonuclease by PCR with primers Cas9-F2/Cas9-R2. The next generation (T2 and T3) of Cas9-positive T1 plants were further analyzed. The WHIRLY1 locus was amplified by PCR using primers CRISPR-WHY1-for/CRISPR-WHY1-rev and then Sanger sequenced using CRISPR-WHY1-for. The mutated-WHIRLY1-containing lines were also checked for the presence of Cas9. As the result, two independent Cas9-free homozygous *WHIRLY1* mutant were identified, *why1-4* and *why1-5*, which contain frameshift mutations (25_27DelAA and 27_28InsT, respectively). The homozygous CRISPR/Cas9 *AtWHIRLY1* knock-out mutants were genotyped by PCR-couple CAPS analysis, which described in Chapter 4.2.4.

The T-DNA insertion line of AtWHIRLY1 (SALK_023713) was obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). Homozygous T-DNA insertion was confirmed by PCR using following primers: why1-t-for, why1-t-rev, and LBa1.

## 4.2.    Methods

### 4.2.1.  Growth condition and oxidative stress experiment for seedlings study

Seeds of *A. thaliana* were sown on half-strength MS medium supplemented with 1% sucrose and 0.7 % plant agar and imbibed in cold-dark room for 3 days. The seeds were germinated under continuous light intensity of 15 µE/m$^2$s at 22 ± 2°C for 5 days. In order to investigate the involvement of AtWHIRLY1 in seedling development, particularly in response to oxidative stress, an abiotic stress experiment was performed following a similar protocol as described in a study by (Kim et al., 2012). The experiment was performed by a 30-min dark-adaptation of 5-day-old seedlings grown under aforemention conditions, followed by a combination of low temperature (12°C) and high light intensity conditions (270 µE/m$^2$s). The whole seedlings were collected before dark adaptation (as control samples) and after one hour and five hours under the stress to perform gene expression analysis and RNA-sequencing. To measure glucosinolate content, the whole seedlings at control, 5-hour-, and 24-hour-oxidative-stress conditions were used. The maximum photosystem II (PSII) efficiency was measured at the control condition as well as 2 days and 4 days under the stress.

### 4.2.2.  Determination of phenotypic parameters

#### *4.2.2.1.    Measurement of maximum PSII efficiency*

The FlourCam7 version 1.2.5.3 system (PSI, Czech) was used to measure the maximum PSII efficiency of young seedlings. First, seedlings were dark-adapted for 15 min to determine the mimimum chlorophyll fluorescence yield (Fo value). Then, plants were exposed to a saturating flash light, leading to the maximum fluorescence emission (Fm value). The maximum PSII value was calculated as the ratio Fv/Fm, where Fv is the variable fluorescence (Fv = Fm - Fo).

#### *4.2.2.2.    Determination of pigment contents*

To determine pigment contents, 20-30 mg of seedling shoot was collected and finely ground in liquid nitrogen. Then, plant material was homogenized with 200 µl of Aceton/Tris buffer (80% aceton in 10 mM Tris-HCl pH7.8) by vigorously vortexing. The mixture was incubated for 10 min in a cold ultrasonic bath before centrifuging at 13 000 rpm for 10 min at 4°C. Next, the supernatant was transferred into a new centrifuge tube. The absorption values of the supernatant were measured using a NanoPhotometer® NP80 (Implen GmbH, Germany) at following wave lengths: $A_{470nm}$, $A_{537nm}$, $A_{647nm}$, $A_{663nm}$, $A_{750nm}$. Then,

pigment concentration was calculated by Excel using equations of (Sims & Gamon, 2002) as followed:

Chlorophyll a = 0.01373*$A_{663}$ - 0.000897*$A_{537}$ - 0.003046*$A_{647}$

Chlorophyll b = 0.02405*$A_{647}$ - 0.004305*$A_{537}$ - 0.005507*$A_{663}$

The unit for all equations is µmol/ml. The pigment content was normalized to total Aceton/Tris buffer volume and sample fresh weigh.

### 4.2.2.3.  Measurement of growth-related parameters

To determine morphological characteristics, seedlings were grown on half-strength MS agar plate vertically and then pictured from the side. First, a scale was set according to a ruler included in the picture using Fiji ImageJ (Schindelin et al., 2012). To measure hypocotyl or root length, cursor was used to define the fragment of interest. Then, the length of fragment was measured automatically.

### 4.2.3.  DNA isolation

To isolate DNA, leaf sample was finely ground in liquid nitrogen and collected into a 2 ml centrifuge tube. Then, the plant material was homogenized in 1 ml extraction buffer (100 mM Tris-HCl pH8.0, 500 mM NaCl, 50 mM EDTA, 1.5% w/v SDS, and 10 mM β-mercaptoethanol) followed by inculation at 65°C for 10 min. Next, 300 µl acetous potassium acetate (3 M potassium acetate and 2 M acetic acid) was added and the mixture was incubated at 4°C for 10 min. The tube was centrifuged at 13 000 rpm for 10 min at 4°C to collect the supernatant into a new 2 ml centrifuge tube. Subsequently, 300 µl phenol:chloroform:isomyl alcohol (25:24:1) solution was added and the mixture was vortexed thoroughly. After that, phase separation was done by centrifuging at 8 000 rpm for 5 min at room temperature, followed by transferring the upper phase to a new 2 ml centrifuge tube. Then, the sample was mixed thoroughly with 500 µl cold isopropanol and incubated at -20°C for 10 min. Nucleic acids were precipitated by centrifugation at 13 000 rpm for 20 min at 4°C. The supernatant was discarded and pellet was washed with 1 ml 70% ethanol by centrifugation at 13 000 rpm for 5 min at 4°C. Pellets was dried completely using a heat-block at 37°C before dissolving in 100 µl 10 mM Tris-HCl pH8.0 supplement with 100 µg/ml RNase A (ThermoFisher Scientific, USA) to remove RNA. A second precipitation step was performed by adding 10 µl 3 M $CH_3COONa$ pH5.0 and 250 µl pure ethanol followed by centrifugation at 13 000 rpm for 20 min at 4°C. Supernatant was removed and pellet was washed with 1 ml 70% ethanol. Finally, dry pellet was dissolved in 100 µl 10 mM Tris-HCl pH8.0. DNA concentration was estimated using a NanoPhotometer® NP80 (Implen, Germany).

### 4.2.4. PCR-coupled CAPS marker analysis to genotype *WHIRLY1* CRISPR/Cas9 knock-out mutants

The CRISPR/Cas9-based mutated *AtWHIRLY1* alleles were confirmed by cleavage amplified polymorphism site (CAPS) approach. Firstly, the segment containing mutated site was amplified by PCR, in which each reaction comprises of 1 µl of diluted DNA (100 ng) as the template, 2 µl 10X *Taq* DNA Polymerase Buffer, 0.4 µl 25 mM MgCl$_2$, 1.2 µl each primer including why1-cc-for and why1-cc-rev, and 0.1 µl 10 U/µl *Taq* DNA Polymerase (EURx, Poland). Then, 5 µl PCR products was digested with restriction enzyme (RE) in a 20 µl reaction with 2 µl 10X NEBuffer 3.1 and 1 µl 10 U/µl *Bsl*I (New England Biolab, USA). The reaction was performed at 37°C for 3 hours. The RE-treated products was visualized by electrophoresis on a agarose gel. The WT allele gave a non-cut fragment of 231 bp while both CRISPR/Cas9-based mutated alleles, *why1-4* and *why1-5*, were digested into two products of 126 bp and 105 bp.

### 4.2.5. Agarose gel electrophoresis

To visualize PCR or RE-treated products, DNA was mix with 6X DNA loading dye (ThermoFisher Scientific, USA). Then, the mixture was loaded on a 1.5% agarose gel supplemented with 4 µl DNA Stain Clear G for 100 ml agarose gel (SERVA, Germany). A GeneRuler DNA ladder mix (ThermoFisher Scientific, USA) was also loaded. DNA were seperated by electrophoresis at 80 V for 35 min in 1X TBE buffer (90 mM Tris base, 90 mM boric acid, 2 mM EDTA). Gel was pictured using E-box Gel Documentation Imaging (Vilber, Germany) system.

### 4.2.6. Isolation of total RNA

To isolate total RNA, a method using TRIzol Reagent (Parasyri et al., 2022) was used with additional Dnase treatment step. Firstly, plant material was finely ground using liquid nitrogen and collected into a 1.5 ml centrifuge tube. Then, 1 ml of 60°C pre-heated TRIzol buffer (38% acidic water-saturated phenol, 0.8 M guanidinium thiocyanate, 0.4 M ammonium thiocyanate, 0.1 M CH$_3$COONa pH5.0, and 5% glycerol) was added and sample was incubated at 60°C for 10 min. After centrifugation at 13 000 rpm for 10 min at 4°C, supernatant was transferred into a new 1.5 ml centrifuge tube and homogenized with 200 µl chloroform by thoroughly vortexing for 1 min. Next, phases were seperated by centrifuging at 13 000 rpm for 15 min at 4°C. The upper aqueous phase containing RNA was collected to a new 1.5 ml centrifuge tube. Then, 250 µl sodium buffer (0.8 M Na-citrate and 1.2 M NaCl) and 250 µl isopropanol were added to precipitate nucleic acid in the solution. The sample was incubated at room temperature for 10 min followed by

centrifugation at 13 000 rpm for 15 min at 4°C. Supernatant was removed and the pellet was washed with 1 ml 70% ethanol. The pellet was dried completely and dissolved in 30 µl DEPC-treated water. Then, a DNase treatment was done by adding 4 µl 10X DNase buffer, 0.5 µl RNase-free 10U/µl DNase (Roche, Germany), and 5.5 µl water into each tube, then incubating at 25°C for 30 min. After that, another purification step was applied to completely remove DNase from the RNA mixture. There, 260 µl DEPC-treated water and 100 µl phenol:chloroform:isoamyl alcohol were added into DNase-treated sample. After vortexing for 15 sec, sample was centrifuged at 13 000 for 15 min at 4°C and the supernatant was transferred to a new 1.5 ml centrifuge tube. RNA was precitated in 600 µl pure ethanol and 30 µl 4M NaCl overnight at -20°C. RNA pellet was formed by centrifuging the tube at 13 000 rpm for 15 min at 4°C. RNA pellet was washed with 70% ethanol and dried completely. Finally, total RNA was dissolved in 30 µl DEPC-treated water. The quatity and quality of total RNA was measured by a NanoPhotometer® NP80 (Implen, Germany).

### 4.2.7. Synthesis of complementary DNA (cDNA)

cDNA synthesis was done using the RevertAidTM H Minus First Strand cDNA Synthesis kit (ThermoFisher Scientific, USA). 1 µg total RNA was mixed with 1 µl 0.5 µg/µl OligodT$_{18}$ Primer, 0.5 µl 0.2 µg/µl Random Hexamer Primer, and nuclease-free water to a final volume of 12.3 µl. Then, sample was incubated at 65°C for 5 min and transferred to ice immidiately. The mixture then was added with 4 µl 5X Reaction Buffer, 2 µl 10 mM dNTPs, 1 µl 20 U/µl RiboLock RNase Inhibitor and 0.7 µl 200 U/µl RevertAid H Minus M-MuLV Reverse Transcriptase. Next, the reverse transcription reaction was performed at 25°C for 5 min, then at 45°C for 60 min, and finally at 70°C for 5 min. The cDNA was stored at -20°C for further experiment.

### 4.2.8. Quantitative Real-Time PCR

The quantitative real-time PCR (qRT-PCR) reaction was performed in a 10-µl reaction consisting of 1 µl diluted cDNA (12.5 ng), 0.4 µl 5mM forward/reverse primers, and 5 µl 2X KAPA SYBR® FAST qPCR Master Mix Universal (KAPABIOSYSTEMS, USA). The qRT-PCR was conducted using a CFX Connect Real-Time PCR Detection System (Bio-Rad, USA). The relative expression level was calculated using method of (Pfaffl, 2001) in which three reference genes were used, including *AtACTIN2, AT5G46630,* and *AtPP2AA2*.

### 4.2.9. Reverse-transcriptase PCR

To validate alternative splicing event, reverse-transcriptase PCR (RT-PCR) with appropriate primers flanking the alternative splicing sites was performed. Each RT-PCR reaction comprised of 1 µl of cDNA (10 ng) as the template, 2 µl 10X *Taq* DNA Polymerase Buffer, 0.4 µl 25 mM MgCl$_2$, 1.2 µl forward/reverse primers, and 0.1 µl 10 U/µl *Taq* DNA Polymerase (EURx, Poland). The thermal program was as followed: initiation by 95°C for 3 min, following by 30-37 cycles of 95°C for 30 sec, 56°C for 15 sec, and 72°C for 30 sec, and last extension of 72°C for 2 min. The PCR product was visualized by electrophoresis on  a agarose gel. The intensity of PCR product bands were analyzed using using Fiji ImageJ (Schindelin et al., 2012).

### 4.2.10. RNA sequencing

To perform RNA-seq, total RNA was isolated using RNeasy Plant Mini Kit (QIAGEN, Netherlands). Then, RNA quantity and quality was measured using the Agilent RNA 6000 Pico Kit with Bioanalyzer 2100 system (Agilent Technologies, USA). Samples with high purity and RNA Integrity Number (RIN) then were sent to Novogene Co., Ltd (UK) for library preparation, RNA sequencing, and bioinformatics analysis according to their standard protocols. RNA-seq were performed using Illumina-based NovaSeq 6000 150 bp paired-end reads sequencing. Reads were checked with FastQC and cleaned up before mapping to the reference *A. thaliana* genome (TAIR10) by HISAT2. The number of reads of each gene was counted using featureCounts. A summary of the quality control of the libraries and RNA-seq samples is presented in the Appendix Table S9. Then, the fpkm value (fragments per kilobase of transcript per million mapped reads) was calculated as the relative expression level of a gene in the sample, in which gene is considered as expressed if the fpkm > 1. Differentially expressed genes (DEGs) between the mutant and the WT were determined by DESeq2 R package using raw read counts. A weak effect of the loss-of-function of WHY1 was expected as the mutant showed no apparent phenotype. Besides, the whole seedlings were used and genes, especially transcription factors, express temporally and spatially at low level in this early stage of development. For these reasons, DEGs were calculated using a low stringent cut-off of |log$_2$fold-change| > 1 and p-value < 0.01 in order to increase the number of DEGs. Furthermore, functional analysis of DEGs was performed using Gene Ontology (GO) enrichment analysis (Aleksander et al., 2023; Thomas et al., 2022; Ashburner et al., 2000), and then was summarized with REVIGO tool (Supek et al., 2011) to reduce functional redundancies, and consequently use Cytoscape for visualization (Shannon et al., 2003). To detect differential alternative splicing (AS) in RNA-seq data, rMATS (replicate multivariate analysis of transcript splicing)

tool was used to calculate the percentage of inclusion of alternative sequence, i.e., percent spliced in (PSI) and compare between samples (ΔPSI). Only gene with at least 10 reads in all samples and with the PSI value of an alternative sequence > 0.1 in at least one sample was included in the analysis. AS event which had the disprecancy in AS isoform ratio between two samples |ΔPSI| > 0.15 and p-value < 0.01 was considered significantly differentially alternative spliced (DAS).

### 4.2.11. Glucosinolates quantification

To determine GSL content, 3-5 mg of *A. thaliana* whole seedlings were ground in liquid nitrogen and tissue powder was subsequently homogenized with extraction solutions. For each sample, 450 µl 2:1 methanol:chloroform extraction solution, supplemented with 20 µg/ml sinigrin as the internal standard and 200 µl deionized water were added. Samples were incubated at room temperature on a shaking platform for 60 min, followed by centrifugation at 10 000 g for 20 min. The upper aqueous phase was used for GSL measurement via Liquid Chromatography coupled with tandem Mass Spectrometry (LC-MS/MS). The separation of substrates was performed with a Nucleoshell RP18 column (50 × 3 mm, particle size 2.7 µm; Macherey-Nagel, Germany) at 35°C using an Agilent 1290 High Pressure Liquid Chromatography (HPLC) system (Agilent, USA). The analytes were detected by Electrospray Ionization MS/MS (ESI-MS/MS) using an API 3200 triple-quadrupole LC-MS/MS system, equipped with an ESI Turbo Ion Spray interface, operated in the negative ion mode (AB Sciex, Germany). GSL-specific signals were acquired via Multiple Reaction Monitoring (MRM), with scan time of 15 msec. Peak areas were calculated automatically by IntelliQuant algorithm of the Analyst 1.6 software (AB Sciex, Germany), with manual adjustment when necessary. GSL content in each sample was calculated and normalized to the internal standard and fresh weight.

### 4.2.12. Statistical analysis

Physiological measurement, qRT-PCR, and GSL quantification approach consisted of three to four independent biological replicates. The mean of biological replicates was used to calculate standard deviation. Two-tail paired Student's t-test was performed to estimate the statistically significant difference between samples using Microsoft Excel. Asterisks indicate statistical significance with * for p-value < 0.05, ** for p-value < 0.01, and *** for p-value < 0.001.

# REFERENCES

Akbudak, M. A., & Filiz, E. (2019). Whirly (Why) transcription factors in tomato (*Solanum lycopersicum* L.): genome-wide identification and transcriptional profiling under drought and salt stresses. *Molecular Biology Reports*, 46(4), 4139–4150. https://doi.org/10.1007/s11033-019-04863-y

Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R, … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *GENETICS*, 224(1). https://doi.org/-10.1093/genetics/iyad031

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., … Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. https://doi.org/10.1038/75556

Barco, B., & Clay, N. K. (2019). Evolution of Glucosinolate Diversity via Whole-Genome Duplications, Gene Rearrangements, and Substrate Promiscuity. *Annual Review of Plant Biology*, 70, 585–604. https://doi.org/10.1146/annurev-arplant-050718

Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Bewley, J. D., Bradford, K. J., Hilhorst, H. W. M., & Nonogaki, H. (2013). Seeds (3rd ed.). *Springer*.

Butt, H., Bazin, J., Prasad, K. V. S. K., Awad, N., Crespi, M., Reddy, A. S. N., & Mahfouz, M. M. (2022). The Rice Serine/Arginine Splicing Factor RS33 Regulates Pre-mRNA Splicing during Abiotic Stress Responses. *Cells*, 11(11). https://doi.org/10.3390/cells11111796

Cai, Q., Guo, L., Shen, Z. R., Wang, D. Y., Zhang, Q., & Sodmergen. (2015). Elevation of pollen mitochondrial DNA copy number by WHIRLY2: Altered respiration and pollen tube growth in *Arabidopsis*. *Plant Physiology*, 169(1), 660–673. https://doi.org/10.1104/pp.15.00437

Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., Knight, H., Nimmo, H. G., Zhang, R., & Brown, J. W. S. (2018). Rapid and dynamic alternative splicing impacts the *Arabidopsis* cold response transcriptome. *Plant Cell*, 30(7), 1424–1444. https://doi.org/10.1105/tpc.18.00177

Cappadocia, L., Maréchal, A., Parent, J. S., Lepage, É., Sygusch, J., & Brisson, N. (2010). Crystal structures of DNA-whirly complexes and their role in *Arabidopsis* organelle genome repair. *Plant Cell*, 22(6), 1849–1867. https://doi.org/10.1105/tpc.109.071399

Cappadocia, L., Parent, J. S., Zampini, É., Lepage, É., Sygusch, J., & Brisson, N. (2012). A conserved lysine residue of plant Whirly proteins is necessary for higher order protein assembly and protection against DNA damage. *Nucleic Acids Research*, 40(1), 258–269. https://doi.org/10.1093/nar/gkr740

Cecchini, N. M., Torres, J. R., López, I. L., Cobo, S., Nota, F., & Alvarez, M. E. (2022). Alternative splicing of an exitron determines the subnuclear localization of the *Arabidopsis* DNA glycosylase MBD4L under heat stress. *Plant Journal*, 110(2), 377–388. https://doi.org/10.1111/tpj.15675

Chandan, R. K., Kumar, R., Swain, D. M., Ghosh, S., Bhagat, P. K., Patel, S., Bagler, G., Sinha, A. K., & Jha, G. (2023). RAV1 family members function as transcriptional regulators and play a positive role in plant disease resistance. *Plant Journal*. https://doi.org/10.1111/tpj.16114

Chen, L. Q., Chhajed, S., Zhang, T., Collins, J. M., Pang, Q., Song, W., He, Y., & Chen, S. (2021). Protein complex formation in methionine chain-elongation and leucine biosynthesis. *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-021-82790-4

Comadira, G., Rasool, B., Kaprinska, B., García, B. M., Morris, J., Verrall, S. R., Bayer, M., Hedley, P. E., Hancock, R. D., & Foyer, C. H. (2015). WHIRLY1 functions in the control of responses to nitrogen deficiency but not aphid infestation in barley. *Plant Physiology*, 168(3), 1140–1151. https://doi.org/10.1104/pp.15.00580

Corbineau, F., Xia, Q., Bailly, C., & El-Maarouf-Bouteau, H. (2014). Ethylene, a key factor in the regulation of seed dormancy. *Frontiers in Plant Science*, 5(OCT). https://doi.org/10.3389/-fpls.2014.00539

Cruz, T. M. D., Carvalho, R. F., Richardson, D. N., & Duque, P. (2014). Abscisic acid (ABA) regulation of *Arabidopsis* SR protein gene expression. *International Journal of Molecular Sciences*, 15(10), 17541–17564. https://doi.org/10.3390/ijms151017541

de Luxán-Hernández, C., Lohmann, J., Tranque, E., Chumova, J., Binarova, P., Salinas, J., & Weingartner, M. (2023). MDF is a conserved splicing factor and modulates cell division and stress response in *Arabidopsis*. *Life Science Alliance*, 6(1), e202201507. https://doi.org/-10.26508/lsa.202201507

Després, C., Subramaniam, R., Matton,', D. P., & Brisson, N. (1995). The Activation of the Potato PR-lOa Gene Requires the Phosphorylation of the Nuclear Factor PBFI. *The Plant Cell*, 7.

Desveaux, D., Allard, J., Brisson, N., & Sygusch, J. (2002). A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nature Structural Biology*, 9(7), 512–517. https://doi.org/10.1038/nsb814

Desveaux, D., Després, C., Joyeux, A., Subramaniam, R., & Brisson, N. (2000). PBF-2 Is a Novel Single-Stranded DNA Binding Factor Implicated in *PR-10a* Gene Activation in Potato. *The Plant Cell*, 12.

Desveaux, D., Maréchal, A., & Brisson, N. (2005). Whirly transcription factors: Defense gene regulation and beyond. *Trends in Plant Science* (Vol. 10, Issue 2, pp. 95–102). Elsevier Ltd. https://doi.org/10.1016/j.tplants.2004.12.008

Desveaux, D., Subramaniam, R., Despré, C., Mess, J.-N., Vesque, C. L., Fobert, P. R., Dangl, J. L., & Brisson, N. (2004). A "Whirly" Transcription Factor Is Required for Salicylic Acid-Dependent Disease Resistance in *Arabidopsis*. *Developmental Cell*, 6.

Dickey, T. H., Altschuler, S. E., & Wuttke, D. S. (2013). Single-stranded DNA-binding proteins: Multiple domains for multiple functions. *Structure*, 21(7), 1074–1084. https://doi.org/-10.1016/j.str.2013.05.013

Egawa, C., Kobayashi, F., Ishibashi, M., Nakamura, T., Nakamura, C., & Takumi, S. (2006). Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. *Genes Genetic System*, 81.

Fasani, E., DalCorso, G., Costa, A., Zenoni, S., & Furini, A. (2019). The *Arabidopsis thaliana* transcription factor MYB59 regulates calcium signalling during plant growth and stress response. *Plant Molecular Biology*, 99(6), 517–534. https://doi.org/10.1007/s11103-019-00833-x

Feng, C. Z., Chen, Y., Wang, C., Kong, Y. H., Wu, W. H., & Chen, Y. F. (2014). *Arabidopsis* RAV1 transcription factor, phosphorylated by SnRK2 kinases, regulates the expression of ABI3, ABI4, and ABI5 during seed germination and early seedling development. *Plant Journal*, 80(4), 654–668. https://doi.org/10.1111/tpj.12670

Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., Yan, S., Zhang, W., He, H., Ma, L., & Zhu, Z. (2015). SKIP Confers Osmotic Tolerance during Salt Stress by Controlling Alternative Gene Splicing in *Arabidopsis*. *Molecular Plant,* 8(7), 1038–1052. https://doi.org/10.1016/-j.molp.2015.01.011

Foyer, C. H., Karpinska, B., & Krupinska, K. (2014). The functions of WHIRLY1 and REDOXRESPONSIVE TRANSCRIPTION FACTOR 1 in cross tolerance responses in plants: A hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369 (1640). https://doi.org/10.1098/rstb.2013.0226

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., & Solano, R. (2014a). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), 2367–2372. https://doi.org/10.1073/pnas.1316278111

Fu, M., Kang, H. K., Son, S. H., Kim, S. K., & Nam, K. H. (2014). A subset of *Arabidopsis* RAV transcription factors modulates drought and salt stress responses independent of ABA. *Plant and Cell Physiology*, 55(11), 1892–1904. https://doi.org/10.1093/pcp/pcu118

Fulnečková, J., Dokládal, L., Kolářová, K., Dadejová, M. N., Procházková, K., Gomelská, S., Sivčák, M., Adamusová, K., Lyčka, M., Peska, V., Dvořáčková, M., & Sýkorová, E. (2022). Telomerase Interaction Partners-Insight from Plants. *International Journal of Molecular Sciences*, 23(1), 368. https://doi.org/10.3390/ijms

Gallie, D. R., & Chen, Z. (2019). Chloroplast-localized iron superoxide dismutases FSD2 and FSD3 are functionally distinct in *Arabidopsis*. *PLoS ONE*, 14(7). https://doi.org/10.1371/-journal.pone.0220078

Gangappa, S. N., & Botto, J. F. (2016). The Multifaceted Roles of HY5 in Plant Growth and Development. *Molecular Plant*, 9(10), 1353–1365. https://doi.org/10.1016/j.molp.2016.07.002

Garavís, M., & Calvo, O. (2017). Sub1/PC4, a multifaceted factor: from transcription to genome stability. *Current Genetics*, 63(6), 1023–1035. https://doi.org/10.1007/s00294-017-0715-6

Golin, S., Negroni, Y. L., Bennewitz, B., Klösgen, R. B., Mulisch, M., La Rocca, N., Cantele, F., Vigani, G., Lo Schiavo, F., Krupinska, K., & Zottini, M. (2020). WHIRLY2 plays a key role in mitochondria morphology, dynamics, and functionality in *Arabidopsis thaliana. Plant Direct*, 4(5). https://doi.org/10.1002/pld3.229

Gupta, R., Sadhale, P. P., & Vijayraghavan, U. (2015). SUB1 plays a negative role during starvation induced sporulation program in *Saccharomyces cerevisiae. PLoS ONE*, 10(7). https://doi.org/10.1371/journal.pone.0132350

Hartmann, L., Drewe-Boß, P., Wießner, T., Wagner, G., Geue, S., Lee, H.-C., Obermüller, D. M., Kahles, A., Behr, J., Sinz, F. H., Rätsch, G., & Wachter, A. (2016). Alternative Splicing Substantially Diversifies the Transcriptome during Early Photomorphogenesis and Correlates with the Energy Availability in *Arabidopsis. The Plant Cell*, 28(11), 2715–2734. https://doi.org/10.1105/tpc.16.00508

Hauvermale, A. L., & Steber, C. M. (2020). GA signaling is essential for the embryo-to-seedling transition during *Arabidopsis* seed germination, a ghost story. *Plant Signaling and Behavior*, 15(1). https://doi.org/10.1080/15592324.2019.1705028

He, H., Van Breusegem, F., & Mhamdi, A. (2018). Redox-dependent control of nuclear transcription in plants. *Journal of Experimental Botany*, 69(14), 3359–3372. https://doi.org/10.1093/jxb/ery130

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., & Zimmermann, P. (2008). Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. *Advances in Bioinformatics*, 2008, 1–5. https://doi.org/10.1155/2008/420747

Hu, Y., & Shu, B. (2021). Identifying Strawberry Whirly Family Transcription Factors and their Expressions in Response to Crown Rot. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 49(2), 1–11. https://doi.org/10.15835/nbha49212323

Hu, Y., Vandenbussche, F., & Van Der Straeten, D. (2017). Regulation of seedling growth by ethylene and the ethylene–auxin crosstalk. *Planta*, 245(3), 467–489. https://doi.org/10.1007/-s00425-017-2651-6

Huang, C., Yu, J., Cai, Q., Chen, Y., Li, Y., Ren, Y., & Miao, Y. (2020). Triple-localized WHIRLY2 influences leaf senescence and silique development via carbon allocation. *Plant Physiology*, 184(3), 1348–1362. https://doi.org/10.1104/pp.20.00832

Huang, D., Lan, W., Li, D., Deng, B., Lin, W., Ren, Y., & Miao, Y. (2018). WHIRLY1 occupancy affects histone lysine modification and WRKY53 transcription in *Arabidopsis* developmental manner. *Frontiers in Plant Science*, 871. https://doi.org/10.3389/fpls.2018.01503

Huang, D., Lan, W., Ma, W., Huang, R., Lin, W., Li, M., Chen, C. Y., Wu, K., & Miao, Y. (2022). WHIRLY1 recruits the histone deacetylase HDA15 repressing leaf senescence and flowering in *Arabidopsis. Journal of Integrative Plant Biology*, 64(7), 1411–1429. https://doi.org/10.1111/-jipb.13272

Huang, D., Lin, W., Deng, B., Ren, Y., & Miao, Y. (2017). Dual-located WHIRLY1 interacting with LHCA1 alters photochemical activities of photosystem i and is involved in light adaptation in *Arabidopsis. International Journal of Molecular Sciences*, 18(11). https://doi.org/10.3390/-ijms18112352

Huang, H., Ullah, F., Zhou, D. X., Yi, M., & Zhao, Y. (2019a). Mechanisms of ROS regulation of plant development and stress responses. *Frontiers in Plant Science*, 10,. https://doi.org/-10.3389/fpls.2019.00800

Huang, W., Chen, X., Guan, Q., Zhong, Z., Ma, J., Yang, B., Wang, T., Zhu, W., & Tian, J. (2019b). Changes of alternative splicing in *Arabidopsis thaliana* grown under different CO2 concentrations. *Gene*, 689, 43–50. https://doi.org/10.1016/j.gene.2018.11.083

Isemer, R., Krause, K., Grabe, N., Kitahata, N., Asami, T., & Krupinska, K. (2012a). Plastid located WHIRLY1 enhances the responsiveness of *Arabidopsis* seedlings toward abscisic acid. *Frontiers in Plant Science*, 3(DEC). https://doi.org/10.3389/fpls.2012.00283

Isemer, R., Mulisch, M., Schäfer, A., Kirchner, S., Koop, H. U., & Krupinska, K. (2012b). Recombinant Whirly1 translocates from transplastomic chloroplasts to the nucleus. *FEBS Letters*, 586(1), 85–88. https://doi.org/10.1016/j.febslet.2011.11.029

Jabre, I., Reddy, A. S. N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L. J., Wilson, C. M., & Syed, N. H. (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Research*, 47(6), 2716–2726. https://doi.org/10.1093/nar/gkz121

Janack, B., Sosoi, P., Krupinska, K., & Humbeck, K. (2016). Knockdown of WHIRLY1 affects drought stress-induced leaf senescence and histone modifications of the senescence-associated gene *HvS40*. *Plants*, 5(3). https://doi.org/10.3390/plants5030037

Janicka, S., Kühn, K., Le Ret, M., Bonnard, G., Imbault, P., Augustyniak, H., & Gualberto, J. M. (2012). A RAD52-like single-stranded DNA binding protein affects mitochondrial DNA repair by recombination. *Plant Journal*, 72(3), 423–435. https://doi.org/10.1111/j.1365-313X.2012.05097.x

Jiang, J., Liu, X., Liu, C., Liu, G., Li, S., & Wang, L. (2017a). Integrating Omics and Alternative Splicing Reveals Insights into Grape Response to High Temperature. *Plant Physiology*, 173(2), 1502–1518. https://doi.org/10.1104/pp.16.01305

Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J., & Zhang, J. (2017b). WRKY transcription factors in plant responses to stresses. *Journal of Integrative Plant Biology*, 59(2), 86–101. https://doi.org/10.1111/jipb.12513

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kagaya, Y., Ohmiya, K., & Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Research*, 27(2). https://academic.oup.com/nar/article/27/2/470/1058626

Karpinska, B., Razak, N., James, E. K., Morris, J. A., Verrall, S. R., Hedley, P. E., Hancock, R. D., & Foyer, C. H. (2022). WHIRLY1 functions in the nucleus to regulate barley leaf development and associated metabolite profiles. *Biochemical Journal*, 479(5), 641–659. https://doi.org/10.1042/BCJ20210810

Kashkan, I., Timofeyenko, K., & Růžička, K. (2022). How alternative splicing changes the properties of plant proteins. *Quantitative Plant Biology*, 3. https://doi.org/10.1017/qpb.2022.9

Kathare, P. K., & Huq, E. (2021). Light-regulated pre-mRNA splicing in plants. *Current Opinion in Plant Biology*, 63. https://doi.org/10.1016/j.pbi.2021.102037

Kim, C., Meskauskiene, R., Zhang, S., Lee, K. P., Ashok, M. L., Blajecka, K., Herrfurth, C., Feussner, I., & Apela, K. (2012). Chloroplasts of *Arabidopsis* are the source and a primary target of a plant-specific programmed cell death signaling pathway. *Plant Cell*, 24(7), 3026–3039. https://doi.org/10.1105/tpc.112.100479

Kim, J. Y., Song, J. T., & Seo, H. S. (2017). COP1 regulates plant growth and development in response to light at the post-translational level. *Journal of Experimental Botany*, 68(17), 4737–4748. https://doi.org/10.1093/jxb/erx312

Krause, K., Kilbienski, I., Mulisch, M., Rödiger, A., Schäfer, A., & Krupinska, K. (2005). DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS Letters*, 579(17), 3707–3712. https://doi.org/10.1016/j.febslet.2005.05.059

Kretschmer, M., Croll, D., & Kronstad, J. W. (2017). Chloroplast-associated metabolic functions influence the susceptibility of maize to *Ustilago maydis*. *Molecular Plant Pathology*, 18(9), 1210–1221. https://doi.org/10.1111/mpp.12485

Krupinska, K., Blanco, N. E., Oetke, S., & Zottini, M. (2020). Genome communication in plants mediated by organelle–nucleus-located proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1801). https://doi.org/10.1098/rstb.2019.0397

Krupinska, K., Braun, S., Nia, M. S., Schäfer, A., Hensel, G., & Bilger, W. (2019). The nucleoid-associated protein WHIRLY1 is required for the coordinate assembly of plastid and nucleus-encoded proteins during chloroplast development. *Planta*, 249(5), 1337–1347. https://doi.org/10.1007/s00425-018-03085-z

Krupinska, K., Dähnhardt, D., Fischer-Kilbienski, I., Kucharewicz, W., Scharrenberg, C., Trösch, M., & Buck, F. (2014a). Identification of WHIRLY1 as a Factor Binding to the Promoter of the Stress- and Senescence-Associated Gene *HvS40*. *Journal of Plant Growth Regulation*, 33(1), 91–105. https://doi.org/10.1007/s00344-013-9378-9

Krupinska, K., Desel, C., Frank, S., & Hensel, G. (2022). WHIRLIES Are Multifunctional DNA-Binding Proteins With Impact on Plant Development and Stress Resistance. *Frontiers in Plant Science*, 13. https://doi.org/10.3389/fpls.2022.880423

Krupinska, K., Oetke, S., Desel, C., Mulisch, M., Schäfer, A., Hollmann, J., Kumlehn, J., & Hensel, G. (2014b). WHIRLY1 is a major organizer of chloroplast nucleoids. *Frontiers in Plant Science*, 5. https://doi.org/10.3389/fpls.2014.00432

Kucharewicz, W., Distelfeld, A., Bilger, W., Müller, M., Munné-Bosch, S., Hensel, G., & Krupinska, K. (2017). Acceleration of leaf senescence is slowed down in transgenic barley plants deficient in the DNA/RNA-binding protein WHIRLY1. *Journal of Experimental Botany*, 68(5), 983–996. https://doi.org/10.1093/jxb/erw501

Laloum, T., Martín, G., & Duque, P. (2018). Alternative Splicing Control of Abiotic Stress Responses. *Trends in Plant Science*, 23(2), 140–150. https://doi.org/10.1016/j.tplants.2017.09.019

Lau, O. S., & Deng, X. W. (2010). Plant hormone signaling lightens up: integrators of light and hormones. *Current Opinion in Plant Biology*, 13(5), 571–577. https://doi.org/10.1016/j.pbi.2010.07.001

Lee, B., Kapoor, A., Zhu, J., & Zhu, J.-K. (2006). STABILIZED1, a Stress-Upregulated Nuclear Protein, Is Required for Pre-mRNA Splicing, mRNA Turnover, and Stress Tolerance in *Arabidopsis*. *The Plant Cell*, 18(7), 1736–1749. https://doi.org/10.1105/tpc.106.042184

Leivar, P., & Monte, E. (2014). PIFs: Systems integrators in plant development. *Plant Cell*, 26(1), 56–78. https://doi.org/10.1105/tpc.113.120857

Li, C., Zheng, L., Zhang, J., Lv, Y., Liu, J., Wang, X., Palfalvi, G., Wang, G., & Zhang, Y. (2017a). Characterization and functional analysis of four HYH splicing variants in *Arabidopsis* hypocotyl elongation. *Gene*, 619, 44–49. https://doi.org/10.1016/j.gene.2017.04.001

Li, M., & Kim, C. (2022). Chloroplast ROS and stress signaling. *Plant Communications*, 3(1), 100264. https://doi.org/10.1016/j.xplc.2021.100264

Li, Q., Ai, G., Shen, D., Zou, F., Wang, J., Bai, T., Chen, Y., Li, S., Zhang, M., Jing, M., & Dou, D. (2019). A *Phytophthora capsici* Effector Targets ACD11 Binding Partners that Regulate ROS-Mediated Defense Response in *Arabidopsis*. *Molecular Plant*, 12(4), 565–581. https://doi.org/10.1016/j.molp.2019.01.018

Li, Q.-F., & He, J.-X. (2016). BZR1 Interacts with HY5 to Mediate Brassinosteroid- and Light-Regulated Cotyledon Opening in *Arabidopsis* in Darkness. *Molecular Plant*, 9(1), 113–125. https://doi.org/10.1016/j.molp.2015.08.014

Li, R., Wang, W., Li, F., Wang, Q., Wang, S., Xu, Y., & Chen, F. (2017b). Response of alternative splice isoforms of *OsRad9* gene from *Oryza sativa* to environmental stress. *Zeitschrift Fur Naturforschung - Section C Journal of Biosciences*, 72(7–8), 325–334. https://doi.org/10.1515/znc-2016-0257

Li, Y., Guo, Q., Liu, P., Huang, J., Zhang, S., Yang, G., Wu, C., Zheng, C., & Yan, K. (2021). Dual roles of the serine/arginine-rich splicing factor SR45a in promoting and interacting with nuclear cap-binding complex to modulate the salt-stress response in *Arabidopsis*. *New Phytologist*, 230(2), 641–655. https://doi.org/10.1111/nph.17175

Li, Z., Jaroszewski, L., Iyer, M., Sedova, M., & Godzik, A. (2020). FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research*, 48(W1), W60–W64. https://doi.org/10.1093/nar/gkaa443

Lin, W., Huang, D., Shi, X., Deng, B., Ren, Y., Lin, W., & Miao, Y. (2019). $H_2O_2$ as a feedback signal on dual-located WHIRLY1 associates with leaf senescence in *Arabidopsis*. *Cells*, 8(12). https://doi.org/10.3390/cells8121585

Lin, W., Zhang, H., Huang, D., Schenke, D., Cai, D., Wu, B., & Miao, Y. (2020). Dual-localized WHIRLY1 affects salicylic acid biosynthesis via coordination of isochorismate synthase1, phenylalanine ammonia lyase1, and S-adenosyl-L-methionine-dependent methyltransferase1. *Plant Physiology*, 184(4), 1884–1899. https://doi.org/10.1104/pp.20.00964

Liu, J., Sun, N., Liu, M., Liu, J., Du, B., Wang, X., & Qi, X. (2013). An autoregulatory loop controlling *Arabidopsis* HsfA2 expression: Role of heat shock-induced alternative splicing. *Plant Physiology*, 162(1), 512–521. https://doi.org/10.1104/pp.112.205864

Liu, W., Yan, Y., Zeng, H., Li, X., Wei, Y., Liu, G., He, C., & Shi, H. (2018). Functional characterization of WHY–WRKY75 transcriptional module in plant response to cassava bacterial blight. *Tree Physiology*, 38(10), 1502–1512. https://doi.org/10.1093/treephys/tpy053

Liu, X. X., Guo, Q. H., Xu, W. B., Liu, P., & Yan, K. (2022). Rapid Regulation of Alternative Splicing in Response to Environmental Stresses. *Frontiers in Plant Science*, 13. https://doi.org/-10.3389/fpls.2022.832177

Luján-Soto, E., & Dinkova, T. D. (2021). Time to wake up: Epigenetic and small-RNA-mediated regulation during seed germination. *Plants*, 10(2), 1–19. https://doi.org/10.3390/plants10020236

Luo, M., Luo, H., Li, H., Liu, D., & Lu, H. (2020). Single-stranded DNA-binding proteins in plant telomeres. *International Journal of Biological Macromolecules*, 165, 1463–1467. https://doi.org/10.1016/j.ijbiomac.2020.09.211

Mandal, D., Datta, S., Raveendar, G., Mondal, P. K., & Nag Chaudhuri, R. (2023). RAV1 mediates cytokinin signaling for regulating primary root growth in *Arabidopsis*. *Plant Journal*, 113(1), 106–126. https://doi.org/10.1111/tpj.16039

Manh, M. B., Ost, C., Peiter, E., Hause, B., Krupinska, K., & Humbeck, K. (2023). WHIRLY1 Acts Upstream of ABA-Related Reprogramming of Drought-Induced Gene Expression in Barley and Affects Stress-Related Histone Modifications. *International Journal of Molecular Sciences*, 24(7), 6326. https://doi.org/10.3390/ijms24076326

Maréchal, A., & Brisson, N. (2010). Recombination and the maintenance of plant organelle genome stability. *New Phytologist*, 186(2), 299–317. https://doi.org/10.1111/j.1469-8137.2010.03195.x

Maréchal, A., Parent, J. S., Sabar, M., Véronneau-Lafortune, F., Abou-Rached, C., & Brisson, N. (2008). Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biology*, 8. https://doi.org/10.1186/1471-2229-8-42

Maréchal, A., Parent, J. S., Véronneau-Lafortune, F., Joyeux, A., Lang, B. F., & Brisson, N. (2009). Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34), 14693–14698. https://doi.org/10.1073/pnas.0901710106

Martín, G., Márquez, Y., Mantica, F., Duque, P., & Irimia, M. (2021). Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biology*, 22(1). https://doi.org/10.1186/s13059-020-02258-y

Mayr, C. (2019). What are 3′ UTRs doing? *Cold Spring Harbor Perspectives in Biology*, 11(10). https://doi.org/10.1101/cshperspect.a034728

McCoy, R. M., Julian, R., Kumar, S. R. V., Ranjan, R., Varala, K., & Li, Y. (2021). A systems biology approach to identify essential epigenetic regulators for specific biological processes in plants. *Plants*, 10(2), 1–16. https://doi.org/10.3390/plants10020364

Meier, K., Ehbrecht, M. D., & Wittstock, U. (2019). Glucosinolate content in dormant and germinating *Arabidopsis thaliana* seeds is affected by non-functional alleles of classical myrosinase and nitrile-specifier protein genes. *Frontiers in Plant Science*, 10. https://doi.org/10.3389/fpls.2019.01549

Melonek, J., Mulisch, M., Schmitz-Linneweber, C., Grabowski, E., Hensel, G., & Krupinska, K. (2010). Whirly1 in chloroplasts associates with intron containing RNAs and rarely co-localizes with nucleoids. *Planta*, 232(2), 471–481. https://doi.org/10.1007/s00425-010-1183-0

Meng, C., Yang, M., Wang, Y., Chen, C., Sui, N., Meng, Q., Zhuang, K., & Lv, W. (2020). SlWHY2 interacts with SlRECA2 to maintain mitochondrial function under drought stress in tomato. *Plant Science*, 301. https://doi.org/10.1016/j.plantsci.2020.110674

Mérillon, J.-M., & Ramawat, · K G. (2017). Glucosinolates. *Springer Cham.* https://doi.org/10.1007/-978-3-319-25462-3

Miao, Y., Jiang, J., Ren, Y., & Zhao, Z. (2013). The single-stranded DNA-binding protein WHIRLY1 represses WRKY53 expression and delays leaf senescence in a developmental stage-dependent manner in *Arabidopsis*. *Plant Physiology*, 163(2), 746–756. https://doi.org/10.1104/-pp.113.223412

Min, J.-H., Park, C.-R., Jang, Y.-H., Ju, H.-W., Lee, K.-H., Lee, S., & Kim, C. S. (2019). A basic helix-loop-helix 104 (bHLH104) protein functions as a transcriptional repressor for glucose and abscisic acid signaling in *Arabidopsis*. *Plant Physiology and Biochemistry*, 136, 34–42. https://doi.org/10.1016/j.plaphy.2019.01.008

Mitreiter, S., & Gigolashvili, T. (2021). Regulation of glucosinolate biosynthesis. *Journal of Experimental Botany*, 72(1), 70–91. https://doi.org/10.1093/jxb/eraa479

Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y. V., Liew, L. C., Lister, R., Lewsey, M. G., & Whelan, J. (2017). Extensive transcriptomic and epigenomic remodelling occurs during *Arabidopsis thaliana* germination. *Genome Biology*, 18(1). https://doi.org/10.1186/s13059-017-1302-3

Oetke, S., Scheidig, A. J., & Krupinska, K. (2022). WHIRLY1 of barley and maize share a PRAPP motif conferring nucleoid compaction. *Plant and Cell Physiology*, 63(2), 234–247.

Oh, E., Zhu, J. Y., Bai, M. Y., Arenhart, R. A., Sun, Y., & Wang, Z. Y. (2014). Cell elongation is regulated through a central circuit of interacting transcription factors in the *Arabidopsis* hypocotyl. *ELife*, 2014(3). https://doi.org/10.7554/eLife.03031

Ohta, M., Sato, A., Renhu, N., Yamamoto, T., Oka, N., Zhu, J.-K., Tada, Y., Suzaki, T., & Miura, K. (2018). MYC-type transcription factors, MYC67 and MYC70, interact with ICE1 and negatively regulate cold tolerance in *Arabidopsis*. *Scientific Reports*, 8(1), 11622. https://doi.org/10.1038/s41598-018-29722-x

O'Malley, R. C., Huang, S. S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., & Ecker, J. R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*, 165(5), 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Palusa, S. G., Ali, G. S., & Reddy, A. S. N. (2007). Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: Regulation by hormones and stresses. *Plant Journal*, 49(6), 1091–1107. https://doi.org/10.1111/j.1365-313X.2006.03020.x

Parasyri, A., Barth, O., Zschiesche, W., & Humbeck, K. (2022). The Barley Heavy Metal Associated Isoprenylated Plant Protein HvFP1 Is Involved in a Crosstalk between the Leaf Development and Abscisic Acid-Related Drought Stress Responses. *Plants*, 11(21), 2851. https://doi.org/10.3390/plants11212851

Petrillo, E., Godoy Herz, M. A., Fuchs, A., Reifer, D., Fuller, J., Yanovsky, M. J., Simpson, C., Brown, J. W. S., Barta, A., Kalyna, M., & Kornblihtt, A. R. (2014). A chloroplast retrograde signal regulates nuclear alternative splicing. *Science*, 344(6182), 427–430. https://doi.org/10.1126/science.1250322

Petrov, V., Vermeirssen, V., De Clercq, I., Van Breusegem, F., Minkov, I., Vandepoele, K., & Gechev, T. S. (2012). Identification of cis-regulatory elements specific for different types of reactive oxygen species in *Arabidopsis thaliana*. *Gene*, 499(1), 52–60. https://doi.org/10.1016/j.gene.2012.02.035

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9).

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K. J., & Oelmüller, R. (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell*, 18(1), 176–197. https://doi.org/10.1105/tpc.105.036392

Prikryl, J., Watkins, K. P., Friso, G., van Wijk, K. J., & Barkan, A. (2008). A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Research*, 36(16), 5152–5165. https://doi.org/10.1093/nar/gkn492

Qiu, Z., Chen, D., Teng, L., Guan, P., Yu, G., Zhang, P., Song, J., Zeng, Q., & Zhu, L. (2022). OsWHY1 Interacts with OsTRX z and is Essential for Early Chloroplast Development in Rice. *Rice*, 15(1). https://doi.org/10.1186/s12284-022-00596-y

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., & Job, D. (2004). The effect of α-amanitin on the *Arabidopsis* seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiology*, 134(4), 1598–1613. https://doi.org/10.1104/pp.103.036293

Reddy, A. S. N. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review of Plant Biology*, 58, 267–294. https://doi.org/10.1146/annurev.arplant.58.032806.103754

Ren, Y., Li, M., Wang, W., Lan, W., Schenke, D., Cai, D., & Miao, Y. (2022). MicroRNA840 ( MIR840 ) accelerates leaf senescence by targeting the overlapping 3′UTRs of PPR and WHIRLY3 in *Arabidopsis thaliana*. *The Plant Journal*, 109(1), 126–143. https://doi.org/10.1111/tpj.15559

Ren, Y., Li, Y., Jiang, Y., Wu, B., & Miao, Y. (2017). Phosphorylation of WHIRLY1 by CIPK14 Shifts Its Localization and Dual Functions in *Arabidopsis*. *Molecular Plant*, 10(5), 749–763. https://doi.org/10.1016/j.molp.2017.03.011

Rosenwasser, S., Fluhr, R., Joshi, J. R., Leviatan, N., Sela, N., Hetzroni, A., & Friedman, H. (2013). ROSMETER: A bioinformatic tool for the identification of transcriptomic imprints related to reactive oxygen species type and origin provides new insights into stress responses. *Plant Physiology*, 163(2), 1071–1083. https://doi.org/10.1104/pp.113.218206

Ruan, Q., Wang, Y., Xu, H., Wang, B., Zhu, X., Wei, B., & Wei, X. (2022). Genome-wide identification, phylogenetic, and expression analysis under abiotic stress conditions of Whirly (WHY) gene family in *Medicago sativa* L. *Scientific Reports*, 12(1). https://doi.org/10.1038/s41598-022-22658-3

Sano, N., Rajjou, L., & North, H. M. (2020). Lost in translation: Physiological roles of stored mRNAs in seed germination. *Plants*, 9(3). https://doi.org/10.3390/plants9030347

Santiago, M. J., Alejandre-Durán, E., & Ruiz-Rubio, M. (2009). Alternative splicing of two translesion synthesis DNA polymerases from *Arabidopsis thaliana*. *Plant Science*, 176(4), 591–596. https://doi.org/10.1016/j.plantsci.2009.01.018

Sanyal, S. K., Kanwar, P., Samtani, H., Kaur, K., Jha, S. K., & Pandey, G. K. (2017). Alternative splicing of CIPK3 results in distinct target selection to propagate ABA signaling in *Arabidopsis*. *Frontiers in Plant Science*, 8. https://doi.org/10.3389/fpls.2017.01924

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682. https://doi.org/10.1038/nmeth.2019

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shikata, H., Hanada, K., Ushijima, T., Nakashima, M., Suzuki, Y., & Matsushita, T. (2014). Phytochrome controls alternative splicing to mediate light responses in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 111(52), 18781–18786. https://doi.org/10.1073/pnas.1407147112

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2–3), 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X

Sohn, K. H., Lee, S. C., Jung, H. W., Hong, J. K., & Hwang, B. K. (2006). Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. *Plant Molecular Biology*, 61(6), 897–915. https://doi.org/10.1007/s11103-006-0057-0

Sønderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J., & Halkier, B. A. (2010). A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in *Arabidopsis*. *Plant Physiology*, 153(1), 348–363. https://doi.org/10.1104/pp.109.149286

Sønderby, I. E., Hansen, B. G., Bjarnholt, N., Ticconi, C., Halkier, B. A., & Kliebenstein, D. J. (2007). A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS ONE*, 2(12). https://doi.org/10.1371/journal.pone.0001322

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, 6(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Taylor, R. E., Waterworth, W., West, C. E., & Foyer, C. H. (2023). WHIRLY proteins maintain seed longevity by effects on seed oxygen signalling during imbibition. *Biochemical Journal*, 480(13), 941–956. https://doi.org/10.1042/BCJ20230008

Taylor, R. E., West, C. E., & Foyer, C. H. (2022). WHIRLY protein functions in plants. *Food and Energy Security*. https://doi.org/10.1002/fes3.379

Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L., & Mi, H. (2022). PANTHER : Making genome-scale phylogenetics accessible to all. *Protein Science*, 31(1), 8–22. https://doi.org/10.1002/pro.4218

Trigg, S. A., Garza, R. M., MacWilliams, A., Nery, J. R., Bartlett, A., Castanon, R., Goubil, A., Feeney, J., O'Malley, R., Huang, S. S. C., Zhang, Z. Z., Galli, M., & Ecker, J. R. (2017). CrY2H-seq: A massively multiplexed assay for deep-coverage interactome mapping. *Nature Methods*, 14(8), 819–825. https://doi.org/10.1038/nmeth.4343

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., … Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), D439–D444. https://doi.org/10.1093/-nar/gkab1061

Wang, W., Li, K., Yang, Z., Hou, Q., Zhao, W. W., & Sun, Q. (2021). RNase H1C collaborates with ssDNA binding proteins WHY1/3 and recombinase RecA1 to fulfill the DNA damage repair in *Arabidopsis* chloroplasts. *Nucleic Acids Research*, 49(12), 6771–6787. https://doi.org/10.1093/-nar/gkab479

Wang, Y., Li, L., Ye, T., Zhao, S., Liu, Z., Feng, Y. Q., & Wu, Y. (2011). Cytokinin antagonizes ABA suppression to seed germination of *Arabidopsis* by downregulating ABI5 expression. *Plant Journal*, 68(2), 249–261. https://doi.org/10.1111/j.1365-313X.2011.04683.x

Waterworth, W. M., Bray, C. M., & West, C. E. (2015). The importance of safeguarding genome integrity in germination and seed longevity. *Journal of Experimental Botany*, 66(12), 3549–3558. https://doi.org/10.1093/jxb/erv080

Willems, P., Mhamdi, A., Stael, S., Storme, V., Kerchev, P., Noctor, G., Gevaert, K., & Van Breusegem, F. (2016). The ROS wheel: Refining ROS transcriptional footprints. *Plant Physiology*, 171(3), 1720–1733. https://doi.org/10.1104/pp.16.00420

Woo, H. R., Kim, J. H., Kim, J., Kim, J., Lee, U., Song, I. J., Kim, J. H., Lee, H. Y., Nam, H. G., & Lim, P. O. (2010). The RAV1 transcription factor positively regulates leaf senescence in *Arabidopsis*. *Journal of Experimental Botany*, 61(14), 3947–3957. https://doi.org/10.1093/-jxb/erq206

Xiao, Y., Chu, L., Zhang, Y., Bian, Y., Xiao, J., & Xu, D. (2022). HY5: A Pivotal Regulator of Light-Dependent Development in Higher Plants. *Frontiers in Plant Science*, 12. https://doi.org/10.3389/fpls.2021.800989

Xiong, J. Y., Lai, C. X., Qu, Z., Yang, X. Y., Qin, X. H., & Liu, G. Q. (2009). Recruitment of AtWHY1 and AtWHY3 by a distal element upstream of the kinesin gene AtKP1 to mediate transcriptional repression. *Plant Molecular Biology*, 71(4–5), 437–449. https://doi.org/10.1007/s11103-009-9533-7

Yadukrishnan, P., & Datta, S. (2021). Light and abscisic acid interplay in early seedling development. *New Phytologist*, 229(2), 763–769. https://doi.org/10.1111/nph.16963

Yan, Y., Liu, W., Wei, Y., & Shi, H. (2020). MeCIPK23 interacts with Whirly transcription factors to activate abscisic acid biosynthesis and regulate drought resistance in cassava. *Plant Biotechnology Journal*, 18(7), 1504–1506. https://doi.org/10.1111/pbi.13321

Yang, B., Song, Z., Li, C., Jiang, J., Zhou, Y., Wang, R., Wang, Q., Ni, C., Liang, Q., Chen, H., & Fan, L.-M. (2018). RSM1, an *Arabidopsis* MYB protein, interacts with HY5/HYH to modulate seed germination and seedling development in response to abscisic acid and salinity. *PLOS Genetics*, 14(12), e1007839. https://doi.org/10.1371/journal.pgen.1007839

Yoo, H. H., Kwon, C., Lee, M. M., & Chung, I. K. (2007). Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in *Arabidopsis*. *Plant Journal*, 49(3), 442–451. https://doi.org/10.1111/j.1365-313X.2006.02974.x

Zampini, É., Lepage, É., Tremblay-Belzile, S., Truche, S., & Brisson, N. (2015). Organelle DNA rearrangement mapping reveals U-turn-like inversions as a major source of genomic instability in *Arabidopsis* and humans. *Genome Research*, 125(5), 645–654. https://doi.org/10.1101/-gr.188573.114

Zhang, W., Du, B., Liu, D., & Qi, X. (2014). Splicing factor SR34b mutation reduces cadmium tolerance in *Arabidopsis* by regulating iron-regulated transporter 1 gene. *Biochemical and*

*Biophysical Research Communications*, 455(3–4), 312–317. https://doi.org/10.1016/j.bbrc.-2014.11.017

Zhao, S. Y., Wang, G. D., Zhao, W. Y., Zhang, S., Kong, F. Y., Dong, X. C., & Meng, Q. W. (2018). Overexpression of tomato WHIRLY protein enhances tolerance to drought stress and resistance to *Pseudomonas solanacearum* in transgenic tobacco. *Biologia Plantarum*, 62(1), 55–68. https://doi.org/10.1007/s10535-017-0714-y

Zhou, D.-X., Kim, Y.-J., Li, Y.-F., Carol, P., & Mache, R. (1998). COP1b, an isoform of COP1 generated by alternative splicing, has a negative effect on COP1 function in regulating light-dependent seedling development in *Arabidopsis*. *Molecular and General Genetics*, 257(4), 387–391. https://doi.org/10.1007/s004380050662

Zhou, Z., Windhorst, A., Schenke, D., & Cai, D. (2023). RNAseq-Based Working Model for Transcriptional Regulation of Crosstalk between Simultaneous Abiotic UV-B and Biotic Stresses in Plants. *Genes*, 14(2). https://doi.org/10.3390/genes14020240

Zhuang, K., Kong, F., Zhang, S., Meng, C., Yang, M., Liu, Z., Wang, Y., Ma, N., & Meng, Q. (2019). Whirly1 enhances tolerance to chilling stress in tomato via protection of photosystem II and regulation of starch degradation. *New Phytologist*, 221(4), 1998–2012. https://doi.org/10.1111/-nph.15532

Zhuang, K., Wang, J., Jiao, B., Chen, C., Zhang, J., Ma, N., & Meng, Q. (2020). WHIRLY1 maintains leaf photosynthetic capacity in tomato by regulating the expression of RbcS1 under chilling stress. *Journal of Experimental Botany*, 71(12), 3653–3663. https://doi.org/10.1093/jxb/eraa145

# APPENDIX

```
                    10        20        30        40        50        60        70        80        90       100
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtWHY1    MSQLLSTPLMAVN---------SNPRFLSSSSVLVTGGFAVKRHGFALKPTTKTVKLFSVKSRQTDYFEKQRFGDSSSSPSPAEGLPARFYVGHSIYKGK
AtWHY3    ......S.P...FSKTFINHKF.DA.....H.I.TS....G.--IIP....AR-LKLT-.....S..............QN-..VSSP............
AtWHY2    ------------------------MMKQARS.LSRSLCDQSKSLFEAS.LRGFASW.NS.TPGRG.P----.KDAAK..------G.LFAPY..F...

                   110       120       130       140       150       160       170       180       190       200
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtWHY1    AALTVDPRAPEFVALDSGAFKLSKDGFLLLQFAPSAGVRQYDWSKKQVFSLSVTEIGTLVSLGPRESCEFFHDPFKGK-SDEGKVRKVLKVEPLPDGSGH
AtWHY3    ....IE.........E......T.E.........A........R...........N....................G...................R
AtWHY2    ...S.E.VL.S.TEI...NLRIDRR.S.MMT.M.AI.E.K...E...K.A..P..V.S.I.M.SKD.S......SMKS-.NA.Q...S.S.K.HA....Y

                   210       220       230       240       250       260       270
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtWHY1    FFNLSVQNKLVNVDESIYIPITRAEFAVLISAFNFVLPYLIGWHAFANSIKPEET--SRVNNASPNYGGDYEWNR
AtWHY3    ..........L.....V.....K.................H....S..........DS--N.L.....K.......S.
AtWHY2    .IS...N.SILKTNDYFVV.V.K.....MKT..S.A..HIM..NRLTGHVNT.ALPSRN.SHLKTEPQLEL..DK
```

**Appendix Figure S1. Alignment of AtWHIRLY protein sequences**. Peptide sequences are aligned using BioEdit version 7.2.5.

**Appendix Figure S2. Validation of the exxpression level of aliphatic and indole GSL genes in seedlings by qRT-PCR.** Normalized expression level by qRT-PCR (bar charts on left vertical axis) and normalized read counts by RNA-seq (line chart on the right vertical axis), both show the average of mean and error bars show standard deviation of three independent biological replicates. Asterisks indicate statistically significant level of Student's t-test between qRT-PCR-based normalized expression level of selected genes in the *why1-5* mutant and WT, (*) p-value < 0.05, (**) p-value < 0.01, (***) p-value < 0.001.
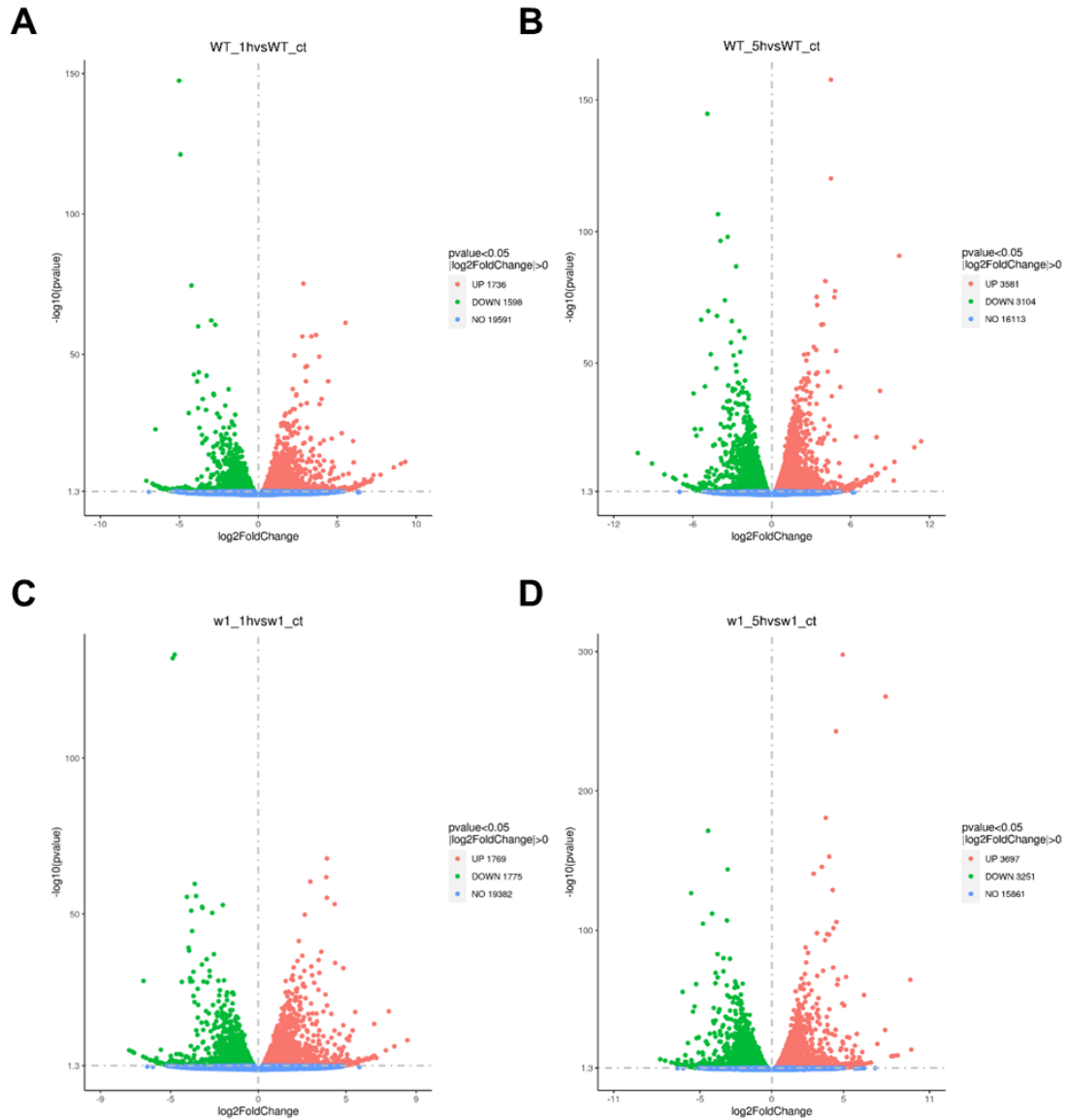
**Appendix Figure S3. Volcano plots show differentially expressed genes in response to oxidative stress.** Each dot presents one gene, in which red means upregulated and green shows downregulated gene. The comparisons between oxidative stress transcriptomes and control transcriptome of both the WT and the WHIRLY1 knock-out backgrounds, include (**A**) WT_1h vs WT_ct, (**B**) WT_5h vs WT_ct, (**C**) why1_1h vs why1_ct, and (**D**) why1_5h vs why1_ct.
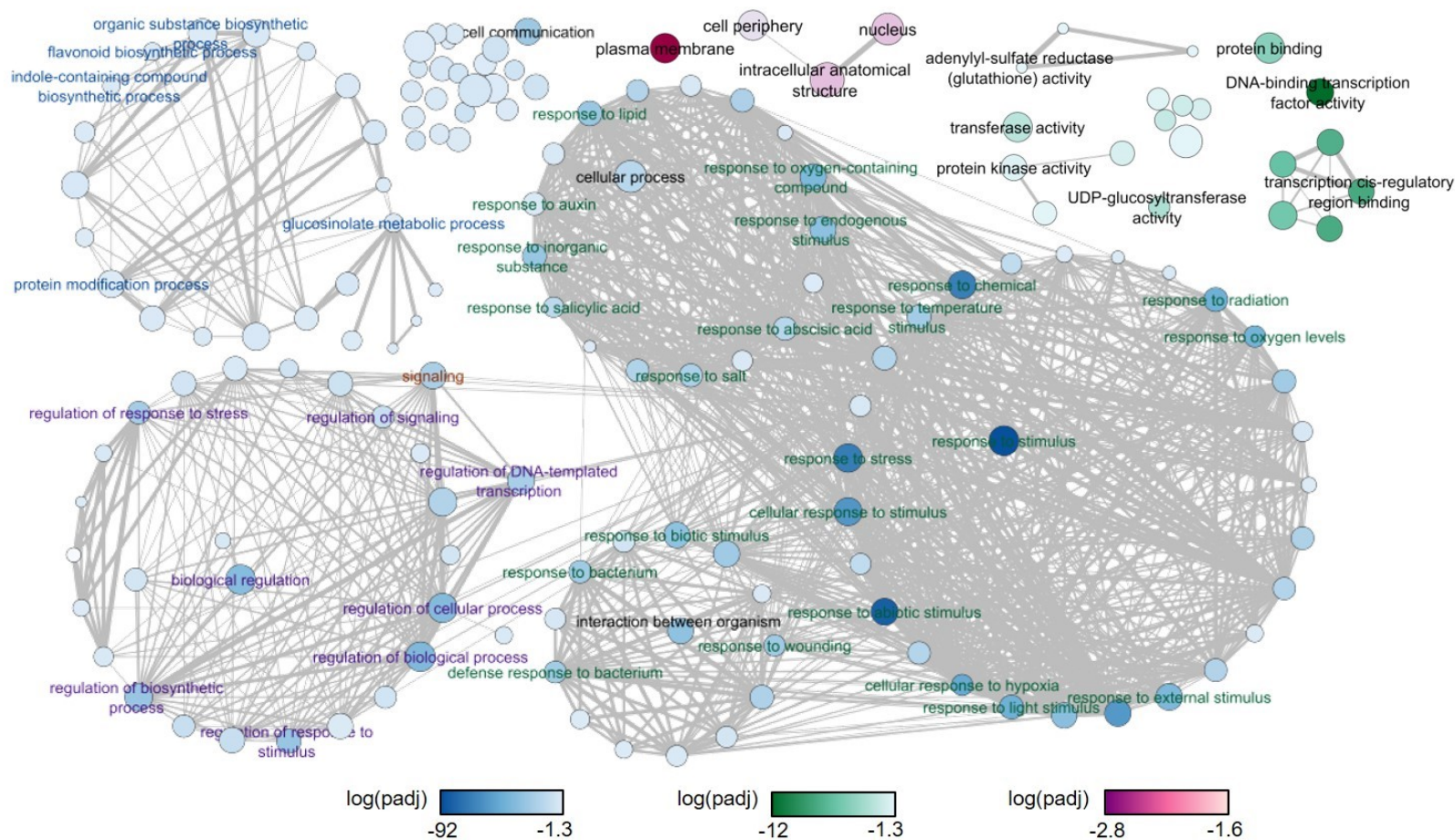
**Appendix Figure S4. GO enrichment analysis of oxidative-stress-associated DEGs after 1 hour under the stress.** Each circle indicates a GO term, and related GO terms are connected by lines. Different colors of circle show GO categories (blue – biological process, green – molecular function, pink – cellular compartment) and different color shades reflect log(p-adj) of enriched GO term.
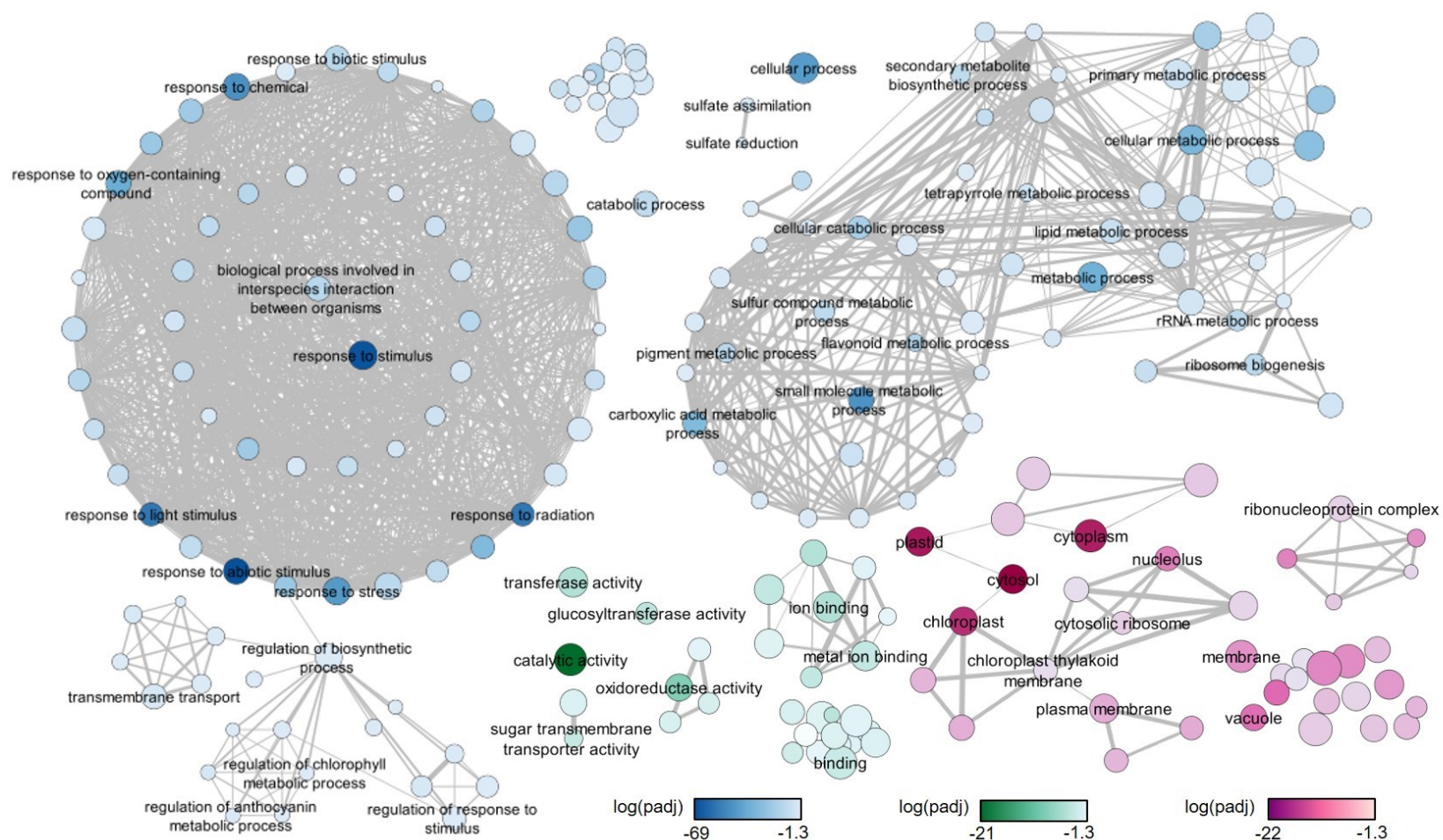
**Appendix Figure S5. GO enrichment analysis of oxidative-stress-associated DEGs after 5 hours under the stress.** Each circle indicates a GO term, and related GO terms are connected by lines. Different colors of circle show GO categories (blue – biological process, green – molecular function, pink – cellular compartment) and different color shades reflect log(p-adj) of enriched GO term.
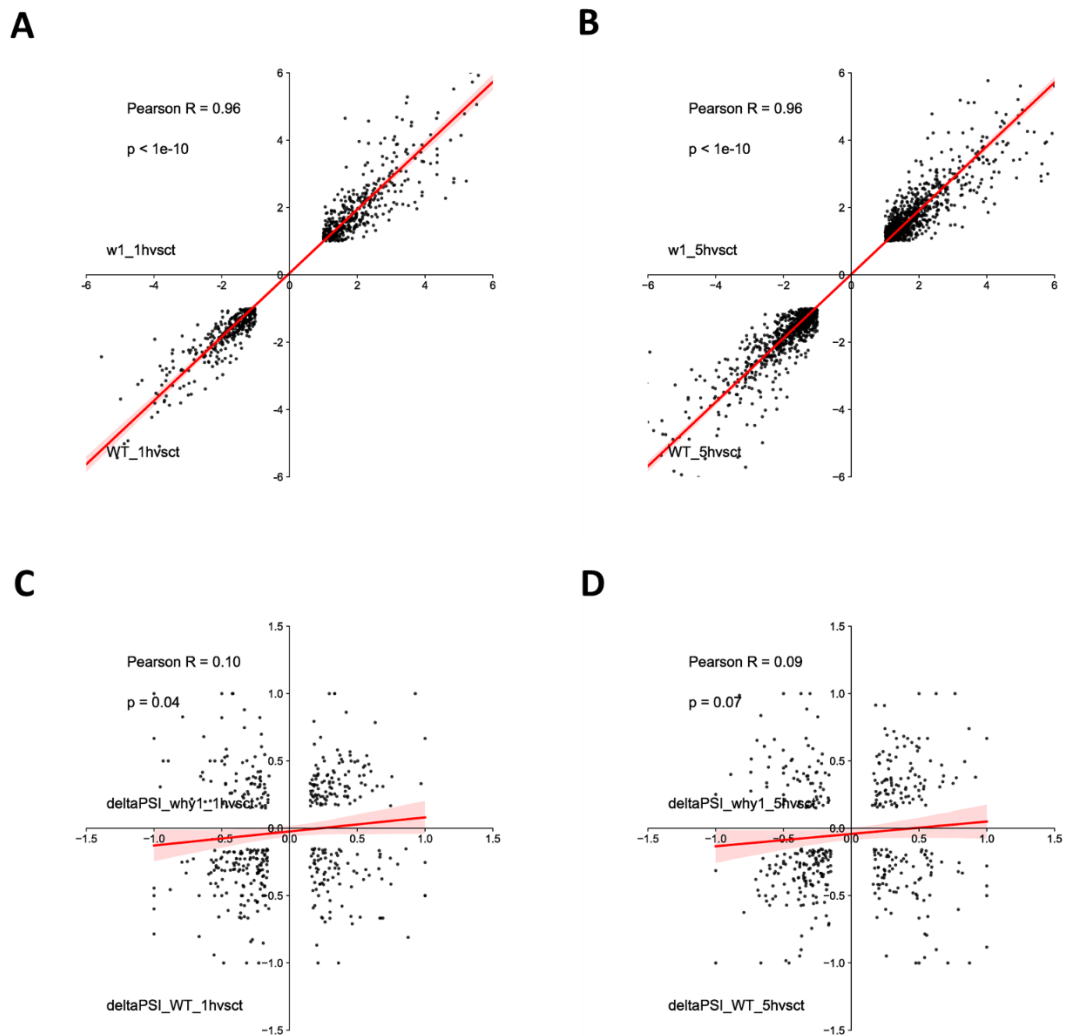
**Appendix Figure S6. Correlation of changes in gene expression level or alternative splicing between the why1 mutant and the WT upon oxidative stress**. Scatter plots showing correlation of gene expression fold-change (**A, B**) and change in percent splice in (**C, D**) in response to oxidative stress between the why1 mutant and the WT using log2FC value and ΔPSI value, respectively.
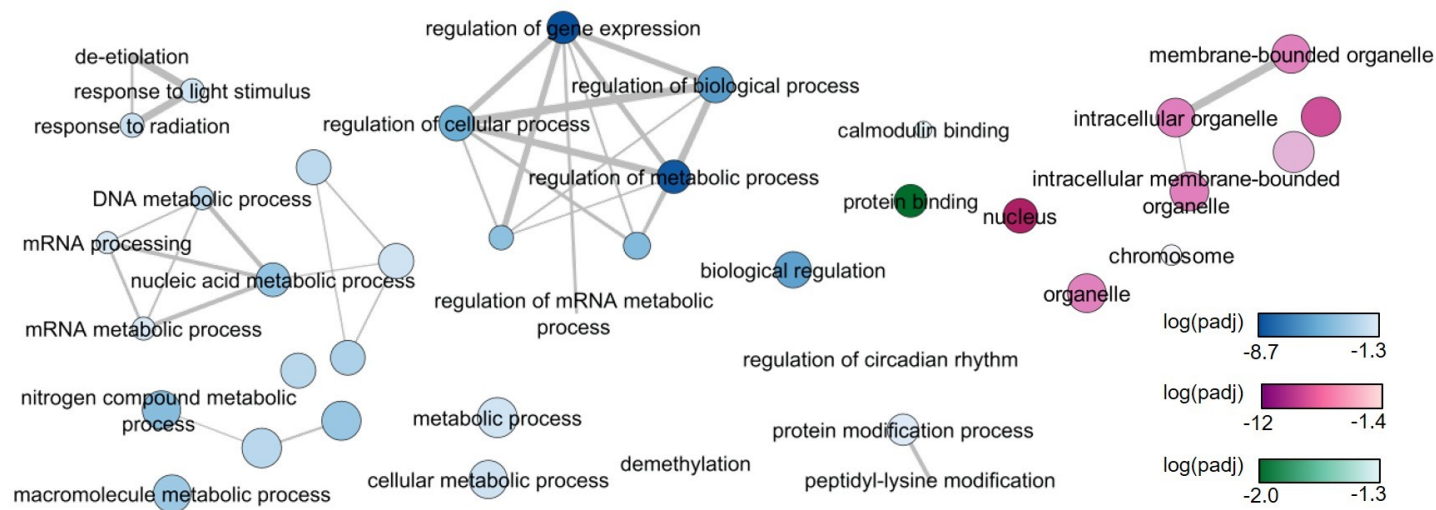
**Appendix Figure S7. GO enrichment analysis of oxidative-stress-associated differentially alternative spliced genes after 1 hour of oxidative stress.** Each circle indicates a GO term, and related GO terms are connected by lines. Different colors of circle show GO categories (blue – biological process, green – molecular function, pink – cellular compartment) and different color shades reflect log(p-adj) of enriched GO term.
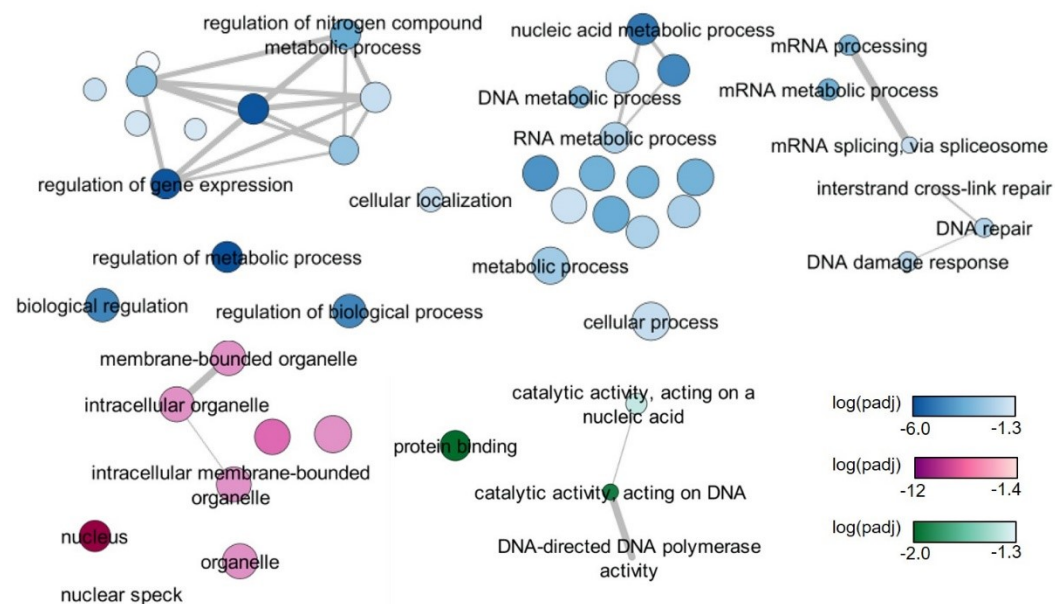
**Appendix Figure S8. GO enrichment analysis of oxidative-stress-associated differentially alternative spliced genes after 5 hours of oxidative stress.** Each circle indicates a GO term, and related GO terms are connected by lines. Different colors of circle show GO categories (blue – biological process, green – molecular function, pink – cellular compartment) and different color shades reflect log(p-adj) of enriched GO term.
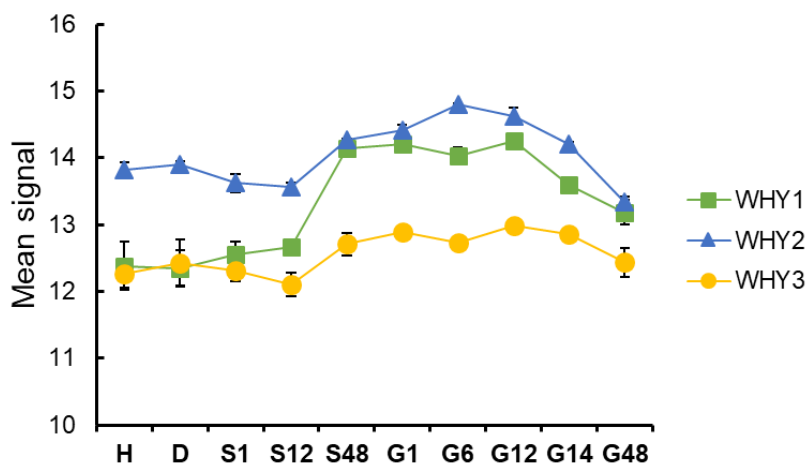
**Appendix Figure S9. Expression of *WHIRLIES* during germination of *A. thaliana* Col-0 seeds.** The microarray data from Narsai et al. (2017) was extracted from Genevestigator (Hruz et al., 2008). Each data point represents the mean of three replicates and bars show standard deviation. H: freshly harvested; D: dried (15 days in darkness); S: stratification (on MS plates, 4°C in the dark); G: germination (22°C under continous light).
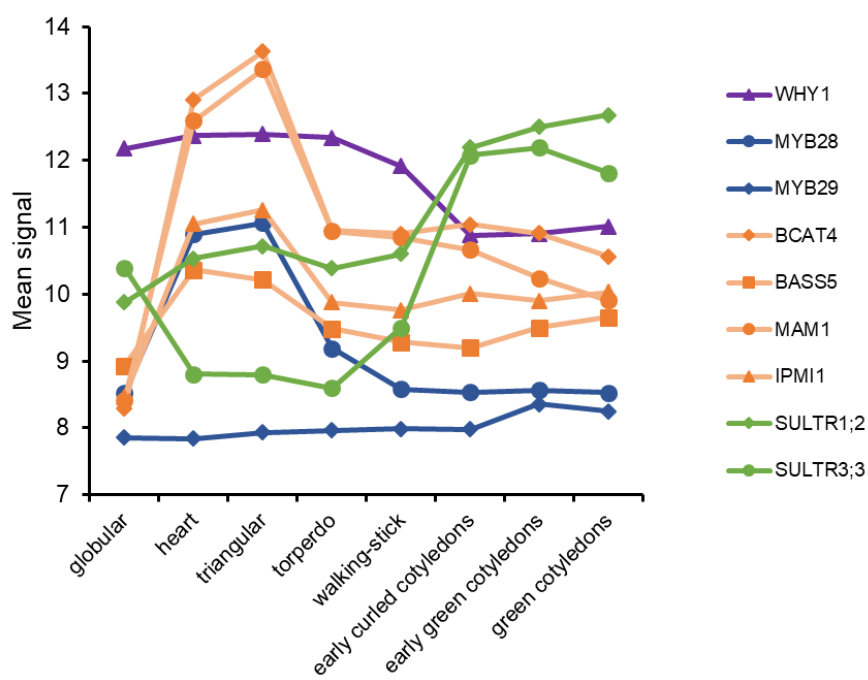


**Appendix Figure S10. Expression of *WHIRLIES* and other aGSL biosynthetic genes during embryogenesis in *A. thaliana*.** The microarray data of AtGenExpress Project number PRJNA96827 was extracted from GeneInvestigator (Hruz et al., 2008). Each data point represents the mean of three replicates.
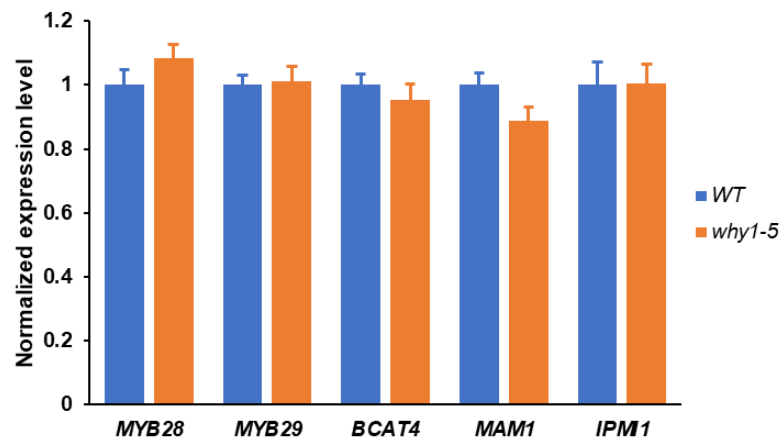
**Appendix Figure S11. Normalized expression level of aGSL genes in mature plants.** WT and the *WHIRLY1* knock-out mutant were grown under short-day condition for 5 weeks and leaves 9th and 11th of three plants were collected to analyze gene expression level. Bar chart shows the average of mean and error bars denote standard deviation of four technical replicates.
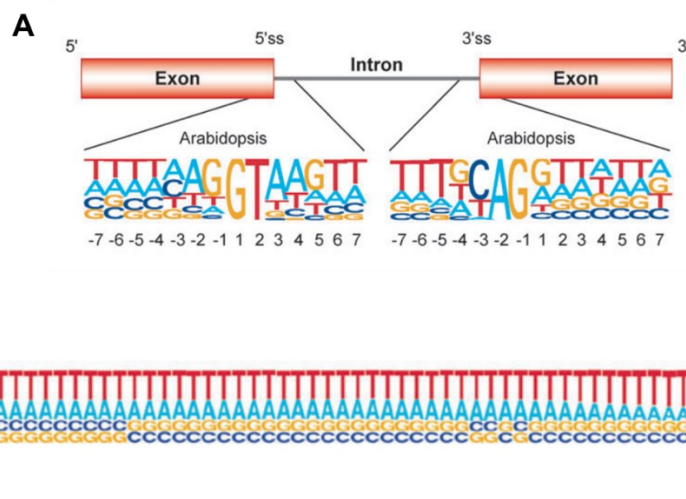


**Appendix Figure S12. Flanking sequences at splicing sites in *A. thaliana*.** (A) The sequence patterns at the 5' and 3' splicing sites of introns in *Arabidopsis*. (B) The sequence pattern of last 70 nucleotides of the *Arabidopsis* introns. The frequencies of A, C, G, and T at each position are indicated by the height of the corresponding letter (Reddy, 2007).

**A**



**B**
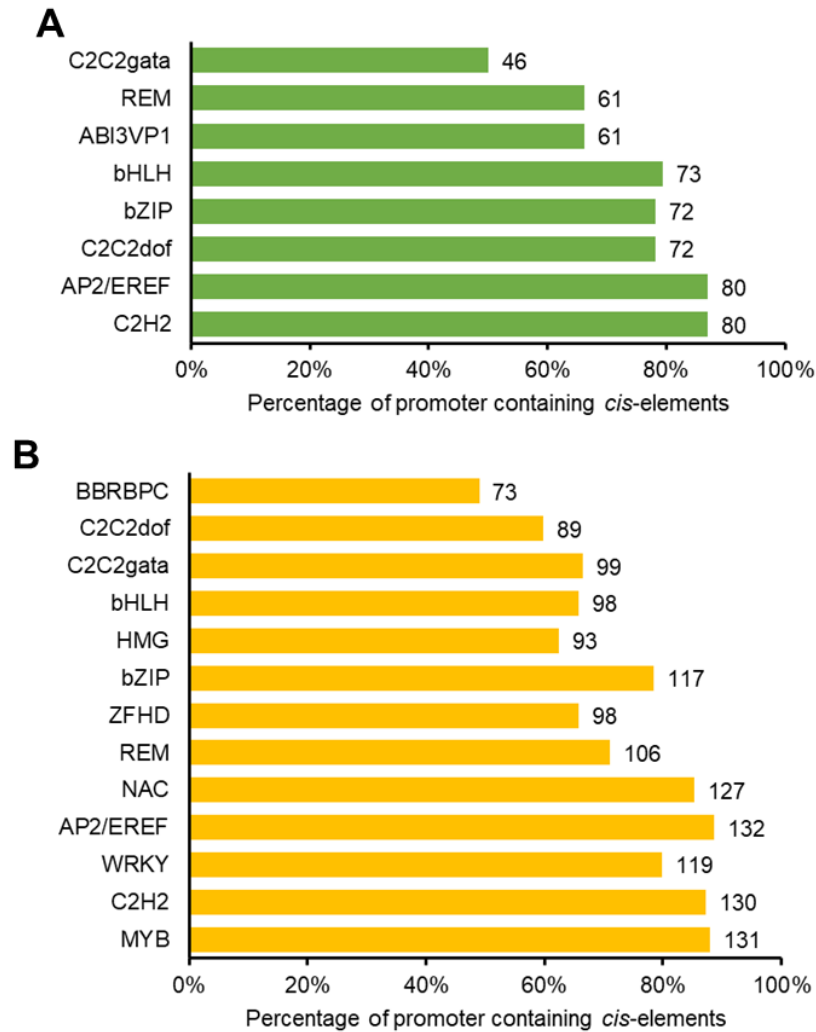


**Appendix Figure S14. Enriched binding sites of transcription factor family on promoters of deregulated genes** in control condition (**A**) and oxidative stress (**B**). 1000-bp promoter region of 92 and 149 specific DEGs between the *WHIRLY1* knock-out and the WT seedling at non-stress and stress condition was analyzed using SEA version 5.5.1, to scan known TF binding sites based on O'Malley database.

**Appendix Table S1. DEGs between the 5-day-old *WHIRLY1* knock-out mutant and WT seedlings.** Genes involved in different pathways are denoted by different colors, in which green resembles biotic stress response (1), orange, abiotic response (2), blue, developmental process (3), and yellow, RNA metabolism (4).

| gene_id | gene_name | l2FC | 1 | 2 | 3 | 4 | gene_description |
|---------|-----------|------|---|---|---|---|------------------|
| AT1G17810 | TIP3-2 | -5.75 | | | x | | BETA-TIP. Involved in transmembrane water transport |
| AT4G35590 | RKD5 | -5.62 | | | | | Protein RKD5 |
| AT1G17380 | JAZ5 | -4.99 | x | | x | | Protein TIFY 11A |
| AT3G03200 | NAC045 | -4.92 | | | | | NAC045 |
| AT4G09030 | AGP10 | -4.90 | x | | | | Classical arabinogalactan protein 10 |
| AT5G59320 | LTP3 | -4.80 | | x | x | | Lipid-transfer protein 3, inlvoled in lipid transport |
| AT5G35375 | - | -4.77 | | | x | | Transmembrane protein. Act upstream or within of seed development |
| AT1G14410 | WHY1 | -4.42 | x | x | x | | Single-stranded DNA-binding protein WHY1, chloroplastic |
| AT5G08760 | STMP8 | -3.16 | x | x | | | Secreted peptide which functions in plant growth and pathogen defense. |
| AT3G28007 | SWEET4 | -3.08 | | | x | | Bidirectional sugar transporter SWEET4 |
| AT3G06355 | - | -3.05 | | | | | long_noncoding_rna |
| AT5G38565 | - | -2.96 | - | - | | | Probable FBD-associated F-box protein At5g38565 |
| AT1G16410 | CYP79F1 | -2.84 | x | | | | Dihomomethionine N-hydroxylase |
| AT1G03920 | NDR2 | -2.30 | | | x | | Non-specific serine/threonine protein kinase, involved in pollen development |
| AT4G33330 | PGSIP3 | -2.28 | | | x | | Encodes a glucuronyltransferase responsible for the addition of GlcA residues onto xylan and for secondary wall deposition |
| novel.44 | - | -2.08 | - | - | | | PF05699:hAT family C-terminal dimerisation region\|PF14372:Domain of unknown function (DUF4413) |
| AT3G09040 | MEF12 | -1.94 | | | x | x | Pentatricopeptide repeat (PPR) superfamily protein, involved in RNA metabolism |
| AT4G13770 | CYP83A1 | -1.89 | x | x | | | Encodes a cytochrome p450 enzyme that catalyzes the initial conversion of aldoximes to thiohydroximates in the synthesis of glucosinolates not derived from tryptophan. Also has a role in auxin homeostasis. |
| AT5G23010 | MAM1 | -1.82 | x | x | | | methylthioalkylmalate synthase 1 |
| AT3G19710 | BCAT4 | -1.77 | x | x | | | Methionine aminotransferase BCAT4 |
| AT4G12030 | BASS5 | -1.76 | x | x | x | | Required for the biosynthesis of methionine-derived glucosinolates. Involved in the transport of 2-keto acids between chloroplasts and the cytosol |
| AT5G03090 | MTO1 | -1.70 | | | | | Mto 1 responding down protein. Act upstream or within macromolecule biosynthetic process, highly expressed in endosperm |
| AT1G16400 | CYP79F2 | -1.63 | x | | | | Encodes cytochrome P450 CYP79F2. Hexahomomethionine N-hydroxylase |
| AT3G02020 | AK3 | -1.61 | | | x | | Aspartokinase 3, chloroplastic. Involved in amino acid biosynthesis |
| AT2G43100 | IPMI2 | -1.59 | x | | x | | IPMI2. Small subunit, which together with IPMI SSU2, IPMI SSU3 and IPMI LSU1, is a member of heterodimeric isopropylmalate isomerase (IPMI). Together with IPMI SSU3 participates in the Met chain elongation pathway. Involved in amino acid biosynthesis |
| AT3G16700 | FAHD1B | -1.57 | | | | | Fumarylacetoacetate hydrolase homolog b |

| gene_id | gene_name | l2FC | 1 | 2 | 3 | 4 | gene_description |
|---|---|---|---|---|---|---|---|
| AT3G58990 | IPMI1 | -1.57 | x | | x | | 3-isopropylmalate dehydratase small subunit 2. Small subunit, which together with IPMI SSU1, IPMISSU2 and IPMI LSU1, is a member of heterodimeric isopropylmalate isomerase (IPMI). Together with IPMI SSU3 participates in the Met chain elongation pathway. Involved in amino acid biosynthesis |
| AT3G15351 | - | -1.48 | | | x | x | P53/DNA damage-regulated protein |
| AT3G58840 | PMD1 | -1.45 | | | x | | Peroxisomal and mitochondrial division factor 1. Involved in the morphogenesis and proliferation of peroxisomes and mitochondria |
| AT1G62560 | FMOGS-OX3 | -1.43 | x | | | | Flavin-containing monooxygenase FMO GS-OX3 |
| AT5G26270 | - | -1.34 | x | x | | | transmembrane protein |
| AT1G78370 | GSTU20 | -1.34 | | x | | | Glutathione S-transferase U20 |
| AT4G15030 | - | -1.33 | | | | | Folate-sensitive fragile site protein, involved in dephosphorylation |
| AT5G46740 | UBP21 | -1.26 | | | x | | Ubiquitin carboxyl-terminal hydrolase 21, act upstream of or within DNA repair |
| AT3G24420 | DLK2 | -1.25 | | x | | | Alpha/beta-Hydrolases superfamily protein; DLK2 is a divergent member of the DWARF14 family in stringolactone signaling |
| AT4G05400 | ML40 | -1.25 | | | x | | ribosomal protein ML40, copper ion binding protein |
| AT5G14200 | IMD1 | -1.24 | x | | x | | 3-isopropylmalate dehydrogenase. Involved in amino acid biosynthesis |
| AT1G80190 | PSF1 | -1.18 | | | x | | partner of SLD five 1, Similar to the PSF1 component of GINS complex, which in other organism was shown to be involved in the initiation of DNA replication. |
| AT2G46650 | CYTB5-C | -1.12 | x | | | | Cytochrome B5 isoform C |
| AT1G17870 | EGY3 | -1.09 | | x | | | S2P-like putative metalloprotease EGY3. Mediates chloroplastic ROS homeostasis and promotes retrograde signaling in response to salt stress. |
| AT4G01935 | - | -1.06 | | x | | | insulin-induced protein |
| AT1G02330 | CSU2 | -1.04 | | x | x | x | Encodes a nuclear coiled-coil domain-containing protein. Involved in RNA splicing, nuclear speck |
| AT4G07820 | - | 1.00 | | | | | CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) superfamily protein. |
| AT5G11250 | BNT1 | 1.03 | x | | | | Disease resistance protein (TIR-NBS-LRR class), Encodes an atypical TIR-NBS-LRR protein that is involved in stress responses. Loss of function alleles overproduce stress hormones JA,SA, ABA, and ET |
| AT1G13260 | RAV1 | 1.07 | x | x | x | | AP2/ERF and B3 domain-containing transcription factor RAV1 which is upregulated in response to low temperature. It contains a B3 DNA binding domain. It has circadian regulation and may function as a negative growth regulator |
| AT5G64110 | PER70 | 1.10 | | x | | | Peroxidase |
| AT5G13270 | RARE1 | 1.10 | | | x | x | Encodes RARE1 (Required for accD RNA Editing 1), a trans-factor essential for C-to-U editing of the chloroplast accD transcript. RARE1 carries 15 PPR (pentatricopeptide repeat) motifs, an E/E+ and a DYW domain (C-terminal tripeptide). |
| AT3G03870 | - | 1.11 | x | x | | | transmembrane protein |
| AT2G15960 | - | 1.11 | | x | | | Expressed decreased in response to proline. Highly expressed in senescence leaves. |
| AT5G54980 | CASPL2D1 | 1.12 | x | | x | x | CASP-like protein Involved in RNA metabolism |
| AT4G07950 | - | 1.12 | | | x | x | DNA-directed RNA polymerase subunit |
| AT1G15920 | CAF1H | 1.16 | | | x | x | CCR4-associated factor 1 homolog 2, deanenylase mRNA |

| gene_id | gene_name | l2FC | 1 | 2 | 3 | 4 | gene_description |
|---|---|---|---|---|---|---|---|
| AT1G69690 | TCP15 | 1.16 | | | x | | AtTCP15 is involved in the regulation of endoreduplication. Modulates GA-dependent stamen filament elongation by direct activation of SAUR63 subfamily genes through conserved target sites in their promoters. Promotes together with TCP8, 14 and 7 endoreduplication-dependent cell expansion in leaf. |
| AT3G54940 | - | 1.17 | | | x | | Papain family cysteine protease |
| AT3G21370 | BGLU19 | 1.20 | | | | | Beta-glucosidase 19 |
| AT3G23150 | ETR2 | 1.23 | | x | | | Ethylene receptor 2 |
| AT3G09830 | PCRK1 | 1.25 | x | | | | Serine/threonine-protein kinase PCRK1, a member of subfamily VIIa of the receptor-like cytoplasmic kinases (RLCKs). It contributes to pattern-triggered immunity in response to P. syringae. |
| AT5G14410 | - | 1.26 | | | | | unknown protein; Ha. |
| AT1G73120 | - | 1.27 | | x | | | F-box/RNI superfamily protein; response to oxidative stress |
| AT5G47060 | - | 1.33 | | x | | | Putative uncharacterized protein |
| AT3G58850 | PAR2 | 1.36 | | | x | | Encodes PHYTOCHROME RAPIDLY REGULATED2 (PAR2), an atypical basic helix-loop-helix (bHLP) proteinActs in the nucleus to control plant development and as a negative regulator of shade avoidance response. Functions as transcriptional repressor of auxin-responsive genes SAUR15 (AT4G38850) and SAUR68 (AT1G29510). |
| AT1G62422 | - | 1.36 | x | | | | F2401.15 |
| AT5G44120 | CRA1 | 1.41 | | | x | | 12S seed storage protein CRA1. Protein is tyrosine-phosphorylated and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds. |
| AT1G11950 | - | 1.43 | | | x | | Transcription factor jumonji (JmjC) domain-containing protein, part of Set1C/COMPASS complex in histone modification |
| AT4G37220 | - | 1.54 | | | | | Cold-regulated 413 plasma membrane protein 4 |
| AT3G21380 | - | 1.68 | | | | | Mannose-binding lectin superfamily protein. |
| AT5G01740 | - | 1.71 | | x | | | AT5g01740/T20L15_10 |
| AT5G35480 | - | 1.75 | | | | | At5g35480 |
| AT5G59940 | - | 1.77 | | | | | Cysteine/Histidine-rich C1 domain family protein |
| AT1G07135 | - | 1.87 | x | x | | | glycine-rich protein |
| AT4G16000 | - | 1.92 | | | | | Uncharacterized protein At4g16000 |
| AT3G46720 | UGT76E5 | 2.07 | | | x | | UDP-glycosyltransferase 76E5, involved in root morphogenesis |
| AT5G45820 | CIPK20 | 2.11 | | | x | | CBL-interacting serine/threonine-protein kinase 20 comprised of an N-terminal kinase catalytic domain similar to SNF1/AMPK and a unique C-terminal regulatory domain. |
| AT2G27380 | EPR1 | 2.12 | | | x | | Proline-rich extensin-like protein EPR1, encodes an extensin like gene involved in seed germination |
| AT1G15870 | - | 2.22 | | | | | Mitochondrial glycoprotein family protein |
| AT1G30040 | GA2OX2 | 2.30 | | x | x | | Gibberellin 2-beta-dioxygenase 2 acts on C-19 GA. Gene is induced in reponse to cytokinin and KNOX activity |
| AT3G15540 | IAA19 | 2.35 | | | x | | Auxin-responsive protein |
| AT2G01008 | - | 2.51 | | | x | x | Maternal effect embryo arrest 38 |
| AT2G41260 | M17 | 2.56 | | | x | | M17, involved in seed maturation |
| AT5G50550 | - | 2.57 | | | x | | Transducin/WD40 repeat-like superfamily protein, involved in protein secretion, regulation of COPII vesicle coating |
| AT4G14590 | DSP2 | 2.87 | | | x | | DEFECTIVE IN SNRNA PROCESSING 2, DSP2, Embryo defective 2739 |

| gene_id | gene_name | l2FC | 1 | 2 | 3 | 4 | gene_description |
|---|---|---|---|---|---|---|---|
| AT1G76800 | VTL2 | 3.13 | | | x | | Vacuolar iron transporter homolog 2. The gene encodes nodulin-like2 whose transcript abundance was repressed under conditions of Fe-deficient growth. |
| AT4G14980 | - | 3.25 | | | x | | Cysteine/Histidine-rich C1 domain family protein. Involved in cell cycle |
| AT4G09200 | - | 3.47 | | | x | | SPla/RYanodine receptor (SPRY) domain-containing protein.. Involved in cytoskeleton organization |
| AT4G33280 | - | 3.62 | | | | | AP2/B3-like transcriptional factor family protein |
| AT3G11840 | PUB24 | 3.68 | x | | | | E3 ubiquitin-protein ligase PUB24 that acts as a negative regulator of PAMP-triggered immunity. |
| AT5G58750 | PRISE | 3.68 | x | | x | | Putative PRISE (progesterone 5β-reductase and/or iridoid synthase-like 1,4-enone reductases). |
| AT5G08150 | SOB5 | 3.70 | | | | | SOB5, SUPPRESSOR OF PHYTOCHROME B 5 |
| AT4G25140 | OLEO1 | 3.72 | | x | x | | Encodes oleosin1, a protein found in oil bodies, involved in seed lipid accumulationFunctions in freezing tolerance of seeds. |
| AT5G45340 | CYP707A3 | 3.87 | | x | x | | Abscisic acid 8'-hydroxylase 3, involved in ABA catabolism. Gene involved in postgermination growth. Plant P450 CYP707A3, ABA 8'-hydroxylase, binds enantioselectively (+)-ABA but not (-)-ABA, whereas the enzyme binds both enantiomers of AHI1 (a structural ABA analogue used as ABA 8'-hydroxylase competitive inhibitor). |
| AT4G22410 | - | 4.30 | | | x | x | U4/U6.U5 tri-snRNP-associated-like protein. |
| AT4G28520 | CRC | 4.82 | | | x | | 12S seed storage protein CRC that is tyrosine-phosphorylated and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds. |
| AT4G29570 | CDA8 | 4.95 | | | x | | Cytidine deaminase 8 |
| AT3G05945 | DEG5 | 4.97 | | | | | long_noncoding_rna |
| AT1G64795 | - | 5.21 | | | | | - |
| AT3G03850 | SAU26 | 5.29 | | x | x | | SAUR-like auxin-responsive protein family., |
| AT5G11930 | GrxC10 | 5.46 | x | x | x | | Encodes a member of the CC-type glutaredoxin (ROXY) family that has been shown to interact with the transcription factor TGA2. Involved in Fe-S cluster assembly |
| AT4G04835 | - | 5.61 | | | | | microRNA ath-MIR841b precursor. |
| AT3G16160 | TCX8 | 5.64 | | | | | Protein tesmin/TSO1-like CXC 8, TCX8 is a transcriptional regulatory protein. It binds the LOX2 promoter and represses its expression. |
| AT1G45616 | AtRLP6 | 5.74 | | | | | Receptor-like protein 6 |
| AT4G22940 | - | 5.80 | | | | | Protein kinase superfamily protein |
| AT5G66815 | CEP5 | 5.83 | | | x | | Counteracts auxin effects by stabilizing AUX/IAA transcriptional repressors. Impact on abiotic stress processes and root development |
| AT3G47180 | CTL16 | 5.90 | | | x | | RING/U-box superfamily protein. Involved in ubiquitination and tissue developement |
| novel.187 | - | 5.91 | | | | | - |
| AT3G05955 | - | 5.95 | | | | | long_noncoding_rna |
| AT1G03880 | CRB | 6.01 | | | x | | CRU2, Protein is tyrosine-phosphorylated and its phosphorylation state is modulated in response to ABA in seeds. |
| AT1G37150 | HCS2 | 6.03 | | | | | Encode a biotin protein ligase / holocarboxylase synthetase (HCS), hcs2 mutants do not show a decrease in HCS activity. |
| AT1G52855 | - | 6.30 | | x | | | At1g52855 |
| AT1G78720 | - | 6.31 | | | | | SecY protein transport family protein. |

| gene_id | gene_name | l2FC | 1 | 2 | 3 | 4 | gene_description |
|---|---|---|---|---|---|---|---|
| AT2G06995 | - | 6.43 | | | | | long_noncoding_rna |
| AT5G52545 | - | 6.49 | | | x | x | RNA-binding protein-like RNA recognition motif protein. |
| AT5G43530 | RAD5B | 6.58 | | | x | | DNA repair protein RAD5B, Helicase protein with RING/U-box domain-containing protein. |
| AT3G32047 | - | 6.96 | | | | | Cytochrome P450 superfamily protein |
| AT2G11405 | - | 7.23 | | | | | transmembrane protein. |
| AT1G75945 | - | 8.00 | | x | x | x | acts upstream of or within RNA metabolic process, response to heat |

**Appendix Table S2. RNA splicing factors/regulators which were differentially alternative spliced in the *WHIRLY1* knock-out mutant at control conditions**

| gene id | name | gene description |
|---|---|---|
| AT1G78290 | SRK2C | SRK2C |
| AT2G35510 | SRO1 | Probable inactive poly [ADP-ribose] polymerase SRO1 |
| AT4G02430 | SR34b | RNA-binding (RRM/RBD/RNP motifs) family protein |
| AT5G37370 | SRL1 | Pre-mRNA splicing factor SR-like 1 |
| AT5G52040 | ATRSP41 | RNA-binding (RRM/RBD/RNP motifs) family protein |
| AT2G46610 | RS31A | Serine/arginine-rich splicing factor RS31A |
| AT4G25500 | RS40 | Serine/arginine-rich splicing factor RS40 |
| AT3G07590 | SMD1A | Small nuclear ribonucleoprotein SmD1a |
| AT2G41500 | LIS | WD-40 repeat family protein / small nuclear ribonucleoprotein Prp4p-related |
| AT3G53500 | RS2Z32 | RNA-binding (RRM/RBD/RNP motifs) family protein with retrovirus zinc finger-like domain |
| AT1G60200 | NA | Splicing factor PWI domain-containing protein / RNA recognition motif (RRM)-containing protein |
| AT1G06960 | NA | U2 small nuclear ribonucleoprotein B" 2 |
| AT5G55100 | NA | SWAP (Suppressor-of-White-APricot)/surp domain-containing protein |
| AT3G03970 | SINE2 | ARM repeat superfamily protein |
| AT5G41690 | NA | RNA-binding (RRM/RBD/RNP motifs) family protein |
| AT1G50300 | TAF15 | Transcription initiation factor TFIID subunit 15 |

**Appendix Table S3. GSL content in 5-day-old knock-out *why1-5* and WT seedlings.** The metabolite concentrations were determined by LC-MS/MS. Descriptive statistics of each biological replicate includes mean, standard deviation (SD), p-value of Student's t-test between why1 and WT samples. Asterisks indicate statistically significant levels, (*) p-value < 0.05, (**) p-value < 0.01, (***) p-value < 0.001. 4MOI3M, 4-methoxy-3-indolylmethyl; I3M, indol-3-ylmethyl; 3MSOP, 3-methylsulfinylpropyl; 3MTP, 3-methylthiopropyl; 4MSOB, 4-methylsulfinylbutyl; 4MTB, 4-methylthiobutyl; 5MSOP, 5-methylsulfinylpentyl; 5MTP, 5-methylthiopentyl; 6MSOH, 6-methylsulfinylhexyl; 6MTH, 6-methylthiohexyl; 7MSOH, 7-methylsulfinylheptyl; 7MTH, 7-methylthioheptyl; 8MSOO, 8-methylsulfinyloctyl; 8MTO, 8-methylthiooctyl

| | GSL | Genotype | Mean | SD | p-value | summary |
|---|---|---|---|---|---|---|
| **Indole GSL** | I3M | WT | 2.45E-01 | 0.02773 | 0.3932 | n.s. |
| | | *why1* | 2.68E-01 | 0.031416 | | |
| | 4MOI3M | WT | 3.71E-02 | 0.003637 | 0.6237 | n.s. |
| | | *why1* | 4.10E-02 | 0.011994 | | |
| **short-chain aliphatic GSL** | 3MSOP | WT | 1.80E-03 | 0.000387 | 0.9450 | n.s. |
| | | *why1* | 1.85E-03 | 0.001181 | | |
| | 3MTP | WT | 6.51E-03 | 0.001287 | 0.0045 | ** |
| | | *why1* | 2.15E-03 | 0.000249 | | |
| | 4MSOB | WT | 1.75E-01 | 0.014207 | 0.0305 | * |
| | | *why1* | 1.21E-01 | 0.024803 | | |
| | 4MTB | WT | 1.81E+00 | 0.281371 | 0.0146 | * |
| | | *why1* | 1.05E+00 | 0.152345 | | |
| | 5MSOP | WT | 3.14E-03 | 0.000315 | 0.0027 | ** |
| | | *why1* | 1.82E-03 | 0.000147 | | |
| | 5MTP | WT | 1.90E-01 | 0.033883 | 0.0467 | * |
| | | *why1* | 1.28E-01 | 0.016483 | | |
| **long-chain aliphatic GSL** | 6MSOH | WT | 5.00E-03 | 0.00049 | 0.1749 | n.s. |
| | | *why1* | 4.04E-03 | 0.000886 | | |
| | 6MTH | WT | 7.16E-02 | 0.011927 | 0.1567 | n.s. |
| | | *why1* | 5.87E-02 | 0.004941 | | |
| | 7MSOH | WT | 2.95E-02 | 0.003928 | 0.2354 | n.s. |
| | | *why1* | 2.52E-02 | 0.003597 | | |
| | 7MTH | WT | 3.94E-01 | 0.053542 | 0.1128 | n.s. |
| | | *why1* | 3.26E-01 | 0.023181 | | |
| | 8MSOO | WT | 7.60E-02 | 0.024914 | 0.3384 | n.s. |
| | | *why1* | 5.88E-02 | 0.011218 | | |
| | 8MTO | WT | 6.27E-01 | 0.089332 | 0.1583 | n.s. |
| | | *why1* | 5.27E-01 | 0.045485 | | |

**Appendix Table S4. Distribution of differentially expressed genes in response to oxidative stress in Core ROS Wheel** (Willems et al., 2016).

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvs WTct | Log2FC_w11hvsw 1ct | Log2FC_WT5hvs WTct | Log2FC_w15hvsw 1ct |
|---|---|---|---|---|---|---|---|
| III | AT1G01430 | TBL25 | TRICHOME BIREFRINGENCE-LIKE 25, involved in the synthesis and deposition of secondary wall cellulose | -1.15634 | -1.21114 | - | - |
| IV | AT1G01720 | NAC002 | NAC transcription activator, induced in response to wounding and abscisic acid. | 1.307242 | 1.694008 | - | 1.049773 |
| I | AT1G03130 | PSAD2 | photosystem I subunit D-2 (PSAD-2) | - | - | -2.07753 | -2.10092 |
| III | AT1G03870 | FLA9 | fasciclin-like arabinogalactan-protein 9 (Fla9) | - | - | -1.00116 | -1.35968 |
| III | AT1G05170 | B3GALT2 | Galactosyltransferase family protein | 1.114134 | 1.203154 | 1.27836 | 1.566012 |
| VI, VII | AT1G05575 | - | transmembrane protein | 3.989066 | 1.738223 | - | - |
| VI | AT1G07135 | - | glycine-rich protein | 3.906203 | 2.330678 | - | -1.57574 |
| III | AT1G09070 | SRC2 | Encodes SRC2, specifically binds the peptide PIEPPPHH, and moves protein from ER to a vacuole | 1.441309 | 1.631335 | - | - |
| III | AT1G10230 | ASK18 | SKP1-like 18 (SK18) | - | - | - | - |
| I | AT1G10360 | GSTU18 | glutathione S-transferase TAU 18 (GSTU18) | - | - | -2.04006 | -2.55233 |
| II | AT1G11260 | STP1 | sugar transporter 1 (STP1) | - | - | -1.70868 | -1.96492 |
| III | AT1G14720 | XTH28 | member of Glycoside Hydrolase Family 16 | - | -1.24932 | - | - |
| III | AT1G14890 | - | Plant invertase/pectin methylesterase inhibitor superfamily protein | - | - | - | -1.10776 |
| I | AT1G17100 | - | SOUL heme-binding family protein | - | - | 1.962421 | 1.999297 |
| III | AT1G17430 | - | alpha/beta-Hydrolases superfamily protein | - | - | - | -2.00621 |
| III | AT1G17870 | EGY3 | Encodes S2P-like putative metalloprotease, located on membrane | - | - | - | 1.4929 |
| III | AT1G18570 | MYB51 | R2R3-MYB transcription family. Involved in indole glucosinolate biosynthesis. | 2.659229 | 2.332517 | - | - |
| III | AT1G18590 | SOT17 | Encodes desulfoglucosinolate sulfotransferase, involved in the final step of glucosinolate core structure biosynthesis. | - | - | 1.972725 | 2.261731 |
| V, VII | AT1G19020 | SDA | SMALL DEFENSE-ASSOCIATED PROTEIN 1, Modulates defense against bacterial pathogens and tolerance to oxidative stress | 4.68208 | 3.164323 | 2.984716 | - |
| III | AT1G20450 | ERD10 | Encodes a gene induced by low temperature and dehydration. | 1.45619 | 1.394053 | 1.517504 | 1.971293 |
| I | AT1G21500 | - | hypothetical protein | - | - | -1.54381 | -1.67894 |
| III | AT1G21550 | CML44 | Calcium-binding EF-hand family protein | - | - | 5.952165 | - |
| I | AT1G22630 | - | SSUH2-like protein | - | - | -1.23457 | - |
| II, III | AT1G25230 | - | Calcineurin-like metallo-phosphoesterase superfamily protein | - | - | -3.35439 | -3.20928 |
| III | AT1G26800 | MPSR1 | RING/U-box superfamily protein | -1.88417 | -2.47938 | -2.49138 | -3.01813 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvsWTct | Log2FC_w11hvsw1ct | Log2FC_WT5hvsWTct | Log2FC_w15hvsw1ct |
|---|---|---|---|---|---|---|---|
| VI | AT1G27730 | ZAT10 | Cys2/His2-type zinc-finger transcription factor, involved in response to photooxidative stress. | 4.002284 | 4.753211 | 1.906011 | 2.254572 |
| III | AT1G29690 | CAD1 | Encodes negative regulator controlling SA-mediated pathway of programmed cell death in plant immunity. | 1.433145 | 1.587204 | - | - |
| III | AT1G29720 | RFK1 | Leucine-rich repeat transmembrane protein kinase | -1.34754 | - | -1.80449 | -3.09036 |
| III | AT1G30070 | - | SGS domain-containing protein | - | - | - | 1.333291 |
| I | AT1G30520 | AAE14 | acyl-activating enzyme 14 (AAE14) | -1.70451 | -1.24372 | -1.79117 | -2.02105 |
| III, IV | AT1G32870 | ANAC13 | NAC domain protein 13 (NAC13) | 1.982808 | 1.612597 | 1.847909 | 1.335274 |
| III | AT1G33240 | AT-GTL1 | Encodes a transcriptional activator that contains two separate but similar trihelix DNA-binding domains, similar to GT-2. | -2.12353 | -2.18652 | -1.93114 | -1.75447 |
| II | AT1G34760 | GRF11 | general regulatory factor 11 (GRF11) | - | - | -2.65349 | -2.50827 |
| I | AT1G44575 | PSBS | NONPHOTOCHEMICAL QUENCHING 4 (NPQ4) | - | - | 1.196243 | 1.15807 |
| III | AT1G48480 | RKL1 | *Arabidopsis* thaliana receptor-like protein kinase (RKL1) gene | - | - | -1.03263 | -1.35122 |
| III | AT1G49500 | - | transcription initiation factor TFIID subunit 1b-like protein | -3.8063 | -3.21144 | -2.89024 | -2.81199 |
| III | AT1G51790 | - | Leucine-rich repeat protein kinase family protein | - | - | -1.02376 | - |
| VIII | AT1G51890 | - | Leucine-rich repeat protein kinase family protein | - | - | -2.13327 | - |
| III | AT1G53440 | - | Leucine-rich repeat transmembrane protein kinase | 1.138898 | 1.443969 | - | - |
| III | AT1G53540 | HSP17.6C | HSP20-like chaperones superfamily protein | - | - | - | 5.101411 |
| II | AT1G54570 | - | Esterase/lipase/thioesterase family protein | - | - | 1.729165 | 2.414509 |
| III | AT1G54820 | - | Protein kinase superfamily protein | -2.16311 | -2.30914 | -1.82167 | -1.16087 |
| V | AT1G56060 | CYSTM3 | Encodes a mitochondrial protein that is induced by salt stress | 7.195002 | 6.479944 | - | - |
| III | AT1G56220 | - | Dormancy/auxin associated family protein | - | - | -3.09003 | -3.11459 |
| III | AT1G56600 | GOLS2 | Encodes a galactinol synthase that catalyzes the formation of galactinol from UDP-galactose and myo-inositol. | 2.275051 | 2.718468 | 1.376782 | 2.501177 |
| I | AT1G58290 | HEMA1 | HEMA1 | - | - | - | 1.135385 |
| III, VI | AT1G61340 | - | A F-box protein induced by various biotic or abiotic stress. | 3.36091 | 3.957971 | 2.381403 | 2.712312 |
| III | AT1G62560 | FMOGS-OX3 | Encodes flavin-monooxygenase glucosinolate S-oxygenase that catalyzes the conversion of methylthioalkyl glucosinolates to methylsulfinylalkyl glucosinolates | - | - | 1.779955 | 2.533087 |
| III | AT1G66890 | - | 50S ribosomal-like protein | -1.43882 | - | - | - |
| III | AT1G67360 | - | Rubber elongation factor protein (REF) | 2.013454 | 2.393009 | 2.559972 | 3.261028 |
| III | AT1G69530 | ATEXPA1 | Member of Alpha-Expansin Gene Family. Naming convention from the Expansin | - | -1.02416 | -1.16645 | -1.40908 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvsWTct | Log2FC_w11hvsw1ct | Log2FC_WT5hvsWTct | Log2FC_w15hvsw1ct |
|---|---|---|---|---|---|---|---|
| III | AT1G70160 | - | zinc finger MYND domain protein | - | - | -1.12258 | -1.07687 |
| III | AT1G73480 | - | alpha/beta-Hydrolases superfamily protein | - | 2.113153 | 1.990201 | 2.83117 |
| III | AT1G74310 | CLPB1 | Encodes ClpB1, which belongs to the Casein lytic proteinase/heat shock protein 100 (Clp/Hsp100) family. Involved in refolding of proteins which form aggregates under heat stress. | 1.458961 | 1.174101 | - | - |
| III | AT1G74320 | - | Encodes a choline kinase, whose expression is induced by high salt and mannitol. | - | - | -1.53762 | -1.26957 |
| III | AT1G74680 | - | Exostosin family protein | -1.12014 | - | - | - |
| I | AT1G75460 | - | ATP-dependent protease La (LON) domain protein | -1.35551 | -1.2496 | -1.46974 | -1.31955 |
| I | AT1G76100 | PETE1 | plastocyanin 1 (PETE1) | - | - | -1.86553 | -1.8145 |
| IV | AT1G77450 | NAC032 | NAC domain containing protein 32 (NAC032) | 2.269302 | 2.50947 | 2.160562 | 1.924804 |
| III | AT1G78070 | - | Transducin/WD40 repeat-like superfamily protein | 1.779678 | 2.312007 | 1.027429 | 1.210947 |
| III | AT1G78460 | - | SOUL heme-binding family protein | - | - | -1.43309 | -1.46339 |
| III | AT1G79380 | RGLG4 | Encodes a ubiquitin ligase that is an essential upstream modulator of JA signaling in response to various stimuli. | - | - | -1.44631 | -1.57483 |
| IV | AT1G80440 | - | Galactose oxidase/kelch repeat superfamily protein | -3.81293 | -3.97655 | -5.93834 | -5.01084 |
| VI | AT1G80840 | WRKY40 | Pathogen-induced transcription factor | 6.002553 | 6.606691 | 3.120035 | 3.292011 |
| II | AT1G80920 | ATJ8 | Chaperone DnaJ-domain superfamily protein, chloroplast stroma | -2.28299 | -3.13761 | -4.47258 | -2.84858 |
| III | AT2G01420 | PIN4 | Encodes an auxin efflux carrier, involved in the maintenance of embryonic auxin gradients. | - | - | -1.12237 | -1.3683 |
| I | AT2G03750 | SOT11 | P-loop containing nucleoside triphosphate hydrolases superfamily protein | -1.51422 | -1.36485 | - | - |
| IV | AT2G04050 | DTX3 | MATE efflux family protein | - | - | 3.693159 | 8.773619 |
| IV, VII | AT2G15480 | UGT73B5 | UDP-glucosyl transferase 73B5 (UGT73B5) | 1.414706 | 1.939439 | - | 1.111257 |
| IV | AT2G15490 | UGT73B4 | UDP-glycosyltransferase 73B4 (UGT73B4) | - | 1.398161 | 1.641288 | 1.78005 |
| III | AT2G20560 | - | DNAJ heat shock family protein | 2.231089 | 1.324589 | - | - |
| III | AT2G20940 | - | transmembrane protein, putative (DUF1279) | - | - | 1.230232 | - |
| III | AT2G21320 | BBX18 | B-box zinc finger family protein | 1.151516 | 1.143201 | 1.423393 | 1.439061 |
| I | AT2G21330 | FBA1 | fructose-bisphosphate aldolase 1 (FBA1) | - | - | 1.174771 | 1.051827 |
| III | AT2G22240 | IPS2 | Myo-inositol-1-phosphate synthase isoform 2. | 2.284858 | 1.783728 | 3.38147 | 3.577082 |
| II | AT2G25900 | ATCTH | ATCTH | -4.07947 | -3.89163 | -5.06406 | -5.36734 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvsWTct | Log2FC_w11hvsw1ct | Log2FC_WT5hvsWTct | Log2FC_w15hvsw1ct |
|---|---|---|---|---|---|---|---|
| III | AT2G26150 | HSFA2 | Encodes a member of Heat Stress Transcription Factor (Hsf) family. Involved in response to misfolded protein accumulation in the cytosol. Regulated by alternative splicing and non-sense-mediated decay. | 4.532875 | 3.60286 | 2.998435 | 3.063217 |
| III | AT2G27050 | EIL1 | ethylene-insensitive3-like1 (EIL1) | -1.03119 | -1.17033 | -1.41141 | -1.2709 |
| III | AT2G28930 | APK1B | protein kinase 1B (PK1B) | -1.13097 | -1.42181 | - | - |
| VIII | AT2G29350 | SAG13 | SENESCENCE-ASSOCIATED GENE 13 encoding a short-chain alcohol dehydrogenase | - | - | 6.594751 | 6.462439 |
| IV | AT2G29420 | GSTU7 | Glutathione transferase belonging to the tau class of GSTs. Induced by SA. | 1.178798 | 1.873187 | 1.226583 | 1.757666 |
| III | AT2G29450 | GSTU5 | Encodes a member of the TAU glutathione S-transferase gene family, induced by exposure to auxin, pathogen and herbicides. | - | - | 1.110695 | 1.396534 |
| III | AT2G30480 | - | hypothetical protein | - | - | - | -1.96029 |
| VI | AT2G32030 | - | Acyl-CoA N-acyltransferases (NAT) superfamily protein | 1.919126 | 4.36283 | - | - |
| VIII | AT2G32680 | AtRLP23 | receptor like protein 23 (RLP23) | 5.252375 | - | - | - |
| III | AT2G33590 | - | NAD(P)-binding Rossmann-fold superfamily protein | 1.283193 | - | 3.033963 | 2.531461 |
| I | AT2G34620 | - | Mitochondrial transcription termination factor family protein | -2.72165 | -2.27377 | -2.53247 | -2.2087 |
| II | AT2G36050 | OFP15 | ovate family protein 15 (OFP15) | -1.29049 | -2.66564 | - | - |
| I | AT2G36145 | - | hypothetical protein | - | - | 1.530015 | 1.124449 |
| III | AT2G36750 | UGT73C1 | UDP-glucosyl transferase 73C1 (UGT73C1) | 2.671216 | 3.098713 | 4.724584 | 4.934118 |
| III | AT2G37130 | PER21 | Peroxidase superfamily protein | - | - | -1.49115 | -1.71442 |
| III | AT2G37710 | LECRK41 | Induced in response to Salicylic acid. | 1.134102 | 1.359156 | - | - |
| V, VI | AT2G38470 | WRKY33 | Member of the plant WRKY transcription factor family. Regulates the antagonistic relationship between defense pathways mediating responses to *P. syringae* and necrotrophic fungal pathogens. Involved in response to various abiotic stresses - especially salt stress. | 3.859053 | 3.907857 | - | - |
| I | AT2G39470 | PNSL1 | PsbP-like protein 2 (PPL2) | - | - | -1.02976 | - |
| III | AT2G39920 | - | HAD superfamily, subfamily IIIB acid phosphatase | - | -4.7755 | -5.90918 | - |
| VI | AT2G40140 | CZF1 | zinc finger (CCCH-type) family protein | 1.973586 | 2.010792 | - | - |
| III | AT2G40750 | WRKY54 | member of WRKY Transcription Factor | - | - | -5.75534 | - |
| III, IV, VII | AT2G41730 | HRG1 | H2O2 response gene, sensor/responder of H2O2, involved in maintaining embryonic root meristem activity | - | - | - | 1.562456 |
| III | AT2G42320 | - | nucleolar protein gar2-related | -1.60959 | -1.66654 | -1.0237 | -1.17107 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvs WTct | Log2FC_w11hvsw 1ct | Log2FC_WT5hvs WTct | Log2FC_w15hvsw 1ct |
|---------|---------|-----------|------------------|--------------------|--------------------|--------------------|--------------------|
| IV | AT2G45170 | ATG8E | Involved in autophagy. Under nutrient starvation the protein localizes to autophagosomes. | -1.45658 | -1.60007 | -2.83735 | -3.12117 |
| III | AT2G46830 | CCA1 | Encodes a transcriptional repressor of TOC1. CCA1 and LHY form a heterodimer and function synergistically in regulating circadian rhythms of *Arabidopsis*. | 1.288751 | 1.382611 | - | 1.027934 |
| IV | AT2G47000 | ABCB4 | Encodes an auxin efflux transmembrane transporter, involved in root hair elongation. | - | - | 1.067413 | 1.049093 |
| III | AT2G47180 | GOLS1 | GolS1 is a galactinol synthase that catalyzes the formation of galactinol from UDP-galactose and myo-inositol. | - | - | -1.08306 | - |
| III | AT3G04640 | - | glycine-rich protein | - | 1.842951 | - | - |
| III | AT3G07090 | - | PPPDE putative thiol peptidase family protein | 1.074809 | 1.004659 | 1.068378 | 1.465903 |
| III | AT3G07770 | HSP90-6 | HEAT SHOCK PROTEIN 89.1 (Hsp89.1) | - | - | - | 1.118716 |
| III | AT3G08590 | - | Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent | - | - | 1.016809 | 1.127967 |
| I | AT3G08940 | LHCB4.2 | light harvesting complex photosystem II (LHCB4.2) | - | - | -2.10836 | -2.18238 |
| III | AT3G08970 | ERDJ3A | J domain protein localized in ER lumen, shows similarity to HSP40 proteins and is induced by heat stress. | - | 1.822572 | - | 1.691493 |
| III | AT3G09350 | Fes1A | Encodes one of the *Arabidopsis* orthologs of the human Hsp70-binding protein 1. Fes1A is cytosolic and associates with cytosolic Hsp70. | 1.919804 | 1.840368 | 1.52137 | 1.585392 |
| III | AT3G09440 | HSP70-3 | Heat shock protein 70 (Hsp-70) family protein | 1.380734 | 1.196538 | 1.129799 | 1.337554 |
| III | AT3G10020 | - | plant/protein | -1.57649 | - | -3.3095 | -3.1361 |
| III | AT3G12320 | LNK3 | Member of a small gene family, functioning redundant with LNK1/2 affects on biomass accumulation and phototrophism. | - | - | 1.198664 | - |
| III | AT3G12580 | HSP70-4 | heat shock protein 70 (HSP70) | 1.844581 | 1.111553 | 1.187608 | 1.671599 |
| III | AT3G12610 | DRT100 | Plays role in DNA-damage repair/toleration. Partially complements RecA-phenotypes. | - | - | -1.06746 | -1.51837 |
| III | AT3G13470 | CPN60B2 | Encodes a subunit of chloroplasts chaperonins that are involved in mediating the folding of newly synthesized, translocated, or stress-denatured proteins. | - | - | - | 1.124165 |
| III | AT3G14200 | - | Chaperone DnaJ-domain superfamily protein | 2.280546 | 2.054153 | 1.198759 | 1.018984 |
| III | AT3G14310 | PME3 | encodes a pectin methylesterase, targeted by a cellulose binding protein (CBP) from the parasitic nematode *Heterodera schachtii* during parasitism. | - | - | - | -1.17999 |
| II | AT3G15450 | - | Aluminium induced protein with YGL and LRDR motifs | -2.26386 | -2.46151 | -4.81596 | -5.07232 |
| II, III | AT3G15770 | - | hypothetical protein | - | -1.136 | - | -2.57813 |
| III | AT3G16050 | PDX12 | Encodes a protein with pyridoxal phosphate synthase activity whose transcripts were detected mostly in roots and accumulate during senescence. | 1.156208 | 1.090811 | - | - |
| I | AT3G16250 | NDF4 | NDH-dependent cyclic electron flow 1 (NDF4) | - | -1.27125 | -1.50748 | -1.41579 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvsWTct | Log2FC_w11hvsw1ct | Log2FC_WT5hvsWTct | Log2FC_w15hvsw1ct |
|---------|---------|-----------|------------------|-------------------|-------------------|-------------------|-------------------|
| III | AT3G17611 | RBL14 | RHOMBOID-like protein 14 (RBL14) | - | 1.105719 | - | - |
| III | AT3G17800 | DUF760-6 | Interacts with chloroplast chaperone CLPC1. Likely involved in senescence | 1.712846 | 1.736736 | - | - |
| III | AT3G18050 | - | GPI-anchored protein | - | - | -1.94523 | -2.15083 |
| I | AT3G19800 | - | Protein of unknown function (DUF177) | - | - | -1.01314 | - |
| III | AT3G19850 | - | Phototropic-responsive NPH3 family protein | - | - | - | -4.5975 |
| II | AT3G21560 | UGT84A2 | UDP-Glycosyltransferase superfamily protein | 1.998088 | 1.684052 | 2.019837 | 1.955261 |
| IV | **AT3G22370** | **AOX1A** | Encodes AOX1a, an isoform of alternative oxidase, detoxify reactive oxygen species production when the cytochrome pathway is inhibited. AOX1a also functions as a marker for mitochondrial retrograde response. | - | - | 1.992233 | 1.771636 |
| II, III | AT3G22840 | ELIP1 | EARLY LIGHT-INDUCABLE PROTEIN (ELIP1) | 3.883072 | 3.186302 | 8.230072 | 7.935592 |
| III | AT3G23170 | PRP | Encodes a proline/serine rich protein, induced by PAMP elicitors | 2.287874 | 2.320149 | 2.705831 | 2.162209 |
| III | AT3G23920 | BAM1 | Encodes a chloroplast beta-amylase. | 1.057204 | - | 1.662796 | 1.58284 |
| III | AT3G23990 | CPN60 | mitochondrial chaperonin HSP, assists in rapid assembly of the oligomeric protein structures in the mitochondria. | - | - | 1.164344 | 1.379136 |
| III | AT3G24500 | MBF1C | One of three genes in A. thaliana encoding multiprotein bridging factor 1, a highly conserved transcriptional coactivator. May serve as a bridging factor between a bZIP factor and TBP. | 1.001528 | - | -2.21581 | - |
| I | AT3G27690 | LHCB2.4 | photosystem II light harvesting complex gene 2.3 (LHCB2.3) | - | - | -1.72729 | -1.88326 |
| III | AT3G28040 | - | Leucine-rich receptor-like protein kinase family protein | -1.20491 | -1.15844 | -1.04575 | -1.12774 |
| III | AT3G28210 | SAP12 | Encodes a putative zinc finger protein (PMZ). | - | - | - | 2.323968 |
| III | AT3G45260 | BIB | C2H2-like zinc finger protein | - | -1.71576 | - | - |
| I | AT3G47070 | - | thylakoid soluble phosphoprotein | - | - | 1.140448 | - |
| III | AT3G47620 | TCP14 | Encodes a transcription factor AtTCP14 that regulates seed germination. | -1.79424 | -1.748 | - | - |
| IV | AT3G48850 | MPT2 | phosphate transporter 3 | - | - | 3.375 | - |
| III | AT3G49160 | PKP4 | Expression of the gene is downregulated in the presence of paraquat, an inducer of photoxidative stress. | - | - | 3.123597 | 2.731496 |
| III | AT3G50700 | GAF1 | zinc finger protein, similar to maize Indeterminate1 (ID1) | -1.39969 | -2.02714 | - | -1.57752 |
| VI, VII | AT3G50930 | HSR4 | cytochrome BC1 synthesis (BCS1) | 1.326255 | 1.707813 | - | - |
| III | AT3G51790 | ATG1 | Encodes a heme-binding protein located in the mitochondrial inner membrane that is involved in cytochrome c maturation. | - | - | 1.335694 | - |
| III | AT3G51910 | HSFA7A | member of Heat Stress Transcription Factor (Hsf) family | - | - | -2.25123 | - |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvsWTct | Log2FC_w11hvsw1ct | Log2FC_WT5hvsWTct | Log2FC_w15hvsw1ct |
|---------|---------|-----------|------------------|-------------------|-------------------|-------------------|-------------------|
| III | AT3G53230 | CDC48D | ATPase, AAA-type, CDC48 protein | - | - | 1.337585 | 1.117639 |
| III | AT3G57630 | - | exostosin family protein | - | 1.59191 | - | 1.021763 |
| III | AT3G58990 | IPMI1 | isopropylmalate isomerase 1 (IPMI1) | - | - | 2.003465 | 2.810689 |
| III | AT3G59280 | PAM16 | Encodes the ortholog of yeast PAM16, part of the mitochondrial inner membrane protein import motor. Single mutant plants exhibit a smaller size and enhanced resistance against virulent pathogens. They also display elevated reactive oxygen species (ROS) accumulation. | - | - | - | 1.041012 |
| III | AT3G61820 | - | Eukaryotic aspartyl protease family protein | - | - | 1.38445 | 1.180302 |
| III | AT3G62260 | - | Protein phosphatase 2C family protein | 1.858589 | 1.628132 | 1.364695 | 1.153545 |
| III | AT3G63210 | MARD1 | encodes a novel zinc-finger protein with a proline-rich N-terminus, identical to senescence-associated protein SAG102 | -1.45325 | -1.52089 | -1.32006 | -1.36463 |
| IV | AT4G01870 | - | tolB protein-related | 2.065234 | 2.349002 | 2.585714 | 2.594971 |
| III | AT4G02050 | STP7 | sugar transporter protein 7 (STP7) | - | - | -1.29463 | - |
| III | AT4G02430 | SR34b | Serine/Arginine-Rich Protein Splicing Factors (SR proteins) 34b | - | - | -1.17238 | - |
| III | AT4G08850 | MIK2 | Leucine-rich repeat receptor-like protein kinase family protein | 1.105694 | 1.008591 | - | - |
| III | AT4G10040 | CYTC-2 | Encodes cytochrome c | - | 1.113612 | 2.22072 | 2.273479 |
| III | AT4G11660 | HSFB2B | member of Heat Stress Transcription Factor (Hsf) family | 1.580871 | 1.289988 | - | - |
| III | AT4G11900 | - | S-locus lectin protein kinase family protein | - | - | -1.60523 | -1.36863 |
| III | AT4G12030 | BASS5 | Required for the biosynthesis of methionine-derived glucosinolates. Involved in the transport of 2-keto acids between chloroplasts and the cytosol. | - | - | - | 1.942069 |
| III | AT4G13770 | CYP83A1 | Encodes a cytochrome p450 enzyme that catalyzes the initial conversion of aldoximes to thiohydroximates in the synthesis of glucosinolates not derived from tryptophan. Also has a role in auxin homeostasis. | - | - | 1.221042 | 2.089161 |
| V | AT4G14365 | XBAT34 | XB3 ortholog 4 in *Arabidopsis thaliana* (XBAT34) | 3.556105 | 4.360459 | 3.321618 | 3.898191 |
| III | AT4G14400 | ACD6 | encodes a novel protein with putative ankyrin and transmembrane regions, involved in resistance to *P. syringae*. | - | 1.464414 | -1.53286 | - |
| III | AT4G14690 | ELIP2 | Encodes an early light-induced protein | 5.519628 | 4.845335 | 9.692462 | 9.653808 |
| IV | AT4G15620 | CASPL1E2 | Uncharacterised protein family (UPF0497) | - | - | 1.074693 | 1.232905 |
| III | AT4G16370 | ATOPT3 | Encodes an oligopeptide transporter involved in metal homeostasis | - | - | -1.41101 | -1.40966 |
| VI | AT4G17490 | ERF6 | Encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family (ATERF-6) | 2.934451 | 2.939327 | - | - |
| III | AT4G19160 | - | transglutaminase family protein | - | - | -2.31922 | -2.12864 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvs WTct | Log2FC_w11hvsw 1ct | Log2FC_WT5hvs WTct | Log2FC_w15hvsw 1ct |
|---|---|---|---|---|---|---|---|
| III | AT4G19530 | - | Encodes a TIR-NB-LRR resistance protein | -1.38685 | -1.77564 | -1.27268 | -1.38235 |
| III | AT4G21320 | HSA32 | Encodes heat-stress-associated 32-kD protein. Upregulated by heat shock. | - | -1.01435 | -2.57175 | - |
| V, VII | AT4G21390 | B120 | S-locus lectin protein kinase family protein | 1.112872 | 1.940585 | - | - |
| III | AT4G22740 | - | glycine-rich protein | - | - | - | - |
| V | AT4G23190 | CRK11 | Encodes putative receptor-like protein kinase | 2.750549 | 3.830926 | - | 1.530605 |
| III | AT4G23300 | CRK22 | Encodes a cysteine-rich receptor-like protein kinase. | - | -1.07291 | -1.59909 | -1.67583 |
| III | AT4G23570 | SGT1A | Closely related to SGT1B, may function in SCF(TIR1) mediated protein degradation | 1.096025 | - | - | - |
| VI | AT4G23810 | WRKY53 | member of WRKY Transcription Factor | 2.859618 | 4.852899 | - | - |
| VI | AT4G24570 | DIC2 | Encodes one of the mitochondrial dicarboxylate carriers (DIC) | 3.640714 | 3.872925 | 1.526243 | 1.307489 |
| III | AT4G25340 | FKBP53 | Encodes a member of the FKBP-type immunophilin family that functions as a histone chaparone. Binds to 18S rDNA and represses its expression. | - | - | - | 1.09098 |
| III | AT4G26270 | PFK3 | phosphofructokinase 3 (PFK3) | - | - | -1.0132 | - |
| III | AT4G26780 | AR192 | Co-chaperone GrpE family protein | - | - | 1.431359 | 1.39636 |
| II | AT4G27450 | - | Aluminium induced protein with YGL and LRDR motifs | - | - | -3.116 | -4.57661 |
| III | AT4G27940 | MTM1 | manganese tracking factor for mitochondrial SOD2 (MTM1) | 1.278565 | 1.435818 | - | 1.320147 |
| II, III | AT4G28240 | - | Wound-responsive family protein | -2.29906 | -2.11397 | -1.6574 | -1.73575 |
| IV | AT4G33560 | - | Wound-responsive family protein | - | - | 1.145725 | - |
| III | AT4G33660 | CYSTM11 | cysteine-rich TM module stress tolerance protein | - | -1.26103 | - | - |
| II | AT4G35770 | STR15 | SENESCENCE 1 (SEN1) | 2.324632 | 2.128223 | -4.24991 | -4.78266 |
| III | AT4G36010 | - | Pathogenesis-related thaumatin superfamily protein | 1.345645 | - | 2.685523 | 2.031358 |
| III, IV, V, VII | AT4G37370 | CYP81D8 | cytochrome P450, family 81, subfamily D, polypeptide 8 | 2.096297 | 3.342623 | 2.700041 | 3.393639 |
| III | AT4G38860 | - | SAUR-like auxin-responsive protein family | -2.76284 | -3.42722 | - | - |
| VII | AT4G39670 | - | Glycolipid transfer protein (GLTP) family protein | 1.711558 | 2.134301 | 2.236025 | 2.196253 |
| III | AT5G01740 | - | Nuclear transport factor 2 (NTF2) family protein | 1.399432 | - | 1.091136 | - |
| III | AT5G02490 | HSP70-2 | Heat shock protein 70 (Hsp 70) family protein | - | - | 1.887065 | 1.807051 |
| III | AT5G03350 | LLP | Legume lectin family protein | - | - | -1.55964 | -1.55172 |
| III | AT5G03720 | HSFA3 | Heat Stress Transcription Factor A3. | - | - | - | -1.30645 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_ WT1hvs WTct | Log2FC_ w11hvsw 1ct | Log2FC_ WT5hvs WTct | Log2FC_ w15hvsw 1ct |
|---|---|---|---|---|---|---|---|
| VI, VII | AT5G04340 | ZAT6 | putative c2h2 zinc finger transcription factor mRNA | 2.671253 | 3.315555 | 2.041797 | 2.140907 |
| III | AT5G04840 | - | bZIP protein | 1.233902 | - | 1.281497 | - |
| III | AT5G05250 | - | hypothetical protein | - | -1.26123 | -3.03357 | -3.9133 |
| III | AT5G05410 | DREB2A | DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN 2 | 2.516603 | 2.288746 | - | - |
| III | AT5G07580 | ERF106 | ethylene response factor subfamily B-3 | - | -1.17739 | - | - |
| III | AT5G09590 | HSP70-10 | heat shock protein 70 (Hsc70-5) | - | - | 1.639427 | 1.894097 |
| VIII | AT5G10380 | RING1 | RING finger domain protein with E3 ligase activity | 2.773485 | 2.958166 | 3.124889 | 2.916238 |
| III | AT5G11260 | HY5 | ELONGATED HYPOCOTYL 5 | 2.392215 | 1.887126 | - | - |
| III | AT5G11610 | - | Exostosin family protein | - | - | - | -1.95146 |
| III | AT5G12020 | HSP17.6 | heat shock protein 17.6 kDa | - | - | 3.448608 | 8.348567 |
| III | AT5G13200 | - | GRAM domain family protein | - | 1.049295 | 1.160686 | 1.246374 |
| IV | AT5G14730 | - | Unknown protein, expression induced by IDL7 and stress | - | 1.024561 | - | - |
| III | AT5G15450 | CLPB3 | CASEIN LYTIC PROTEINASE B3,a chloroplast-targeted Hsp101 homologue | - | 1.032633 | 1.590417 | 2.223294 |
| II | AT5G16030 | - | mental retardation GTPase activating protein | -2.60134 | -2.50957 | -3.30838 | -3.37435 |
| III | AT5G16200 | - | 50S ribosomal protein-related | 1.318116 | 2.539539 | - | 1.103147 |
| IV | AT5G16960 | - | Zinc-binding dehydrogenase family protein | 5.37206 | 7.753747 | - | 4.966904 |
| IV | AT5G16980 | - | Zinc-binding dehydrogenase family protein | 1.651125 | 1.189226 | 1.552277 | 1.425742 |
| III | AT5G17350 | - | PADRE protein upregulated after infection by *S. sclerotiorum*. | 1.848861 | 1.113209 | - | - |
| III | AT5G17400 | ER-ANT1 | ER-localized adenine nucleotide transporter 1 | - | - | - | -1.04457 |
| III | AT5G18340 | PUB48 | ARM repeat superfamily protein | - | - | - | -5.27169 |
| VIII | AT5G18470 | - | Curculin-like (mannose-binding) lectin family protein | 5.25825 | 6.730735 | - | - |
| III | AT5G19875 | - | transmembrane protein | 1.605591 | 2.073742 | 1.080884 | 1.262478 |
| II | AT5G21170 | AKINBETA1 | 5'-AMP-activated protein kinase beta-2 subunit protein, a subunit of the SnRK1 kinase | - | - | -2.91337 | -3.0566 |
| V | AT5G25930 | - | Protein kinase family protein with leucine-rich repeat domain | 1.738609 | 2.157042 | - | - |
| VI | AT5G27420 | CNI1 | Carbon/Nitrogen Insensitive1 | 2.637711 | 2.848591 | - | - |
| II | AT5G28770 | BZIP63 | bZIP transcription factor | -3.36873 | -3.23403 | -3.11262 | -3.34107 |
| III | AT5G35460 | - | membrane protein | - | - | - | 1.021043 |
| III | AT5G39020 | - | Malectin/receptor-like protein kinase family protein | 1.47162 | 1.217465 | - | - |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvs WTct | Log2FC_w11hvsw 1ct | Log2FC_WT5hvs WTct | Log2FC_w15hvsw 1ct |
|---|---|---|---|---|---|---|---|
| IV | AT5G39050 | PMAT1 | Encodes a malonyl-transferase that may play a role in phenolic xenobiotic detoxification. | - | - | 1.460612 | 1.130489 |
| II | AT5G40450 | - | unknown protein | - | - | -1.92152 | -3.08231 |
| III | AT5G42150 | - | Glutathione S-transferase family protein | - | - | 1.374846 | 1.853546 |
| I | AT5G42760 | - | Leucine carboxyl methyltransferase | 2.41848 | 2.643461 | 1.593384 | 1.928979 |
| IV | AT5G43450 | - | encodes a protein whose sequence is similar to ACC oxidase | - | - | 1.802464 | 1.946133 |
| III | AT5G45630 | - | senescence regulator (Protein of unknown function, DUF584) | - | 3.907065 | - | 4.363564 |
| VI | AT5G47230 | ERF5 | ethylene response factor 5 | 2.000884 | 2.549709 | - | - |
| III | AT5G48570 | FKBP65 | Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins with potential to interact with Hsp90/Hsp70 as co-chaperones. | - | - | 1.568155 | - |
| II | AT5G49360 | BXL1 | beta-xylosidase 1 (BXL1) | - | - | -5.36055 | -5.6099 |
| IV | AT5G49450 | BZIP1 | Transcription factor BASIC-LEUCIN ZIPPER 1 | -3.06905 | -2.13597 | -4.01971 | -2.53523 |
| III | AT5G49910 | HSP70-7 | Stromal heat shock protein involved in protein import into chloroplast. | - | - | 1.087659 | 1.333805 |
| III | AT5G50160 | FRO8 | Encodes a ferric chelate reductase that is expressed in shoots and flowers. | - | -1.23332 | - | - |
| III | AT5G51440 | HSP23.5 | HSP20-like chaperones superfamily protein | - | - | 3.037652 | 2.910377 |
| I | AT5G51720 | NEET | 2 iron, 2 sulfur cluster binding | 1.47679 | 1.371733 | 1.987286 | 2.307749 |
| IV | AT5G51830 | - | pfkB-like carbohydrate kinase family protein | - | - | 1.475741 | 1.351274 |
| I | AT5G52570 | BETA-OHASE 2 | beta-carotene hydroxylase 2 (BETA-OHASE 2) | 1.170878 | 1.364678 | - | - |
| III | AT5G52640 | HSP90-1 | heat shock protein 90.1 | 1.353799 | - | - | - |
| VI | AT5G52750 | HIPP13 | Heavy metal transport/detoxification superfamily protein | 5.722032 | 8.494864 | 3.226559 | 5.692024 |
| III | AT5G53400 | BOB1 | Encodes a non-canonical small heat shock protein required for both development and thermotolerance. | - | - | 1.046259 | 1.198379 |
| III | AT5G53450 | FBN1 | FBN11 contains a lipid-binding FBN domain and a kinase domain. It is induced by osmotic stress | -1.65648 | -1.79357 | -3.34355 | -3.01439 |
| VII | AT5G54490 | PBP1 | PINOID (PID)-binding protein 1 | 3.655245 | - | - | - |
| III | AT5G54630 | - | zinc finger protein-related | -3.54542 | -2.9007 | -3.48693 | -2.68235 |
| III | AT5G56030 | HSP81-2 | heat shock protein 81.2 | - | - | 1.012161 | 1.192586 |
| I | AT5G57345 | OxR | Encodes a transmembrane protein localized in the ER. It is expressed throughout the plant and expression is induced in response to abiotic stress. | - | - | - | 1.258417 |
| III | AT5G57630 | CIPK21 | CBL-interacting protein kinase | - | -1.31208 | - | -1.40431 |

| Cluster | gene_id | gene_name | gene_description | Log2FC_WT1hvs WTct | Log2FC_w11hvsw 1ct | Log2FC_WT5hvs WTct | Log2FC_w15hvsw 1ct |
|---|---|---|---|---|---|---|---|
| III | AT5G58770 | - | Undecaprenyl pyrophosphate synthetase family protein | 2.188748 | 1.653627 | 1.908539 | 1.963887 |
| III | AT5G58787 | - | RING/U-box superfamily protein | -1.05142 | - | -1.59964 | -1.73544 |
| II | AT5G59080 | - | hypothetical protein | -1.60298 | -2.70358 | -1.94576 | -2.77704 |
| IV | AT5G59220 | SAG113 | highly ABA-induced PP2C gene 1 (HAI1) | - | 2.436166 | 1.940463 | 3.246223 |
| III, VI, VII | AT5G59820 | ZAT12 | Encodes a zinc finger protein involved in high light and cold acclimation. | 2.039524 | 1.582757 | 1.584961 | - |
| III | AT5G60710 | - | Zinc finger (C3HC4-type RING finger) family protein | - | - | - | -1.16868 |
| IV | AT5G61820 | - | Stress upregulated Nod 19 protein | 1.370665 | 1.654681 | 2.035611 | 2.187126 |
| II | AT5G62210 | ATS3 | Embryo-specific protein 3 | - | - | 3.058086 | 4.21565 |
| II | AT5G62280 | - | Protein of unknown function (DUF1442) | - | - | -2.49459 | - |
| III | AT5G62520 | SRO5 | SRO5 and P5CDH generate 24-nt and 21-nt siRNAs, which together are components of a regulatory loop controlling reactive oxygen species (ROS) production and stress response. | - | - | - | 2.768969 |
| III | AT5G62730 | NPF4.7 | Major facilitator superfamily protein | - | -3.04272 | - | - |
| III | AT5G63130 | - | Octicosapeptide/Phox/Bem1p family protein | 3.72645 | 3.448798 | 2.834257 | 2.531301 |
| III | AT5G64170 | LNK1 | dentin sialophosphoprotein-related | 1.797469 | 1.458464 | 2.603244 | 2.833587 |
| II, III | AT5G64410 | OPT4 | oligopeptide transporter 4 (OPT4) | - | - | -1.96773 | -1.9183 |
| III | AT5G64570 | BXL4 | beta-d-xylosidase 4 | - | - | -1.05509 | -1.33283 |
| III | AT5G66620 | DAR6 | DA1-related protein 6 (DAR6) | 1.4714 | 2.476911 | 1.161567 | 1.259973 |
| III | ATMG01090 | ORF262 | ATP synthase 9 mitochondrial | - | -2.62632 | - | - |

**Appendix Table S5. Result of two-way ANOVA tests of the influence of knock-out of *WHIRLY1* versus oxidative stress, on aGSL expression level.** p-value < 0.05 are indicated in red color, meaning there was significant effect of investigated factor on gene expression level.

| Source of Variation | MYB28 | MYB29 | BCAT4 | IPMI1 | IPMI2 | MAM1 |
|---|---|---|---|---|---|---|
| Conditions | 8.65E-05 | 9.02E-07 | 0.057843 | 5E-09 | 0.000219 | 0.011084 |
| Genotype | 0.591482 | 0.385828 | 0.001113 | 0.00025 | 0.000283 | 2.65E-05 |
| Interaction | 0.916982 | 0.140976 | 0.966171 | 0.261285 | 0.581084 | 0.901748 |

**Appendix Table S6. Oxidative stress associated differentially alternative spliced genes.** All common genes between the *WHIRLY1* knock-out mutant and the WT that exhibited changes in alternative splicing pattern in response to short (1 hour, 1h) or prolonged (5 hours, 5h) oxidative stress.

| 1h | 5h | gene_id | gene_name | gene_description |
|----|----|---------|-----------|------------------|
|    | x  | AT1G01020 | ARV1 | ARV1 family protein |
| x  |    | AT1G01060 | LHY | Encodes a myb-related putative transcription factor involved in circadian rhythm |
| x  |    | AT1G01220 | FKGP | Bifunctional fucokinase/fucose pyrophosphorylase |
| x  |    | AT1G01240 | - | transmembrane protein |
|    | x  | AT1G01448 | - | other RNA |
|    | x  | AT1G01520 | ASG4 | Homeodomain-like superfamily protein |
| x  |    | AT1G01740 | BSK4 | Serine/threonine-protein kinase BSK4 |
|    | x  | AT1G02260 | - | Divalent ion symporter |
| x  | x  | AT1G02960 | - | Kinetochore protein |
|    | x  | AT1G03457 | - | RNA-binding (RRM/RBD/RNP motifs) family protein |
| x  |    | AT1G03905 | ABCI19 | ABC transporter I family member 19 |
|    | x  | AT1G03960 | B"EPSILON | Serine/threonine protein phosphatase 2A regulatory subunit B"epsilon |
|    | x  | AT1G04160 | XI-B | Myosin-8 |
| x  |    | AT1G04300 | TRAF1B | TNF receptor-associated factor homolog 1b |
| x  |    | AT1G04970 | LBR-1 | lipid-binding serum glycoprotein family protein |
| x  | x  | AT1G05230 | HDG2 | homeodomain GLABROUS 2 |
| x  |    | AT1G05540 | DOA1 | Protein of unknown function (DUF295) |
| x  | x  | AT1G05710 | - | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
|    | x  | AT1G05750 | PDE247 | Pentatricopeptide repeat-containing protein, chloroplastic |
|    | x  | AT1G05860 | - | INO80 complex subunit D-like protein |
| x  |    | AT1G05950 | - | hypothetical protein |
| x  |    | AT1G06150 | EMB1444 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
| x  |    | AT1G06630 | - | F-box/RNI-like superfamily protein |
| x  | x  | AT1G06710 | - | Tetratricopeptide repeat (TPR)-like superfamily protein |
| x  | x  | AT1G07119 | - | other RNA |
| x  | x  | AT1G07170 | - | PHD finger-like domain-containing protein 5A |
| x  |    | AT1G07350 | SR45A | Serine/arginine-rich splicing factor SR45a |
|    | x  | AT1G07780 | PAI1 | N-(5'-phosphoribosyl)anthranilate isomerase 1, chloroplastic |
|    | x  | AT1G08230 | GAT1 | GABA transporter 1 |
|    | x  | AT1G08320 | TGA9 | Transcription factor TGA9 |
| x  |    | AT1G08530 | - | Chitinase-like protein |
| x  |    | AT1G08680 | ZIGA4 | ARF GAP-like zinc finger-containing protein ZIGA4 |
|    | x  | AT1G08710 | SKIP24 | F-box protein SKIP24 |
| x  |    | AT1G08820 | VAP27-2 | vamp/synaptobrevin-associated protein 27-2 |
|    | x  | AT1G09023 | - | Natural antisense transcript overlaps with AT1G68940 |
|    | x  | AT1G09140 | SR30 | Serine/arginine-rich splicing factor SR30 |
| x  | x  | AT1G09195 | - | Ppx/GppA phosphatase |
| x  |    | AT1G09530 | PIF3 | Transcription factor PIF3 |
| x  |    | AT1G09710 | - | Homeodomain-like superfamily protein |
| x  | x  | AT1G09840 | ASK10 | Shaggy-related protein kinase kappa |
| x  | x  | AT1G09890 | - | Rhamnogalacturonate lyase family protein |
|    | x  | AT1G10240 | FRS11 | Protein FAR1-RELATED SEQUENCE 11 |
|    | x  | AT1G10450 | SNL6 | Paired amphipathic helix protein Sin3-like 6 |
| x  |    | AT1G10657 | - | Transmembrane protein |
|    | x  | AT1G10920 | LOV1 | Putative inactive disease susceptibility protein LOV1 |
| x  |    | AT1G11280 | - | S-locus lectin protein kinase family protein |
| x  |    | AT1G13000 | - | Transmembrane protein |
| x  |    | AT1G13450 | GT-1 | Trihelix transcription factor GT-1 |
| x  |    | AT1G13860 | QUL1 | Probable methyltransferase PMT4 |
| x  |    | AT1G15490 | - | Alpha/beta-Hydrolases superfamily protein |
|    | x  | AT1G16020 | CCZ1A | Vacuolar fusion protein CCZ1 homolog A |

| 1h | 5h | gene_id | gene_name | gene_description |
|----|----|---------|-----------|------------------|
| x |   | AT1G16560 | PGAP3B | Per1-like family protein |
| x | x | AT1G16930 | - | F-box/RNI-like/FBD-like domains-containing protein |
| x |   | AT1G17940 | - | Endosomal targeting BRO1-like domain-containing protein |
| x | x | AT1G18335 | - | Acyl-CoA N-acyltransferases (NAT) superfamily protein |
| x |   | AT1G18610 | - | Galactose oxidase/kelch repeat superfamily protein |
| x |   | AT1G18950 | - | DDT domain superfamily |
|   | x | AT1G19025 | - | DNA repair metallo-beta-lactamase family protein |
| x | x | AT1G19400 | - | Erythronate-4-phosphate dehydrogenase family protein |
|   | x | AT1G20920 | RH42 | DEAD-box ATP-dependent RNA helicase 42 |
|   | x | AT1G21360 | GLTP2 | Glycolipid transfer protein 2 |
|   | x | AT1G21610 | - | Wound-responsive family protein |
|   | x | AT1G22140 | - | zinc finger CCCH domain protein |
|   | x | AT1G25230 | - | Calcineurin-like metallo-phosphoesterase superfamily protein |
| x |   | AT1G25550 | HHO3 | Transcription factor HHO3 |
|   | x | AT1G26208 | - | other RNA |
|   | x | AT1G26230 | CPN60B4 | TCP-1/cpn60 chaperonin family protein, chloroplast |
| x |   | AT1G26440 | UPS5 | Ureide permease 5 |
|   | x | AT1G27460 | NPGR1 | Protein NPGR1 |
| x | x | AT1G27630 | CYCT1-3 | Cyclin-T1-3 |
| x |   | AT1G27921 | - | other RNA |
| x | x | AT1G28090 | - | Polynucleotide adenylyltransferase family protein |
| x | x | AT1G28395 | - | hypothetical protein |
|   | x | AT1G28610 | GGL3 | GDSL-like Lipase/Acylhydrolase superfamily protein |
| x |   | AT1G29170 | SCAR3 | Protein SCAR3 |
| x | x | AT1G29410 | PAI3 | N-(5'-phosphoribosyl)anthranilate isomerase 3, chloroplastic |
| x | x | AT1G31050 | - | bHLH DNA-binding superfamily protein |
| x |   | AT1G31360 | RECQL2 | RECQ helicase L2 |
| x |   | AT1G31500 | - | DNAse I-like superfamily protein |
| x |   | AT1G31600 | TRM9 | RNA-binding (RRM/RBD/RNP motifs) family protein |
| x | x | AT1G32740 | - | SBP (S-ribonuclease binding protein) family protein |
|   | x | AT1G33790 | JAL4 | Jacalin-related lectin 4 |
|   | x | AT1G34010 | - | unknown protein |
|   | x | AT1G34760 | GRF11 | general regulatory factor 11 |
| x |   | AT1G34844 | - | other RNA |
|   | x | AT1G35250 | ALT2 | Acyl-acyl carrier protein thioesterase ATL2, chloroplastic |
| x | x | AT1G36070 | - | Transducin/WD40 repeat-like superfamily protein |
|   | x | AT1G42540 | GLR3.3 | Glutamate receptor 3.3 |
|   | x | AT1G43620 | UGT80B1 | UDP-glucose:sterol-glucosyltransferase |
| x |   | AT1G43650 | UMAMIT22 | nodulin MtN21-like transporter family protein |
|   | x | AT1G43770 | - | RING/FYVE/PHD zinc finger superfamily protein |
| x | x | AT1G44125 | - | Natural antisense gene, locus overlaps with AT1G44120 |
| x |   | AT1G45248 | - | Nucleolar histone methyltransferase-related protein |
|   | x | AT1G45249 | ABF2 | abscisic acid responsive elements-binding factor 2 |
|   | x | AT1G48030 | LPD1 | Dihydrolipoyl dehydrogenase 1, mitochondrial |
|   | x | AT1G48175 | TAD2 | tRNA-specific adenosine deaminase TAD2 |
| x | x | AT1G48360 | FAN1 | Fanconi-associated nuclease 1 homolog |
| x |   | AT1G48500 | TIFY6A | Protein TIFY 6A |
|   | x | AT1G49360 | - | F-box protein family |
| x | x | AT1G49980 | POLK | DNA polymerase kappa |
| x |   | AT1G50300 | TAF15 | Transcription initiation factor TFIID subunit 15 |
|   | x | AT1G50410 | CHR28 | Helicase-like transcription factor CHR28 |
|   | x | AT1G50700 | CPK33 | calcium-dependent protein kinase 33 |
|   | x | AT1G50730 | - | hypothetical protein |
|   | x | AT1G50840 | POLGAMMA2 | polymerase gamma 2 |
|   | x | AT1G50970 | - | Vacuolar protein sorting-associated protein 53 B |
| x |   | AT1G51270 | - | Vesicle-associated protein 1-4 |
| x |   | AT1G51840 | - | protein kinase-related |

| 1h | 5h | gene_id | gene_name | gene_description |
|----|----|---------|-----------|------------------|
| x | | AT1G52370 | - | Ribosomal protein L22p/L17e family protein |
| x | | AT1G52630 | OFUT13 | O-fucosyltransferase 13 |
| x | | AT1G53165 | ATMAP4K ALPHA1 | Protein kinase superfamily protein |
| | x | AT1G53570 | MAPKKK3 | Mitogen-activated protein kinase kinase kinase 3 |
| | x | AT1G54310 | - | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein |
| x | x | AT1G54730 | - | Sugar transporter ERD6-like 5 |
| x | | AT1G55110 | IDD7 | Protein indeterminate-domain 7 |
| x | x | AT1G55350 | DEK1 | Calpain-type cysteine protease family |
| x | | AT1G55500 | ECT4 | evolutionarily conserved C-terminal region 4 |
| x | | AT1G55610 | BRL1 | Serine/threonine-protein kinase BRI1-like 1 |
| | x | AT1G55870 | PARN | Poly(A)-specific ribonuclease PARN |
| x | | AT1G56612 | - | other RNA |
| | x | AT1G57870 | ATSK42 | Shaggy-like kinase 42 |
| x | | AT1G58025 | - | DNA-binding bromodomain-containing protein |
| x | | AT1G58180 | BCA6 | Beta carbonic anhydrase 6, mitochondrial |
| | x | AT1G58350 | ZW18 | Putative serine esterase family protein |
| x | | AT1G59520 | CW7 | CW7 |
| x | | AT1G59890 | SNL5 | SIN3-like 5 |
| x | | AT1G60200 | - | Splicing factor PWI domain-containing protein / RNA recognition motif (RRM)-containing protein |
| | x | AT1G60430 | ARPC3 | Actin-related protein 2/3 complex subunit 3 |
| x | x | AT1G60505 | - | other RNA |
| x | | AT1G60545 | - | other RNA |
| x | | AT1G60640 | - | stress response protein |
| x | | AT1G60850 | ATRPAC42 | DNA-directed RNA polymerase family protein |
| x | x | AT1G61010 | CPSF73-I | Cleavage and polyadenylation specificity factor subunit 3-I |
| | x | AT1G61040 | VIP5 | VIP5 |
| x | | AT1G61360 | - | Serine/threonine-protein kinase |
| | x | AT1G61430 | - | G-type lectin S-receptor-like serine/threonine-protein kinase |
| x | | AT1G61970 | mTERF29 | Mitochondrial transcription termination factor family protein |
| x | x | AT1G63260 | TET10 | Tetraspanin-10 |
| x | x | AT1G63480 | AHL12 | AT-hook motif nuclear-localized protein 12 |
| x | x | AT1G65950 | - | Protein kinase superfamily protein |
| | x | AT1G66730 | LIG6 | DNA ligase 6 |
| x | x | AT1G66840 | PMI2 | plastid movement impaired 2 |
| x | | AT1G67040 | TRM22 | DnaA initiator-associating protein |
| | x | AT1G67792 | - | other RNA |
| x | x | AT1G67850 | - | lysine ketoglutarate reductase trans-splicing protein (DUF707) |
| | x | AT1G68080 | - | 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein |
| x | x | AT1G68872 | - | other RNA |
| x | x | AT1G69310 | WRKY57 | WRKY DNA-binding protein 57 |
| x | | AT1G70490 | ARF2-A | ADP-ribosylation factor 2-B |
| x | | AT1G70895 | CLE17 | CLAVATA3/ESR (CLE)-related protein 17 |
| | x | AT1G71220 | EBS1 | UDP-glucose:glycoprotein glucosyltransferases |
| | x | AT1G72560 | PSD | Exportin-T |
| | x | AT1G72770 | HAB1 | Protein phosphatase 2C 16 |
| | x | AT1G72855 | - | other RNA |
| | x | AT1G72990 | BGAL17 | Beta-galactosidase 17 |
| | x | AT1G73170 | - | P-loop containing nucleoside triphosphate hydrolases superfamily protein |
| x | | AT1G73310 | SCPL4 | Serine carboxypeptidase-like 4 |
| | x | AT1G73350 | - | Ankyrin repeat protein |
| | x | AT1G73650 | - | Protein of unknown function (DUF1295) |
| | x | AT1G73970 | - | obscurin-like protein |
| x | x | AT1G75295 | - | other RNA |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x | x | AT1G76170 | - | 2-thiocytidine tRNA biosynthesis protein, TtcA |
| x |  | AT1G76620 | - | Serine/Threonine-kinase, putative (DUF547) |
| x |  | AT1G77220 | - | Protein LAZ1 homolog 1 |
|  | x | AT1G77290 | TCHQD | Glutathione S-transferase TCHQD |
| x |  | AT1G77460 | CSI3 | Protein CELLULOSE SYNTHASE INTERACTIVE 3 |
| x | x | AT1G77800 | - | PHD finger family protein |
|  | x | AT1G78200 | - | Probable protein phosphatase 2C 17 |
| x | x | AT1G78690 | - | N-acylphosphatidylethanolamine synthase |
|  | x | AT1G78810 | - | hypothetical protein |
| x |  | AT1G78910 | - | RNA pseudouridine synthase 3, mitochondrial |
|  | x | AT1G79270 | ECT8 | Evolutionarily conserved C-terminal region 8 |
|  | x | AT1G79630 | - | Probable protein phosphatase 2C 18 |
| x |  | AT1G79640 | - | Protein kinase superfamily protein |
| x | x | AT1G79950 | RTEL1 | Regulator of telomere elongation helicase 1 homolog |
| x | x | AT1G80245 | - | Spc97 / Spc98 family of spindle pole body (SBP) component |
| x |  | AT1G80360 | ISS1 | Aromatic aminotransferase ISS1 |
| x |  | AT1G80650 | RTL1 | Ribonuclease 3-like protein 1 |
|  | x | AT1G80810 | - | Tudor/PWWP/MBT superfamily protein |
| x |  | AT1G80860 | PLMT | Phosphatidyl-N-methylethanolamine N-methyltransferase |
|  | x | AT2G01830 | AHK4 | Histidine kinase 4 |
| x |  | AT2G01930 | BPC1 | Protein BASIC PENTACYSTEINE1 |
| x |  | AT2G02148 | - | PPR containing protein |
|  | x | AT2G02480 | STI | Protein STICHEL |
|  | x | AT2G02710 | TLP1 | Protein TWIN LOV 1 |
| x |  | AT2G03640 | - | Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain-containing protein |
|  | x | AT2G03810 | - | 18S pre-ribosomal assembly protein gar2-like protein |
|  | x | AT2G04100 | DTX6 | Protein DETOXIFICATION 6 |
|  | x | AT2G04270 | RNE | Ribonuclease E/G-like protein, chloroplastic |
|  | x | AT2G04690 | - | Pyridoxamine 5'-phosphate oxidase family protein |
| x | x | AT2G04790 | - | PTB domain engulfment adapter |
| x |  | AT2G04800 | - | hypothetical protein |
| x |  | AT2G04845 | - | Acyl-CoA N-acyltransferases (NAT) superfamily protein |
| x | x | AT2G05210 | POT1A | Protection of telomeres protein 1a |
| x |  | AT2G06005 | FIP1 | FRIGIDA interacting protein 1 |
| x | x | AT2G06025 | - | Acyl-CoA N-acyltransferases (NAT) superfamily protein |
| x | x | AT2G09795 | - | other RNA |
| x | x | AT2G15000 | - | Caspase-6 protein |
|  | x | AT2G15050 | LTP7 | Non-specific lipid-transfer protein |
| x | x | AT2G15530 | - | RING/U-box superfamily protein |
| x |  | AT2G16920 | UBC23 | Probable ubiquitin-conjugating enzyme E2 23 |
| x | x | AT2G16940 | - | Splicing factor, CC1-like protein |
| x | x | AT2G16990 | - | Major facilitator superfamily protein |
| x | x | AT2G17150 | NLP1 | Protein NLP1 |
| x |  | AT2G17320 | - | pantothenate kinase, uncharacterized conserved protein (UCP030210) |
| x | x | AT2G17780 | MCA2 | Protein MID1-COMPLEMENTING ACTIVITY 2 |
| x | x | AT2G17970 | ALKBH9B | RNA demethylase ALKBH9B |
| x |  | AT2G18100 | - | transmembrane/coiled-coil protein (DUF726) |
| x |  | AT2G18750 | CBP60C | Calmodulin-binding protein 60 C |
| x |  | AT2G18760 | CHR8 | chromatin remodeling 8 |
|  | x | AT2G19390 | - | Expressed protein |
| x | x | AT2G20585 | NFD6 | Nuclear fusion defective 6 |
| x | x | AT2G20815 | QWRF3 | QWRF motif protein (DUF566) |
|  | x | AT2G20950 | - | Arabidopsis phospholipase-like protein (PEARLI 4) family |
| x |  | AT2G21710 | MTERF2 | Transcription termination factor MTERF2, chloroplastic |
|  | x | AT2G22260 | - | oxidoreductase, 2OG-Fe(II) oxygenase family protein |
| x |  | AT2G22980 | SCPL13 | Serine carboxypeptidase-like 13 |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| | x | AT2G23040 | - | long noncoding RNA |
| x | | AT2G23910 | - | NAD(P)-binding Rossmann-fold superfamily protein |
| x | x | AT2G23985 | - | hypothetical protein |
| x | | AT2G24650 | REM13 | B3 domain-containing protein REM13 |
| x | x | AT2G25350 | - | Phox (PX) domain-containing protein |
| | x | AT2G25480 | - | TPX2 (targeting protein for Xklp2) protein family |
| x | | AT2G25670 | - | hypothetical protein |
| | x | AT2G25850 | PAPS2 | poly(A) polymerase 2 |
| | x | AT2G26210 | - | Ankyrin repeat family protein |
| | x | AT2G26590 | RPN13 | regulatory particle non-ATPase 13 |
| x | x | AT2G26770 | SCAB1 | Stomatal closure-related actin-binding protein 1 |
| | x | AT2G28760 | UXS6 | UDP-XYL synthase 6 |
| x | x | AT2G29140 | APUM3 | Pumilio homolog 3 |
| x | | AT2G29910 | - | F-box/RNI-like superfamily protein |
| | x | AT2G30120 | FLX | protein FLC EXPRESSOR |
| x | | AT2G30460 | UXT2 | UDP-xylose transporter 2 |
| x | | AT2G30695 | - | bacterial trigger factor |
| x | x | AT2G30740 | PTI12 | PTI1-like tyrosine-protein kinase 2 |
| x | | AT2G31130 | - | hypothetical protein |
| | x | AT2G31960 | CALS2 | Callose synthase 2 |
| x | | AT2G32000 | - | DNA topoisomerase, type IA, core |
| | x | AT2G32120 | HSP70-8 | Heat shock 70 kDa protein 8 |
| | x | AT2G32235 | - | hypothetical protein |
| x | x | AT2G32250 | FRS2 | Protein FAR1-RELATED SEQUENCE 2 |
| x | x | AT2G32415 | RRP6L3 | Protein RRP6-like 3 |
| x | | AT2G32640 | - | Lycopene beta/epsilon cyclase protein |
| x | x | AT2G32760 | USL1 | UV radiation resistance-associated protein |
| | x | AT2G33205 | - | Serinc-domain containing serine and sphingolipid biosynthesis protein |
| | x | AT2G33500 | COL14 | Zinc finger protein CONSTANS-LIKE 14 |
| x | | AT2G33540 | CPL3 | RNA polymerase II C-terminal domain phosphatase-like 3 |
| x | x | AT2G34310 | - | Expressed protein |
| x | | AT2G35940 | BLH1 | BEL1-like homeodomain protein 1 |
| x | x | AT2G37150 | - | RING/U-box superfamily protein |
| | x | AT2G38460 | IREG1 | Solute carrier family 40 member 1 |
| | x | AT2G38660 | FOLD1 | Bifunctional protein FolD 1, mitochondrial |
| x | | AT2G38780 | - | Cytochrome C oxidase subunit |
| | x | AT2G38880 | NF-YB1 | Nuclear factor Y, subunit B1 |
| | x | AT2G39280 | - | Ypt/Rab-GAP domain of gyp1p superfamily protein |
| x | x | AT2G39690 | - | ternary complex factor MIP1 leucine-zipper protein |
| x | | AT2G39780 | RNS2 | Ribonuclease 2 |
| x | x | AT2G39805 | - | Integral membrane Yip1 family protein |
| x | | AT2G39890 | PROT1 | Proline transporter 1 |
| x | x | AT2G39950 | - | flocculation protein |
| x | | AT2G40630 | - | Uncharacterised conserved protein (UCP030365) |
| x | x | AT2G40640 | PUB63 | U-box domain-containing protein 63 |
| x | | AT2G40690 | GLY1 | Glycerol-3-phosphate dehydrogenase |
| x | x | AT2G40800 | - | Import inner membrane translocase subunit |
| x | | AT2G40830 | RHC1A | Probable E3 ubiquitin-protein ligase RHC1A |
| x | | AT2G41150 | - | plant/protein |
| | x | AT2G41350 | AUG1 | AUGMIN subunit 1 |
| x | x | AT2G41880 | GK-1 | guanylate kinase 1 |
| | x | AT2G42170 | - | Actin family protein |
| x | x | AT2G42810 | PAPP5 | Serine/threonine-protein phosphatase 5 |
| x | | AT2G43465 | - | RNA-binding ASCH domain protein |
| x | | AT2G43490 | - | Ypt/Rab-GAP domain of gyp1p superfamily protein |
| | x | AT2G43500 | - | Plant regulator RWP-RK family protein |
| | x | AT2G43760 | MOCS2 | Molybdopterin synthase catalytic subunit |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x | x | AT2G43920 | HOL2 | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein |
| x | x | AT2G44090 | - | Ankyrin repeat family protein |
|  | x | AT2G44798 | - | other RNA |
| x |  | AT2G44900 | FBX5 | Protein ARABIDILLO 1 |
|  | x | AT2G45460 | - | SMAD/FHA domain-containing protein |
| x |  | AT2G45500 | - | AAA-type ATPase family protein |
| x |  | AT2G45540 | BCHC2 | BEACH domain-containing protein C2 |
| x |  | AT2G45830 | DTA2 | Downstream target of AGL15 2 |
| x |  | AT2G46560 | - | Transducin family protein / WD-40 repeat family protein |
| x | x | AT2G46572 | - | other RNA |
| x | x | AT2G46610 | RS31A | Serine/arginine-rich splicing factor RS31A |
| x |  | AT2G46790 | APRR9 | Two-component response regulator-like APRR9 |
| x | x | AT2G46810 | BHLH70 | Transcription factor bHLH70 |
| x |  | AT2G46915 | - | Protein of unknown function (DUF3754) |
|  | x | AT2G47410 | - | WD40/YVTN repeat-like-containing domain;Bromodomain |
| x |  | AT2G47860 | - | Phototropic-responsive NPH3 family protein |
| x |  | AT2G48020 | - | Sugar transporter ERD6-like 7 |
| x |  | AT3G01310 | - | Phosphoglycerate mutase-like family protein |
| x | x | AT3G01770 | GTE11 | Transcription factor GTE11 |
| x | x | AT3G01990 | ACR6 | ACT domain repeat 6 |
| x |  | AT3G02150 | TCP13 | Transcription factor TCP13 |
| x | x | AT3G02170 | LNG2 | Protein LONGIFOLIA 2 |
| x | x | AT3G02290 | - | RING/U-box superfamily protein |
|  | x | AT3G02600 | LPP3 | Putative lipid phosphate phosphatase 3, chloroplastic |
| x |  | AT3G02860 | ARS1 | nucleus zinc ion binding protein, is translocated to the cytoplasm in response to ABA or oxidative stress |
|  | x | AT3G03120 | ATARFB1C | ADP-ribosylation factor B1C |
| x | x | AT3G03970 | SINE2 | Protein SINE2 |
|  | x | AT3G04500 | - | Putative RRM-containing protein |
|  | x | AT3G05250 | - | RING/U-box superfamily protein |
| x |  | AT3G05270 | FPP3 | Filament-like plant protein 3 |
| x |  | AT3G05365 | - | hypothetical protein |
| x | x | AT3G05750 | - | Serine-rich adhesin for platelets-like protein |
|  | x | AT3G05920 | HIPP43 | Heavy metal-associated isoprenylated plant protein 43 |
|  | x | AT3G06190 | BPM2 | BTB/POZ and MATH domain-containing protein 2 |
| x |  | AT3G06250 | FRS7 | Protein FAR1-RELATED SEQUENCE 7 |
| x |  | AT3G06630 | - | Protein kinase family protein |
|  | x | AT3G06880 | - | Transducin/WD40 repeat-like superfamily protein |
|  | x | AT3G07360 | PUB9 | RING-type E3 ubiquitin transferase |
|  | x | AT3G07890 | - | Ypt/Rab-GAP domain of gyp1p superfamily protein |
|  | x | AT3G08710 | TRX9 | Thioredoxin H9 |
| x | x | AT3G08840 | - | D-alanine-D-alanine ligase family |
| x |  | AT3G09540 | - | Pectate lyase |
| x |  | AT3G09745 | - | Natural antisense transcript overlaps with AT3G62090 |
| x | x | AT3G09800 | - | SNARE-like superfamily protein |
| x | x | AT3G09920 | PIP5K9 | Phosphatidylinositol 4-phosphate 5-kinase 9 |
|  | x | AT3G10250 | - | histidine-tRNA ligase |
|  | x | AT3G10500 | NAC053 | NAC domain-containing protein 53 |
|  | x | AT3G10572 | APEM9 | Protein APEM9 |
|  | x | AT3G11240 | ATE2 | Arginyl-tRNA--protein transferase 2 |
| x |  | AT3G11470 | - | 4'-phosphopantetheinyl transferase superfamily |
| x |  | AT3G11820 | SYP121 | syntaxin of plants 121 |
| x |  | AT3G12250 | TGA6 | TGACG motif-binding factor 6 |
| x |  | AT3G12977 | - | NAC domain transcriptional regulator superfamily protein |
| x | x | AT3G13060 | ECT5 | evolutionarily conserved C-terminal region 5 |
| x | x | AT3G13190 | - | WEB family protein (DUF827) |
| x |  | AT3G13420 | - | Transmembrane protein |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x |  | AT3G13800 | - | Metallo-hydrolase/oxidoreductase superfamily protein |
| x |  | AT3G14150 | GLO3 | Aldolase-type TIM barrel family protein |
|  | x | AT3G14880 | - | DNA-binding protein-like protein |
| x |  | AT3G14980 | IDM1 | Increased DNA methylation 1 |
|  | x | AT3G15395 | - | HrpN-interacting protein from malus protein |
| x | x | AT3G15970 | NUP50B | Nuclear pore complex protein NUP50B |
|  | x | AT3G15980 | - | Coatomer subunit beta |
| x |  | AT3G16030 | CES101 | Lectin protein kinase family protein |
|  | x | AT3G16800 | - | Probable protein phosphatase 2C 41 |
|  | x | AT3G17310 | DRM3 | Encodes a mutated Domains Rearranged Methyltransferase 3, required for normal maintenance of non-CG DNA methylation |
|  | x | AT3G17609 | HYH | Encodes a homolog of HY5 (HYH). Involved in phyB signaling pathway |
|  | x | AT3G17950 | - | transmembrane protein |
|  | x | AT3G18500 | - | DNAse I-like superfamily protein |
| x |  | AT3G19150 | KRP6 | Cyclin-dependent kinase inhibitor |
| x |  | AT3G19330 | - | Transmembrane protein, putative (DUF677) |
| x |  | AT3G19370 | - | Filament-like protein (DUF869) |
|  | x | AT3G19780 | DUF179 | Interacts with chloroplast chaperone CLPC1, which unfolds and delivers substrates to the stromal CLPPRT protease complex for degradation. |
| x | x | AT3G20070 | TTN9 | titan9 |
| x | x | AT3G20280 | - | RING/FYVE/PHD zinc finger superfamily protein |
| x |  | AT3G20810 | JMJ30 | Lysine-specific demethylase JMJ30 |
|  | x | AT3G21175 | GATA24 | GATA transcription factor 24 |
| x |  | AT3G21810 | - | Zinc finger CCCH domain-containing protein 40 |
| x |  | AT3G22190 | IQD5 | IQ-domain 5 |
|  | x | AT3G22220 | - | HAT transposon superfamily |
| x |  | AT3G23200 | - | CASP-like protein 5B3 |
|  | x | AT3G23255 | - | tRNA dimethylallyltransferase |
| x |  | AT3G23610 | DSPTP1 | dual specificity protein phosphatase 1 |
| x |  | AT3G24740 | - | cellulose synthase, putative (DUF1644) |
| x |  | AT3G24840 | - | Sec14p-like phosphatidylinositol transfer family protein |
| x |  | AT3G24929 | - | hypothetical protein |
| x | x | AT3G26020 | - | Serine/threonine protein phosphatase 2A regulatory subunit |
| x | x | AT3G26115 | - | D-cysteine desulfhydrase 2, mitochondrial |
| x |  | AT3G26490 | - | BTB/POZ domain-containing protein At3g26490 |
| x | x | AT3G26510 | - | Octicosapeptide/Phox/Bem1p family protein |
| x | x | AT3G26670 | - | Probable magnesium transporter |
| x | x | AT3G26700 | - | Protein kinase superfamily protein |
|  | x | AT3G26890 | - | meiosis chromosome segregation family protein |
| x |  | AT3G27110 | - | Peptidase family M48 family protein |
|  | x | AT3G27580 | D6PKL3 | Serine/threonine-protein kinase D6PKL3 |
|  | x | AT3G27610 | - | Nucleotidylyl transferase superfamily protein |
| x |  | AT3G27990 | - | other RNA |
| x | x | AT3G28130 | UMAMIT44 | nodulin MtN21 /EamA-like transporter family protein |
| x |  | AT3G29100 | VTI13 | Vesicle transport v-SNARE 13 |
|  | x | AT3G29160 | KIN11 | Non-specific serine/threonine protein kinase |
| x |  | AT3G29330 | - | Zinc finger RNA-binding-like protein |
| x | x | AT3G29644 | - | other RNA |
|  | x | AT3G32940 | - | RNA-binding KH domain-containing protein |
|  | x | AT3G42150 | - | transmembrane protein |
| x |  | AT3G44530 | HIRA | homolog of histone chaperone HIRA |
| x |  | AT3G46200 | NUDT9 | Nudix hydrolase 9 |
|  | x | AT3G46658 | - | other RNA |
| x |  | AT3G46930 | - | Protein kinase superfamily protein |
| x |  | AT3G47560 | - | Alpha/beta-Hydrolases superfamily protein |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x | x | AT3G47630 | - | translocator assembly/maintenance protein |
| x | x | AT3G47910 | - | Ubiquitin carboxyl-terminal hydrolase-related protein |
| | x | AT3G48190 | ATM | Serine/Threonine-kinase ATM-like protein |
| x | | AT3G48360 | BT2 | BTB/POZ and TAZ domain-containing protein 2 |
| | x | AT3G49430 | SR34A | Serine/arginine-rich splicing factor SR34A |
| x | | AT3G49640 | - | Aldolase-type TIM barrel family protein |
| x | | AT3G50240 | KIN4B | Kinesin-like protein KIN-4B |
| x | x | AT3G51075 | - | other RNA |
| | x | AT3G51360 | - | Eukaryotic aspartyl protease family protein |
| | x | AT3G51640 | - | Stress response NST1-like protein |
| | x | AT3G52050 | - | 5'-3' exonuclease family protein |
| x | | AT3G52250 | - | Duplicated homeodomain-like superfamily protein |
| | x | AT3G52340 | SPP2 | Sucrose-6F-phosphate phosphohydrolase 2 |
| x | | AT3G52720 | ACA1 | Alpha carbonic anhydrase 1, chloroplastic |
| x | | AT3G53270 | - | Small nuclear RNA activating complex (SNAPc), subunit SNAP43 protein |
| x | x | AT3G53570 | AFC1 | serine/threonine-protein kinase AFC1 |
| | x | AT3G53830 | - | Regulator of chromosome condensation (RCC1) family protein |
| x | | AT3G54230 | SUA | suppressor of abi3-5 |
| x | | AT3G54510 | - | Hyperosmolality-gated Ca2+ permeable channel 2.5 |
| x | x | AT3G54750 | - | Downstream neighbor of Son |
| | x | AT3G55080 | - | SET domain-containing protein |
| x | x | AT3G55850 | LAF3 | Protein LONG AFTER FAR-RED 3 |
| | x | AT3G56720 | - | pre-mRNA-splicing factor |
| | x | AT3G56730 | - | Putative endonuclease or glycosyl hydrolase |
| | x | AT3G56830 | - | YCF20-like protein (DUF565) |
| | x | AT3G56860 | UBA2A | UBP1-associated protein 2A |
| | x | AT3G57680 | CTPA3 | Carboxyl-terminal-processing peptidase 3, chloroplastic |
| x | | AT3G57880 | - | Anthranilate phosphoribosyltransferase-like protein |
| | x | AT3G58160 | XI-J | Myosin-16 |
| x | | AT3G58930 | - | F-box/RNI-like superfamily protein |
| x | x | AT3G59210 | - | F-box/RNI-like superfamily protein |
| x | x | AT3G59320 | - | Anthocyanin-related membrane protein 2 |
| | x | AT3G59330 | - | Solute carrier family 35 protein (DUF914) |
| x | | AT3G59470 | - | Far-red impaired responsive (FAR1) family protein |
| | x | AT3G59490 | - | hypothetical protein |
| x | | AT3G59810 | LSM6A | Sm-like protein LSM6A |
| x | | AT3G60070 | - | AT3g60070/T2O9_50 |
| x | | AT3G60910 | - | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein |
| | x | AT3G61610 | - | Glucose-6-phosphate 1-epimerase |
| x | | AT3G62390 | TBL6 | Protein trichome birefringence-like 6 |
| | x | AT3G62620 | - | At3g62620 |
| | x | AT3G62940 | - | Cysteine proteinases superfamily protein |
| | x | AT3G63180 | ATTKL | TIC-like protein |
| | x | AT3G63450 | - | RNA-binding (RRM/RBD/RNP motifs) family protein |
| | x | AT4G00440 | - | GPI-anchored adhesin-like protein, putative (DUF3741) |
| x | | AT4G00520 | - | Acyl-CoA thioesterase family protein |
| x | | AT4G00560 | - | NAD(P)-binding Rossmann-fold superfamily protein |
| x | | AT4G00680 | ADF8 | actin depolymerizing factor 8 |
| x | | AT4G00990 | - | Transcription factor jumonji (JmjC) domain-containing protein |
| x | | AT4G01280 | RVE5 | Protein REVEILLE 5 |
| | x | AT4G01510 | ARV2 | Arv1-like protein |
| x | x | AT4G01915 | - | hypothetical protein |
| x | | AT4G02075 | PIT1 | RING/FYVE/PHD zinc finger superfamily protein |
| x | | AT4G02210 | - | Myb/SANT-like DNA-binding domain protein |
| | x | AT4G02430 | SR34b | RNA-binding (RRM/RBD/RNP motifs) family protein |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| | x | AT4G02460 | PMS1 | DNA mismatch repair protein PMS1 |
| x | | AT4G02880 | - | ELKS/Rab6-interacting/CAST family protein |
| | x | AT4G03000 | RF298 | Putative E3 ubiquitin-protein ligase RF298 |
| | x | AT4G03230 | - | S-locus lectin protein kinase family protein |
| x | | AT4G04221 | - | other RNA |
| | x | AT4G04570 | CRK40 | Cysteine-rich receptor-like protein kinase 40 |
| x | | AT4G04740 | CPK23 | Calcium-dependent protein kinase 23 |
| x | x | AT4G04750 | - | Major facilitator superfamily protein |
| | x | AT4G04790 | - | Pentatricopeptide repeat-containing protein At4g04790, mitochondrial |
| | x | AT4G04880 | - | adenosine/AMP deaminase family protein |
| x | | AT4G05040 | - | ankyrin repeat family protein |
| x | | AT4G05590 | MCP2L | Encodes NRGA1, a putative mitochondrial pyruvate carrier that mediates ABA regulation of guard cell ion channels and drought stress responses. |
| x | | AT4G08170 | ITPK3 | Inositol-tetrakisphosphate 1-kinase 3 |
| x | x | AT4G08460 | - | hypothetical protein |
| | x | AT4G08470 | MAPKKK10 | Mitogen-activated protein kinase kinase kinase 3 |
| x | | AT4G09920 | - | F-box protein At4g09920 |
| x | | AT4G10170 | PHYL1.1 | Phytolongin Phyl1.1 |
| x | x | AT4G10360 | - | TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein |
| x | x | AT4G10430 | - | TMPIT-like protein |
| x | x | AT4G11830 | PLDGAMMA2 | Phospholipase D gamma 2 |
| x | | AT4G11840 | PLDGAMMA3 | Phospholipase D gamma 3 |
| x | | AT4G11900 | - | G-type lectin S-receptor-like serine/threonine-protein kinase At4g11900 |
| x | | AT4G11970 | - | YTH family protein |
| x | | AT4G12020 | WRKY19 | Probable WRKY transcription factor 19 |
| | x | AT4G12460 | ORP2B | OSBP(oxysterol binding protein)-related protein 2B |
| | x | AT4G12560 | CPR1 | F-box protein CPR1 |
| | x | AT4G13030 | - | P-loop containing nucleoside triphosphate hydrolases superfamily protein |
| | x | AT4G13040 | - | Integrase-type DNA-binding superfamily protein |
| x | | AT4G13330 | - | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein |
| x | x | AT4G14385 | - | Histone acetyltransferase subunit NuA4-domain protein |
| | x | AT4G14410 | BHLH104 | Transcription factor bHLH104 |
| x | x | AT4G14520 | NRPB7L | DNA-directed RNA polymerase subunit 7-like protein |
| x | x | AT4G15090 | FAR1 | Protein FAR-RED IMPAIRED RESPONSE 1 |
| x | | AT4G15563 | - | F-box-like protein |
| | x | AT4G15765 | - | FAD/NAD(P)-binding oxidoreductase family protein |
| | x | AT4G16420 | ADA2B | Transcriptional adapter ADA2b |
| x | x | AT4G16710 | - | glycosyltransferase family protein 28 |
| x | x | AT4G16765 | - | 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein |
| x | x | AT4G16920 | - | Disease resistance protein (TIR-NBS-LRR class) family |
| x | x | AT4G16990 | RLM3 | disease resistance protein (TIR-NBS class), putative |
| x | x | AT4G17310 | - | hypothetical protein |
| | x | AT4G17410 | - | DWNN domain, a CCHC-type zinc finger |
| | x | AT4G18070 | - | suppressor |
| | x | AT4G18140 | SSP4b | SCP1-like small phosphatase 4b |
| | x | AT4G18600 | WAVE5 | Scar-like domain-containing protein WAVE 5 |
| x | x | AT4G18975 | - | Pentatricopeptide repeat-containing protein, chloroplastic |
| x | | AT4G19160 | - | transglutaminase family protein |
| | x | AT4G19440 | - | Pentatricopeptide repeat-containing protein, chloroplastic |
| | x | AT4G19670 | - | RBR-type E3 ubiquitin transferase |
| | x | AT4G19870 | - | F-box/kelch-repeat protein |
| x | x | AT4G20310 | S2P | Site-2 protease |

| 1h | 5h | gene_id | gene_name | gene_description |
|----|----|---------|-----------|------------------|
|    | x  | AT4G20320 | - | CTP synthase |
| x  | x  | AT4G20380 | LSD1 | LSD1 zinc finger family protein |
| x  |    | AT4G21310 | - | transmembrane protein, putative (DUF1218) |
| x  | x  | AT4G21400 | CRK28 | Cysteine-rich receptor-like protein kinase 28 |
| x  | x  | AT4G21585 | ENDO4 | endonuclease 4 |
| x  |    | AT4G22120 | CSC1 | Hyperosmolality-gated Ca2+ permeable channel 1.2 |
|    | x  | AT4G22990 | - | Major Facilitator Superfamily with SPX (SYG1/Pho81/XPR1) domain-containing protein |
| x  |    | AT4G23000 | - | Calcineurin-like metallo-phosphoesterase superfamily protein |
| x  |    | AT4G23250 | CRK17 | cysteine-rich receptor-like protein kinase 17 |
| x  |    | AT4G23330 | - | hypothetical protein |
|    | x  | AT4G24180 | TLP1 | THAUMATIN-LIKE PROTEIN 1 |
|    | x  | AT4G24460 | CLT2 | Protein CLT2, chloroplastic |
|    | x  | AT4G24610 | - | pesticidal crystal cry8Ba protein |
|    | x  | AT4G24680 | MOS1 | modifier of snc1 |
|    | x  | AT4G24740 | AFC2 | LAMMER-type protein kinase AFC2 |
| x  |    | AT4G25315 | - | expressed protein |
| x  | x  | AT4G25434 | ATNUDT10 | nudix hydrolase homolog 10 |
|    | x  | AT4G25440 | ZFWD1 | Zinc finger CCCH domain-containing protein 48 |
|    | x  | AT4G25500 | RS40 | Serine/arginine-rich splicing factor RS40 |
| x  | x  | AT4G26140 | BGAL12 | beta-galactosidase 12 |
| x  | x  | AT4G26600 | - | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein |
| x  | x  | AT4G27040 | VP22-1 | Vacuolar protein sorting-associated protein |
| x  | x  | AT4G27610 | - | Intracellular protein transporter |
| x  |    | AT4G28110 | AtMYB41 | MYB41 |
|    | x  | AT4G28760 | - | Methyl-coenzyme M reductase II subunit gamma, putative (DUF3741) |
| x  | x  | AT4G28790 | BHLH23 | basic helix-loop-helix DNA-binding superfamily protein |
| x  |    | AT4G28820 | - | HIT-type Zinc finger family protein |
|    | x  | AT4G30340 | DGK7 | Diacylglycerol kinase 7 |
| x  | x  | AT4G30560 | CNGC9 | Putative cyclic nucleotide-gated ion channel 9 |
| x  |    | AT4G30630 | - | hypothetical protein |
| x  | x  | AT4G30820 | - | cyclin-dependent kinase-activating kinase assembly factor-related / CDK-activating kinase assembly factor-related |
| x  |    | AT4G31150 | - | endonuclease V family protein |
| x  |    | AT4G31170 | - | Protein kinase superfamily protein |
| x  |    | AT4G31200 | - | SWAP (Suppressor-of-White-APricot)/surp RNA-binding domain-containing protein |
| x  |    | AT4G31240 | NRX2 | protein kinase C-like zinc finger protein |
| x  | x  | AT4G31480 | - | Coatomer, beta subunit |
| x  | x  | AT4G31870 | GPX7 | glutathione peroxidase 7, chloroplastic |
|    | x  | AT4G31890 | - | ARM repeat superfamily protein |
| x  |    | AT4G32000 | - | Protein kinase superfamily protein |
| x  | x  | AT4G32360 | MFDR | NADPH:adrenodoxin oxidoreductase, mitochondrial |
| x  | x  | AT4G32850 | nPAP | nuclear poly(a) polymerase |
| x  | x  | AT4G33000 | CBL10 | Calcineurin B-like 10 |
| x  |    | AT4G33200 | XI-I | myosin, putative |
| x  |    | AT4G33740 | - | Myb-like protein X |
| x  | x  | AT4G34140 | - | D111/G-patch domain-containing protein |
| x  | x  | AT4G34310 | - | Alpha/beta-Hydrolases superfamily protein |
|    | x  | AT4G34570 | THY-2 | Bifunctional dihydrofolate reductase-thymidylate synthase 2 |
| x  |    | AT4G34840 | MTN2 | 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase 2 |
| x  |    | AT4G34900 | XDH2 | Xanthine dehydrogenase 2 |
| x  | x  | AT4G35300 | TMT2 | tonoplast monosaccharide transporter2 |
| x  |    | AT4G35770 | SEN1 | Senescence-associated gene 1 |
| x  |    | AT4G36050 | APE2 | DNA-(apurinic or apyrimidinic site) lyase 2 |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x |  | AT4G36380 | ROT3 | A cytochrome P-450 gene that is involved in brassinosteroid biosynthesis, most likely in the conversion step of typhasterol (TY) to castasterone (CS) |
|  | x | AT4G36690 | U2AF65A | U2 snRNP auxiliary factor large subunit |
| x |  | AT4G37110 | - | Zinc-finger domain of monoamine-oxidase A repressor R1 |
|  | x | AT4G37670 | NAGS2 | N-acetyl-l-glutamate synthase 2 |
| x | x | AT4G38200 | BIG1 | Brefeldin A-inhibited guanine nucleotide-exchange protein 1 |
|  | x | AT4G38380 | DTX45 | Protein DETOXIFICATION 45, chloroplastic |
| x | x | AT4G38960 | - | B-box type zinc finger family protein |
| x |  | AT4G39140 | - | RING finger family protein |
| x | x | AT4G39270 | - | Leucine-rich repeat protein kinase family protein |
| x | x | AT4G40030 | - | Histone superfamily protein |
| x |  | AT5G01175 | - | other RNA |
|  | x | AT5G01730 | SCAR4 | Protein SCAR4 |
| x |  | AT5G01770 | RAPTOR2 | Regulatory-associated protein of TOR 2 |
| x | x | AT5G01780 | - | 2-oxoglutarate-dependent dioxygenase family protein |
| x | x | AT5G02080 | - | DNA / pantothenate metabolism flavoprotein |
|  | x | AT5G02860 | - | Pentatricopeptide repeat-containing protein At5g02860 |
|  | x | AT5G03180 | - | RING/U-box superfamily protein |
| x |  | AT5G03440 | - | zinc finger protein |
| x | x | AT5G03500 | - | Mediator complex, subunit Med7 |
| x |  | AT5G03830 | - | CDK inhibitor P21 binding protein |
| x |  | AT5G03960 | IQD12 | IQ-domain 12 |
|  | x | AT5G04220 | SYT3 | Synaptotagmin-3 |
|  | x | AT5G04280 | RZ1C | Glycine-rich RNA-binding protein RZ1C |
|  | x | AT5G05310 | - | TLC ATP/ADP transporter |
| x |  | AT5G05435 | - | other RNA |
| x |  | AT5G05550 | - | Sequence-specific DNA binding transcription factor |
| x |  | AT5G05670 | - | signal recognition particle binding protein |
|  | x | AT5G05800 | LIMYB | L10-interacting MYB domain-containing protein |
| x |  | AT5G06100 | MYB33 | Transcription factor MYB33 |
|  | x | AT5G06120 | - | ARM repeat superfamily protein |
|  | x | AT5G06265 | - | hyaluronan mediated motility receptor-related |
| x |  | AT5G06440 | - | Polyketide cyclase/dehydrase and lipid transport superfamily protein |
| x |  | AT5G06690 | WCRKC1 | WCRKC thioredoxin 1 |
|  | x | AT5G06800 | - | myb-like HTH transcriptional regulator family protein |
|  | x | AT5G06830 | - | CDK5RAP3-like protein |
|  | x | AT5G06960 | TGA5 | OCS-element binding factor 5 interacts with NPR1 to promote expression of salicylic acid induced genes |
| x |  | AT5G07440 | GDH2 | Glutamate dehydrogenase |
| x |  | AT5G07890 | - | myosin heavy chain-like protein |
| x | x | AT5G08130 | BIM1 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
|  | x | AT5G08190 | NF-YB12 | nuclear factor Y, subunit B12 |
|  | x | AT5G08440 | - | Transmembrane protein |
|  | x | AT5G08450 | - | Zinc finger CCCH domain protein |
| x |  | AT5G08560 | WDR26 | WD repeat-containing protein 26 homolog |
|  | x | AT5G08750 | - | RING/FYVE/PHD zinc finger superfamily protein |
| x |  | AT5G09230 | SRT2 | sirtuin 2 |
| x |  | AT5G09410 | EICBP.B | Ethylene induced calmodulin binding protein |
| x | x | AT5G09690 | ATMGT7 | magnesium transporter 7 |
|  | x | AT5G10350 | PABN3 | Polyadenylate-binding protein 3 |
|  | x | AT5G10450 | GRF6 | G-box regulating factor 6 |
|  | x | AT5G10470 | KCA1 | kinesin like protein for actin based chloroplast movement 1 |
| x |  | AT5G10710 | - | Centromere protein O |
| x |  | AT5G10800 | - | Protein RRC1-like |
|  | x | AT5G11010 | - | Polynucleotide 5'-hydroxyl-kinase NOL9 |
| x | x | AT5G11380 | DXPS3 | 1-D-deoxyxylulose 5-phosphate synthase-like protein |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x | | AT5G12440 | - | CCCH-type zinc fingerfamily protein with RNA-binding domain |
| x | x | AT5G13220 | JAZ10 | jasmonate-zim-domain protein 10 |
| | x | AT5G13590 | - | hypothetical protein |
| x | | AT5G13950 | NFRKB2 | nuclear factor kappa-B-binding protein |
| x | | AT5G14210 | - | Leucine-rich repeat protein kinase family protein |
| x | | AT5G14260 | - | Rubisco methyltransferase family protein |
| | x | AT5G14610 | - | DEAD box RNA helicase family protein |
| x | x | AT5G14740 | BCA2 | Beta carbonic anhydrase 2, chloroplastic |
| | x | AT5G15460 | MUB2 | Membrane-anchored ubiquitin-fold protein 2 |
| | x | AT5G16520 | - | transmembrane protein |
| | x | AT5G16880 | TOL1 | TOM1-like protein 1 |
| x | | AT5G16980 | - | Zinc-binding dehydrogenase family protein |
| x | x | AT5G18245 | - | other RNA |
| x | | AT5G18420 | - | CCR4-NOT transcription complex subunit |
| x | | AT5G18440 | - | Encodes NUFIP that directs assembly of C/D snoRNP (small nucleolar ribonucleoprotein) |
| | x | AT5G18630 | - | Alpha/beta-Hydrolases superfamily protein |
| x | | AT5G18980 | - | ARM repeat superfamily protein |
| x | | AT5G19030 | - | RNA-binding (RRM/RBD/RNP motifs) family protein |
| | x | AT5G19210 | RH58 | DEAD-box ATP-dependent RNA helicase 58, chloroplastic |
| x | x | AT5G19480 | MED19B | Probable mediator of RNA polymerase II transcription subunit 19b |
| | x | AT5G20220 | - | zinc knuckle (CCHC-type) family protein |
| | x | AT5G20360 | Phox3 | Protein PHOX3 |
| | x | AT5G20450 | - | Myosin family protein with Dil domain |
| | x | AT5G22300 | NIT4 | Bifunctional nitrilase/nitrile hydratase NIT4 |
| | x | AT5G22700 | - | F-box/RNI-like/FBD-like domains-containing protein |
| x | | AT5G22760 | - | PHD finger family protein |
| x | | AT5G22820 | - | ARM repeat superfamily protein |
| x | x | AT5G22860 | - | Prolylcarboxypeptidase-like protein |
| | x | AT5G23480 | - | SWIB/MDM2 domain;Plus-3;GYF |
| | x | AT5G23720 | PHS1 | Dual specificity protein phosphatase PHS1 |
| | x | AT5G23870 | PAE9 | Pectin acetylesterase 9 |
| x | | AT5G24310 | ABIL3 | Protein ABIL3 |
| x | | AT5G24318 | - | O-Glycosyl hydrolases family 17 protein |
| | x | AT5G24350 | MIP1 | Member of MAG2 complex on the ER that is responsible for efficient transport of seed storage proteins |
| x | x | AT5G25520 | - | SPOC domain / Transcription elongation factor S-II protein |
| | x | AT5G25540 | CID6 | Polyadenylate-binding protein-interacting protein 6 |
| x | | AT5G26180 | - | S-adenosyl-L-methionine-dependent methyltransferases superfamily protein |
| x | x | AT5G26680 | - | Flap endonuclease 1 |
| | x | AT5G26740 | - | organic solute transporter ostalpha protein (DUF300) |
| x | x | AT5G26770 | NEAP2 | coiled-coil domain protein |
| x | | AT5G27950 | KIN14U | Kinesin-like protein KIN-14U |
| x | | AT5G28770 | BZIP63 | Basic leucine zipper 63 |
| x | x | AT5G35995 | - | F-box/LRR-repeat protein |
| | x | AT5G36180 | SCPL1 | Serine carboxypeptidase-like 1 |
| x | x | AT5G37370 | SRL1 | Pre-mRNA splicing factor SR-like 1 |
| x | x | AT5G37380 | - | Chaperone DnaJ-domain superfamily protein |
| x | | AT5G37710 | - | alpha/beta-Hydrolases superfamily protein |
| x | | AT5G37850 | SOS4 | pfkB-like carbohydrate kinase family protein |
| | x | AT5G40820 | ATRAD3 | Ataxia telangiectasia-mutated and RAD3-related |
| | x | AT5G41690 | - | RNA-binding (RRM/RBD/RNP motifs) family protein |
| | x | AT5G41910 | MED10A | Mediator of RNA polymerase II transcription subunit 10a |
| | x | AT5G42770 | - | Maf-like protein |
| x | | AT5G42900 | COR27 | Cold regulated protein 27 |
| x | | AT5G43500 | ARP9 | Actin-related protein 9 |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x | x | AT5G43630 | TZP | Zinc knuckle (CCHC-type) family protein |
|  | x | AT5G43910 | - | PfkB-like carbohydrate kinase family protein |
| x | x | AT5G43930 | - | Transducin family protein / WD-40 repeat family protein |
| x | x | AT5G43990 | SUVR2 | SET-domain containing protein lysine methyltransferase family protein |
|  | x | AT5G44010 | - | fanconi anemia group F protein (FANCF) |
|  | x | AT5G44290 | - | Protein kinase superfamily protein |
|  | x | AT5G44750 | REV1 | DNA repair protein REV1 |
|  | x | AT5G45115 | - | Transcription termination factor family protein |
| x |  | AT5G45410 | - | zinc finger B-box protein, chloroplast |
| x | x | AT5G45472 | - | Potential natural antisense gene, locus overlaps with AT5G45470 |
|  | x | AT5G45500 | - | RNI-like superfamily protein |
|  | x | AT5G45560 | EDR2L | Protein ENHANCED DISEASE RESISTANCE 2-like |
| x | x | AT5G45920 | - | GDSL esterase/lipase At5g45920 |
| x |  | AT5G45940 | NUDT11 | Nudix hydrolase 11 |
| x |  | AT5G47380 | - | electron transporter, putative (DUF547) |
|  | x | AT5G47455 | - | hypothetical protein |
| x |  | AT5G47650 | ATNUDT2 | nudix hydrolase homolog 2 |
| x |  | AT5G47690 | PDS5A | Binding protein |
|  | x | AT5G47900 | - | heparan-alpha-glucosaminide N-acetyltransferase-like protein (DUF1624) |
| x |  | AT5G48150 | PAT1 | Scarecrow-like transcription factor PAT1 |
| x | x | AT5G48560 | BHLH78 | Transcription factor bHLH78 |
| x | x | AT5G48610 | - | myb-like protein X |
|  | x | AT5G48620 | RPP8L4 | Probable disease resistance RPP8-like protein 4 |
|  | x | AT5G48640 | CYCC1-1 | Cyclin-C1-1 |
| x |  | AT5G48775 | - | other RNA |
|  | x | AT5G48880 | PKT2 | peroxisomal 3-keto-acyl-CoA thiolase 2 |
| x |  | AT5G49960 | - | ion channel POLLUX family member |
| x | x | AT5G50130 | - | NAD(P)-binding Rossmann-fold superfamily protein |
| x |  | AT5G50160 | FRO8 | Ferric reduction oxidase 8, mitochondrial |
|  | x | AT5G50565 | - | hypothetical protein |
| x | x | AT5G51180 | - | alpha/beta-Hydrolases superfamily protein |
|  | x | AT5G51620 | - | Uncharacterised protein family (UPF0172) |
| x |  | AT5G52040 | ATRSP41 | RNA-binding (RRM/RBD/RNP motifs) family protein |
| x | x | AT5G52070 | - | Agenet domain-containing protein |
| x |  | AT5G52230 | MBD13 | Methyl-CpG-binding domain-containing protein 13 |
| x |  | AT5G52790 | CBSDUF5 | CBS domain protein with a domain protein (DUF21) |
| x | x | AT5G53050 | - | Alpha/beta-Hydrolases superfamily protein |
|  | x | AT5G53180 | ATPTB2 | Polypyrimidine tract-binding protein homolog 2 |
|  | x | AT5G53420 | - | CCT motif family protein |
| x | x | AT5G53486 | - | transmembrane protein |
| x | x | AT5G53550 | YSL3 | YELLOW STRIPE like 3 |
| x |  | AT5G53760 | MLO11 | MILDEW RESISTANCE LOCUS O 11 |
| x | x | AT5G53850 | - | haloacid dehalogenase-like hydrolase family protein |
|  | x | AT5G54569 | - | other RNA |
| x |  | AT5G55040 | - | DNA-binding bromodomain-containing protein |
| x | x | AT5G55100 | - | SWAP (Suppressor-of-White-APricot)/surp domain-containing protein |
| x | x | AT5G55550 | - | RNA-binding (RRM/RBD/RNP motifs) family protein |
| x |  | AT5G56140 | - | RNA-binding KH domain-containing protein |
|  | x | AT5G56190 | - | Transducin/WD40 repeat-like superfamily protein |
|  | x | AT5G56320 | ATEXPA14 | expansin A14 |
|  | x | AT5G56850 | - | hypothetical protein |
| x | x | AT5G57060 | - | 60S ribosomal L18a-like protein |
| x |  | AT5G57450 | XRCC3 | DNA repair protein XRCC3 homolog |
| x | x | AT5G57565 | - | Protein kinase superfamily protein |

| 1h | 5h | gene_id | gene_name | gene_description |
|---|---|---|---|---|
| x |  | AT5G57790 | - | Encodes a nuclear localized protein of unknown function that is involved in pollen and embryo sac development |
| x |  | AT5G58370 | EngB-3 | P-loop containing nucleoside triphosphate hydrolases superfamily protein |
| x |  | AT5G58790 | - | hypothetical protein |
| x |  | AT5G58940 | CRCK1 | Calmodulin-binding receptor-like cytoplasmic kinase 1 |
| x |  | AT5G59780 | MYB59 | Transcription factor MYB59 |
| x | x | AT5G59830 | - | unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G13660.2); Ha. |
|  | x | AT5G60210 | RIP5 | ROP interactive partner 5 |
|  | x | AT5G60430 | - | Antiporter/ drug transporter |
|  | x | AT5G60870 | - | Regulator of chromosome condensation (RCC1) family protein |
|  | x | AT5G60930 | - | P-loop containing nucleoside triphosphate hydrolases superfamily protein |
|  | x | AT5G61190 | - | putative endonuclease or glycosyl hydrolase with C2H2-type zinc finger domain |
| x |  | AT5G61270 | BHLH72 | Transcription factor PIF7 |
| x | x | AT5G61310 | - | Cytochrome c oxidase subunit 5C |
| x |  | AT5G61530 | - | Uncharacterized Rho GTPase-activating protein At5g61530 |
|  | x | AT5G61960 | ML1 | Protein MEI2-like 1 |
| x |  | AT5G62760 | - | P-loop containing nucleoside triphosphate hydrolases superfamily protein |
|  | x | AT5G62950 | - | RNA polymerase II, Rpb4, core protein |
| x | x | AT5G63120 | RH30 | DEAD-box ATP-dependent RNA helicase 30 |
|  | x | AT5G63220 | GET4 | golgi-to-ER traffic-like protein |
| x |  | AT5G63370 | CDKG1 | Cyclin-dependent kinase G1 |
|  | x | AT5G63440 | - | hypothetical protein (DUF167) |
| x |  | AT5G63700 | - | Zinc ion binding / DNA binding protein |
|  | x | AT5G64170 | LNK1 | Protein LNK1 |
| x |  | AT5G64460 | - | Phosphoglycerate mutase-like protein 1 |
|  | x | AT5G64470 | TBL12 | Protein trichome birefringence-like 12 |
| x | x | AT5G64630 | FAS2 | Chromatin assembly factor 1 subunit FAS2 |
| x | x | AT5G65070 | MAF4 | K-box region and MADS-box transcription factor family protein |
|  | x | AT5G65440 | - | Transmembrane protein |
| x | x | AT5G65685 | - | UDP-Glycosyltransferase superfamily protein |
| x | x | AT5G65740 | - | zinc ion binding protein |
|  | x | AT5G66005 | - | Expressed protein |
|  | x | AT5G66600 | - | Putative uncharacterized protein |
| x |  | AT5G66675 | - | transmembrane protein, putative (DUF677) |
|  | x | AT5G67110 | ALC | Transcription factor ALC |
|  | x | AT5G67540 | - | Arabinanase/levansucrase/invertase |

**Appendix Table S7. Alternative splicing in response to various abiotic stresses.**

| Stress | Conditions | No. DAS events | No. DASGs | No. (%) DEGs in DASGs | Reference |
|---|---|---|---|---|---|
| Oxidative stress | Seedlings_5d_1h_WT vs Seedlings_5d_0h_WT | 1514 | 1106 | 49 (4.43%) | **This work** |
| Oxidative stress | Seedlings_5d_5h_WT vs Seedlings_5d_0h_WT | 1472 | 1093 | 129 (11.80%) | |
| Oxidative stress | Seedlings_5d_1h_why1 vs Seedlings_5d_0h_why1 | 1339 | 1010 | 67 (6.63%) | |
| Oxidative stress | Seedlings_5d_5h_why1 vs Seedlings_5d_0h_why1 | 1379 | 1093 | 132 (12.08%) | |
| ABA | Seedlings_2wLD_6h_mock_WT vs Seedlings_2wLD_6h_ABA_WT | 498 | 394 | 46 (11.68%) | Martin et al., 2021 and references within |
| Drought | Seedlings_9d_24h_control_WT vs Seedlings_9d_24h_mannitol_WT | 1184 | 898 | 111 (12.36%) | |
| Salt | Seedlings_11d_4h_control_WT vs Seedlings_11d_4h_NaCl_WT | 1342 | 1007 | 113 (11.22%) | |
| Heat | Leaves_5wSD_30min_23C_WT vs Leaves_5wSD_30min_37C_WT | 1684 | 1254 | 141 (11.24%) | |
| Cold | Seedlings_9dWL_0h_10C_WT vs Seedlings_9dWL_24h_10C_WT | 1262 | 947 | 130 (13.72%) | |
| Cold | Leaves_5wSD_4d_4C_WT vs Leaves_5wSD_0d_20C_WT | n.a. | 2442 | 795 (32.56%) | Calixto et al., 2018 |
| $CO_2$ | Seedling_5dLD_150ppm_CO2_WT vs Seedling_5dLD_500ppm_CO2_WT | n.a. | 251 | 5 (2%) | Huang et al., 2019b |

**Appendix Table S8. Primers used in the present work.**

| Name | Sequence (5' to 3') |
| --- | --- |
| *For screening WHIRLY1 knock-out mutants* | |
| Cas9-F2 | CAGCTCGTGCAGACCTACAAC |
| Cas9-R2 | TGCCTTCTAAGGATAGCGTG |
| CRISPR-WHY1-for | TTGTCTCGTCATCTCTCTCG |
| CRISPR-WHY1-rev | CCTTTTGTGTTTTTAAGCTTCC |
| *For genotyping WHIRLY1 mutants* | |
| why1-cc-for | AGTCTCTGAGTGTGTAAAGGAAGC |
| why1-cc-rev | TCTTCGTCGTCGGTTTCAAGG |
| LBa1 | TGGTTCACGTAGTGGGCCATCG |
| why1-t-for | AGTCTCTGAGTGTGTAAAGGAAGC |
| why1-t-rev | ATGACCCACGTAAAATCTAGCAG |
| *For quantitative real-time PCR* | |
| AtACTIN2_qF1 | AACTCTCCCGCTATGTATGTCG |
| AtACTIN2_qR1 | CCTCGTAGATTGGCACAGTG |
| AT5G46630_qF1 | TTCCAAGTGACAACTGGTCG |
| AT5G46630_qR1 | CAATTTCTGCACTTAGCGTGG |
| AtPP2AA2_qF2 | CCATTAGATCTTGTCTCTCTGCT |
| AtPP2AA2_qR2 | GACAAAACCCGTACCGAG |
| AtWhy1_qF2 | GAGTTTGCTGTTCTTATCTCTGC |
| AtWhy1_qR2 | TTGTTCACACGGCTTGTCTC |
| AtWhy2_qF1 | TTTGCAGTGATGAAGACAGC |
| AtWhy2_qR1 | AGATGAGAAACATTCCTCGACG |
| AtWhy3_qF1 | TGTGGCTTTGGAATCTGGTGC |
| AtWhy3_qR1 | AGAAGACCTGCTTCCTACTCC |
| AtRAV1_qF1 | AGAGCAGAGTAAACGGCGTC |
| AtRAV1_qR1 | AACCCCGTCGTAGAAACACC |
| AtFSD2_qF1 | AGGAAAGCCAACTGGAGAGC |
| AtFSD2_qR1 | GCAGCCGACTTGAACCTTTC |
| AtMYB28_qF1 | AAAAAGCTGGGTTGAAACGGTG |
| AtMYB28_qR1 | CTCGCCTCTTTTGATCTCAGGT |
| AtMYB29_qF1 | GAGCAATCGCAATCGGGTTC |
| AtMYB29_qR1 | GTGTCTCGTCACTGCTGGTT |
| AtBCAT4_qF1 | TCCTGAGACCACCTTCACAGC |
| AtBCAT4_qR1 | CACACCACCAGTTCCACTAGG |
| AtMAM1_qF1 | TGATTGAGATGGCCGTGAGT |
| AtMAM1_qR1 | GAAATCCTTGTCCGACCTGC |
| AtIPMI1_qF1 | ACACGCTCCAGTTTGTCTCG |
| AtIPMI1_qR1 | CGGGAAAATCTCACCTGTAGC |
| AtIPMI2_qF1 | GACTCTAACGAAGCCCTAGCC |
| AtIPMI2_qR1 | GGATGATCTGGTCGGTGTCTA |
| AtBASS5_qF1 | TCTGACTCAAACGAGCTGTATC |

| Name | Sequence (5' to 3') |
|---|---|
| AtBASS5_qR1 | ACAAAGTACCTTGGCTTGAACC |
| AtCYP79F1_qF1 | TTGCGCGTCAAGATACCACC |
| AtCYP79F1_qR1 | GGTTACGACCTAGTCCAGGG |
| AtCYP79F2_qF1 | CTTTTGTCTGTTGAGCCACGC |
| AtCYP79F2_qR1 | AACGACAAAGCGTCGAAACAC |
| AtMYB34_qF1 | CTGACCGGGATCATCAGTGG |
| AtMYB34_qR1 | TTGAGCAATGTGGAGGTCGG |
| AtMYB51_qF1 | CCCTTCACGGCAACAAATGG |
| AtMYB51_qR1 | CCGGAGGTTATGCCCTTGTG |
| AtCYP83A1_qF1 | TACCAAGATCGCCGGTTACG |
| AtCYP83A1_qR1 | CGGGCCTAAACTCATCAGGG |
| AtSUR1_qF1 | TGCTGCTTACAGTGGTCTCG |
| AtSUR1_qR1 | AGCCACAGTGTTCTCGTCTG |
| *For reverse-transcriptase PCR* | |
| AtBPL2_asf1 | ACACAAGAGAGTCAGGTCGC |
| AtBPL2_asr1 | CCATTAAAACGAATGTTCCTGC |
| SR34b_asf3 | TCTCCTCGCTCTCGCTCCC |
| SR34b_asr3 | TGTCCTAGCTCGATGGGTCC |
| FSD2_asf1 | GGCGAATAGACTTGACGTTGC |
| FSD2_asr1 | ACCTTGTGCTTACAGTTTCCC |

**Appendix Table S9. Summary of RNA quality, library quality, and mapping results**. RIN value: RNA Integrity Number calculated from Agilent 2100 Bioanalyzer system. Raw reads: Raw reads: reads count from the raw data; Clean reads: reads count after filtering reads containing adaptors or N > 10% or Qscore of over 50% bases below 5. Raw bases: number of bases in the raw data, (number of raw read) * (sequence length), converting unit to G; Clean bases: number of bases in clean data, (number of clean reads) * (sequence length), converting unit to G; Error rate (%): base error rate of whole sequencing; Q30(%): The percentage of the bases whose Q Phred values is greater than 30. (Number of bases with Q Phred value > 30) / (Number of total bases) *100; GC content (%): The percentage of G&C base numbers of total bases. (G&C base number) / (Total base number) *100. Total mapped reads: numbers of clean reads being mapped on the genome; Uniquely mapped reads: numbers of reads being mapped on single position of the genome; Total mapping rate: (mapped reads)/(total reads)*100; Uniquely mapping rate: (uniquely mapped reads)/(total reads)*100; Exonic, intronic, intergenic regions: percentage of clean reads mapped in exonic, intronic or intergenic regions of reference genome.

| Sample | WT_ct_1 | WT_ct_2 | WT_ct_3 | WT_1h_1 | WT_1h_2 | WT_1h_3 | WT_5h_1 | WT_5h_2 | WT_5h_3 |
|---|---|---|---|---|---|---|---|---|---|
| RIN value | 8.1 | 8.2 | 8.4 | 8.4 | 8.4 | 8.4 | 8.1 | 8.5 | 8.5 |
| Raw reads | 24258106 | 19246368 | 21534927 | 21532838 | 25056003 | 23736720 | 20321164 | 23378697 | 23772515 |
| Clean reads | 23610330 | 18821582 | 21059695 | 21173840 | 24589960 | 23268720 | 20052211 | 22962917 | 23136736 |
| Raw bases | 7.3G | 5.8G | 6.5G | 6.5G | 7.5G | 7.1G | 6.1G | 7.0G | 7.1G |
| Clean bases | 7.1G | 5.6G | 6.3G | 6.4G | 7.4G | 7.0G | 6.0G | 6.9G | 6.9G |
| Error rate | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 | 0.03 | 0.03 |
| Q30 | 91.61 | 91.68 | 91.76 | 91.75 | 92.36 | 92.24 | 94.81 | 91.96 | 92.02 |
| GC content | 47.45 | 47.53 | 47.3 | 47.36 | 47.52 | 47.14 | 47.31 | 47.02 | 46.96 |
| Total mapping rate | 96.88% | 96.77% | 96.74% | 96.65% | 97.17% | 97.13% | 98.19% | 96.92% | 96.93% |
| Uniquely mapping rate | 93.16% | 93.09% | 92.25% | 92.65% | 92.89% | 93.61% | 93.84% | 93.41% | 93.50% |
| Multiple mapping rate | 3.71% | 3.68% | 4.49% | 4.00% | 4.28% | 3.52% | 4.35% | 3.51% | 3.43% |
| Percentage of reads mapped to exon | 97.90% | 97.90% | 96.99% | 97.53% | 97.32% | 97.80% | 97.79% | 98.03% | 98.24% |
| Percentage of reads mapped to intron | 0.40% | 0.33% | 0.38% | 0.43% | 0.40% | 0.42% | 0.24% | 0.31% | 0.31% |
| Percentage of reads mapped to intergenic region | 1.70% | 1.77% | 2.64% | 2.05% | 2.27% | 1.78% | 1.96% | 1.66% | 1.45% |

| Sample | w1_ct_1 | w1_ct_2 | w1_ct_3 | w1_1h_1 | w1_1h_2 | w1_1h_3 | w1_5h_1 | w1_5h_2 | w1_5h_3 |
|---|---|---|---|---|---|---|---|---|---|
| **RIN value** | 8.3 | 8.4 | 8.4 | 7.8 | 8.3 | 8.2 | 8.3 | 8.2 | 8.1 |
| **Raw reads** | 24913730 | 21706075 | 20030280 | 21734810 | 23076325 | 24075907 | 25671411 | 21812188 | 24467891 |
| **Clean reads** | 24365409 | 21407189 | 19596839 | 21386854 | 22627512 | 23593970 | 25079220 | 21421753 | 24083394 |
| **Raw bases** | 7.5G | 6.5G | 6.0G | 6.5G | 6.9G | 7.2G | 7.7G | 6.5G | 7.3G |
| **Clean bases** | 7.3G | 6.4G | 5.9G | 6.4G | 6.8G | 7.1G | 7.5G | 6.4G | 7.2G |
| **Error rate** | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| **Q30** | 91.82 | 94.09 | 91.48 | 92.85 | 92.13 | 93.45 | 92.95 | 91.99 | 92.29 |
| **GC content** | 45.71 | 47.01 | 46.45 | 46.58 | 46.35 | 46.68 | 45.62 | 46.41 | 46.34 |
| **Total mapping rate** | 62.26% | 85.71% | 80.92% | 84.56% | 78.90% | 84.45% | 71.27% | 85.76% | 86.08% |
| **Uniquely mapping rate** | 59.80% | 82.23% | 78.04% | 81.59% | 75.76% | 81.58% | 69.05% | 83.04% | 82.98% |
| **Multiple mapping rate** | 2.45% | 3.49% | 2.87% | 2.97% | 3.14% | 2.88% | 2.22% | 2.72% | 3.10% |
| **Percentage of reads mapped to exon** | 97.62% | 97.65% | 97.86% | 97.88% | 97.53% | 97.96% | 98.60% | 98.29% | 98.08% |
| **Percentage of reads mapped to intron** | 0.33% | 0.32% | 0.40% | 0.36% | 0.40% | 0.38% | 0.25% | 0.28% | 0.34% |
| **Percentage of reads mapped to intergenic region** | 2.05% | 2.03% | 1.74% | 1.76% | 2.06% | 1.66% | 1.15% | 1.43% | 1.57% |

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGEMENT

I feel so lucky and grateful to have many people who have supported me throughout my PhD. So I would like to begin by saying thank you to everyone who has helped me along the way.

Thank you to my supervisor, Klaus Humbeck. I am grateful that you gave me the opportunity to come here, to join our working group AG Humbeck, and to be part of RTG2498. Thank you for patiently guiding me at the very beginning of my PhD. I appreciate our many engaging discussions about the project and also many 'therapy sessions", supporting me in both research and everyday life. I am glad that you always listen to all of my craziest ideas and believe in me. I have learned so much from you, not only in doing research and writing but also in supervising and supporting a fellow researcher.

I would also like to thank Prof. Karin Krupinska, who, from the beginning until today, has inspired me about WHIRLIES. And my gratitude also goes to Prof. Katharina Markmann, who encourages me to be more curious in research. You both have given me valuable input in every thesis committee meeting and also in many other spontaneous discussions.

I want to further express my appreciation to reviewers for their time and constructive feedbacks on the thesis.

Furthermore, I am truly thankful to all current and former members of AG Humbeck for their support. Olaf, I am grateful for your guidance when I started working with *Arabidopsis*, for our many discussions which gave new ideas for my work, and for helping troubleshoot many issues in my experiments. Athina and Charlotte, I am glad that you welcomed me into the office, and I always cherish all our work and non-work time together. Your input, especially in my writing, was very helpful. Siska and Nancy, your organization in the lab made my work much easier. I thank you both also for helping me with genotyping and *N. benthamina* work. Michael, I very much appreciate your help in the greenhouse and I always remember our fun little experiments. Minh, thank you for sharing the bench with me and also for many input on the project. Ivana, I am very glad that you continue the work on WHIRLIES and I thank you for many morning walks together. Wiebke and Sylvie, although we didn't see each other often, I enjoyed every talk we had. Many of you have also helped me with booking appointments and translating documents, and I am truly

grateful for all the assistance I received. Last but not least, I would like to special thank Elena, your cheerful assistance in the project has been very helpful and enjoyable.

I also would like to express my gratitude to all members of RTG2498, I am lucky to be a small part of the RTG. I am thankful for all discussions during regular seminars and the supportive and collaborative working atmosphere, especially to Pinelopi and AG Abel for helping me in glucosinolate measurement. And I am very grateful to RTG for granting me the start-up funding, as it encourages and supports me to stay in academia. To all PhD students, I remember many social events we had over the last 4 years, and many of you have become my friends. I wish you all the best in the future.

Of course, I would like to say thank you to all my friends who have completed my PhD experience. Yunjing, you sent me a book called 'Surviving your PhD' at the beginning and I am glad that we have survived it together with a lot of laughter and food. You encouraged me to try new things and inspired me with your dedication and passion. I believe you will be very successful in the future. Athina, I always have a good time with you and I wish that we could have hung out more. But I am very happy for your next steps and I sincerely wish you all the best. Huyền and Cường, you have supported me throughout my PhD as well as my normal life with hours of talking. I truly appreciate our friendship and I know we will stick together for a while. Nikola and Floki, you have endured my complaining about my writing and made it much easier, meow. And to many more with whom I shared experiences and good times during my PhD journey, Bèo Nước, Gremlins, Grandma...

Con cảm ơn bố mẹ và em trai đã luôn thông cảm và luôn khích lệ con từ xa.

Once again, I sincerely thank you all very much. Your guidance, assistance, collaboration, and friendship have profoundly influenced both my academic journey and my personal growth.

# CURRICULUM VITAE

Name: Nguyen, Thuy Linh

Nationality: Vietnamese

Date of birth: -

Place of birth: Vietnam

*Education and research experience*

| | |
|---|---|
| 10/2019 – present | **PhD student** |
| | AG Humbeck, Plant Physiology Department, Faculty of Natural Science I, Martin-Luther University of Halle-Wittenberg, Germany |
| | Thesis title: Investigations of *Arabidopsis thaliana* WHIRLY1 function in early seedling development |
| | Supervisor: Prof. Dr. Klaus Humbeck |
| | |
| 10/2015 – 08/2019 | **Researcher** |
| | Genome Biodiversity Laboratory |
| | Institute of Genome Research, Vietnam Academy of Science and Technology, Vietnam |
| | |
| 11/2016 - 04/2019 | **Master student** |
| | Faculty of Biology, Vietnam National University of Science, Vietnam |
| | Thesis title: Research on the transgenic maize harboring *DREB2.7* gene related to drought tolerance |
| | Supervisors: Dr. Huynh Thi Thu Hue; Dr. Le Hong Diep |
| | |
| 09/2011 – 06/2015 | **Bachelor student** |
| | Faculty of Biology, Vietnam National University of Science, Vietnam |
| | Thesis title: "Construction of yeast *Pichia pastoris* expresssing the recombinant HIV-1 protease" |
| | Supervisor: Dr. Dinh Nho Thai |

*Conference attendances*

| | |
|---|---|
| 02/2023 | poster presentation at Molecular Biology of Plants MBP2023 |
| 11/2022 | oral presentation at WORKSHOP – AMU/CAU: ''Strategies aiming to understand the role of WHIRLIES in plant development and stress resistance'' |
| 03/2022 | online symposium organization in RTG (2498) plant cell compartments: stay connected |
| 09/2021 | oral presentation at Mitteldeutsche Pflanzenphysiologie Tagung 2021 |

*List of publications*

1. Ost, C., Cao, X. H., **Nguyen, T. L.**, Himmelbach, A., Mascher, M., Stein, N., Humbeck, K. (2023) Drought stress-related reprogramming of gene expression in barley leaves involves loading of ABA-related genes with euchromatic marks H3K4me3 and H3K9ac. *Int J Mol Sci.* 24(15):12065. doi: 10.3390/ijms241512065

2. Hoang, T. T. Y., Luu, H. L., **Nguyen, T. L.**, Duong, T. D., Nguyen, H. D., Huynh, T. T. H. (2019) A comparison of the production of polyphenol contents and the expression of genes involved in Vietnamese tea cultivars. *Int Food Res J* 26(6):1781-1788.

3. Huynh, T. T. H., **Nguyen, T. L.**, Luu, H. L., Nguyen, H. H., Le H. D., Bui, M. M., Pham, T. H., Doan, T. B. T., Le, T. T. H., Ha, H. H., Nong, V. H. (2019) Isolation and characterization of a *DREB* homolog gene from a local drought-tolerant maize cultivar. *ACTA BIOLOGICA CRACOVIENSIA Series Botanica* 61(2). doi:10.24425/abcsb.2019.127746.

4. Bui, M. M., **Nguyen, T. L.**, Ha, H. H., Le, T. T. H., Nguyen, X. T., Nong, V. H., Huynh, T. T. H. (2019) A *LEA* gene from a Vietnamese maize landrace can enhance the drought tolerance of transgenic maize and tobacco. *Agronomy* 9(2):62. doi: 10.3390/agronomy9020062.

5. **Nguyen, T. L.**, Luu, H. L., Nguyen, T. D., Huynh, T. T. H. (2019) Isolation and characterization of a C-repeat binding factor gene from Tevang-1 maize cultivar. *Academia Journal of Biology* 41(3): 85–100. doi: 10.15625/2615-0923/v41n3.13872

6. **Nguyen, T. L.**, Pham, T. H., Do, V. T., Huynh, T. T. H. (2017) Evaluating the systematic position of *Ehretia asperula* Zoll. & Moritzi based on ITS1, matK, and trnL-trnF DNA sequences. *Vietnam Journal of Science, Technology and Engineering* 59(4): 61-65. doi: 10.31276/VJSTE.59(4).61

# EIDESSTATTLICHE ERKLÄRUNG (STATUTORY DECLARATION)

Hiermit erkläre ich an Eides statt, dass ich mich mit der vorliegenden wissenschaftlichen Arbeit erstmals um die Erlangung des Doktorgrades bewerbe, die Arbeit selbstständig und ohne fremde Hilfe verfasst, nur die angegebenen Quellen und Hilfsmittel genutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Halle (Saale),den _____         _____

                                        Nguyen, Thuy Linh