Comprehensive transcriptome analysis of arabidopsis thaliana DNA polymerase epsilon catalytic subunit A and B mutants

Article

Accepted Version

Comprehensive transcriptome analysis of *Arabidopsis thaliana DNA polymerase epsilon catalytic subunit A* and *B* mutants

Anushka M. Wickramasuriya[1]*, Thulani M. Hewavithana[1], Kithmee K. de Silva[1], Ihsan Ullah[2], Jim M. Dunwell[2]

[1] Department of Plant Sciences, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka

[2] School of Agriculture, Policy and Development, University of Reading, UK

*Corresponding author:

Department of Plant Sciences, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka

Email: anushka@pts.cmb.ac.lk

Abstract

In *Arabidopsis*, the catalytic subunit of DNA polymerase ε (POLE) is encoded by two genes: *DNA polymerase epsilon catalytic subunit A* (*AtPOL2A*) and *B* (*AtPOL2B*). Although studies have shown *AtPOL2A* to be involved in various biological processes, the role of *AtPOL2B* remains to be determined. In the present study, leaf cDNA libraries of one *AtPOL2A* mutant (*atpol2a-1*) and three *AtPOL2B* mutants (*atpol2b-1*, *-2* and *-3*) were sequenced using the Illumina platform. Analysis of gene expression profiles identified a total of 198, 76, 141 and 67 differentially expressed genes (DEGs) in *atpol2a-1*, *atpol2b-1*, *atpol2b-2* and *atpol2b-3*, respectively. It was noted that the majority of pericentromeric transposable elements were transcriptionally active in *atpol2a-1* as compared to *atpol2b* mutants and wild-type plants. Computational analysis of the proteins encoded by the DEGs identified CER1, Replication Protein A 1E (RPA1E) and AT5G60250 as potential interactors of AtPOL2A, and Pathogenesis-related gene 1 (PR1) and AT5G48490 as potential interactors of AtPOL2B. Interestingly, all these proteins showed a significant interaction with the POLE catalytic subunit of *Saccharomyces cerevisiae*. Furthermore, the *in silico* promoter analysis showed that the *AtPOL2A* promoter sequence is overrepresented with *cis*-acting regulatory elements associated with cell cycle regulation, meristematic/reproductive tissue-specific pattern of expression and MYB protein recognition, whereas the *AtPOL2B* promoter sequence was mainly enriched with stress-responsive elements; defense-

responsive elements were only detected in the *AtPOL2B* promoter. Our data support the idea that *AtPOL2B* may coexpress with stress-responsive genes. The findings of the present study begin to unravel the potential molecular interactors of *AtPOL2* genes at the molecular level and suggest new avenues for future studies.

Keywords: *Arabidopsis*, DNA polymerase epsilon, protein-protein interactions, transcriptome

Running head: Transcriptome analysis of *AtPOL2A* and *AtPOL2B* mutants

Key Message

Transcriptome analysis of *Arabidopsis DNA polymerase epsilon catalytic subunit A* and *B* mutants provided novel insights into molecular interactors of *AtPOL2*s. Furthermore, the combined analysis of protein-protein interactions and promoter sequences revealed that the duplicated copy of *AtPOL2A* may play a role in mediating stress responses.

Abbreviations:

ABA	Abscisic acid
ABO4	ABSCISIC ACID OVERSENSITIVE 4
AGL	AGAMOUS-LIKE
ATHB2	*Arabidopsis* HOMEOBOX 2
AtPOL2A	DNA polymerase epsilon catalytic subunit A
AtPOL2B	DNA polymerase epsilon catalytic subunit B
ChiC	Class V chitinase
cis-acting regulatory elements	CREs
DEGs	Differentially expressed genes
DNA Pols	DNA polymerases
ESD7	EARLY IN SHORT DAYS 7
FC	Fold change
GLL23	GDSL-like lipase 23
GMI1	GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1
GO	Gene Ontology
GOMo	Gene Ontology for Motifs
HBP1	HISTONE GENE-BINDING PROTEIN 1
HD-Zip	Homeodomain-leucine zipper
KEGG	Kyoto Encyclopedia of Genes and Genomes
KTI1	KUNITZ TRYPSIN INHIBITOR 1
MAF	MADS AFFECTING FLOWERING
MEME	Multiple Em for Motif Elicitation

NIT2	NITRILASE 2
OSM34	Osmotin 34
PARP	POLY (ADP-RIBOSE) POLYMERASE
PDF1.3	Plant defensin 1.3
POLE	DNA polymerase ε
PR1	Pathogenesis-related gene 1
RPA1E	Replication Protein A 1E
SMR7	SIAMESE-RELATED 7
T-DNA	Transfer DNA
TEs	Transposable elements
TFs	Transcription factors
TIL	TILTED
WT	Wild-type

Introduction

In all living organisms, DNA polymerases (DNA Pols) play a central role in the regulation of DNA replication and repair to ensure faithful transmission of genetic information. In most eukaryotes, the members of the family B DNA Pols – α, δ and ε are the main replicative DNA Pols. The knock-out mutants of the main catalytic subunit of DNA polymerase ε (POLE), for example, *tilted* (*til*) *1-1* to *-3* in *Arabidopsis* are embryo lethal (Jenik et al. 2005; Pedroza-Garcia et al. 2019). The present study focuses on the catalytic subunit of the *Arabidopsis* POLE complex that has been shown to have diverse regulatory roles, for example in DNA replication and repair, and epigenetic gene silencing (Ronceret et al. 2005; Yin et al. 2009).

In *Arabidopsis thaliana*, the catalytic subunit of POLE is encoded by two genes (Jenik et al. 2005; Ronceret et al. 2005), *AtPOL2A* (also known as *TIL1*, *ABSCISIC ACID OVERSENSITIVE 4* (*ABO4*), or *EARLY IN SHORT DAYS 7* (*ESD7*); TAIR locus id: AT1G08260) and *AtPOL2B* (also called *TIL2*; TAIR locus id: AT2G27120). The *AtPOL2A* gene is approximately 16 kb long (49 exons) and encodes a protein of 2161 amino acids, whereas the approximately 13 kb long *AtPOL2B* (49 exons) encodes a protein of 2138 amino acids which is 78.5% identical to AtPOL2A (Jenik et al. 2005; Ronceret et al. 2005; del Olmo et al. 2010; Yin et al. 2009). Both AtPOL2A and B contain all the motifs that are necessary for a functional POLE catalytic subunit.

Gene expression studies have detected *AtPOL2A* transcripts in a range of plant tissues i.e. actively growing tissues (floral meristem and flowers until anthesis) and mature tissues (5-week old leaves) but at a relatively low level (Jenik et al. 2005). However, the expression of *AtPOL2B* has been detected mostly under adverse environmental conditions and loss of function of *AtPOL2B* has resulted in no visible phenotype, suggesting that *AtPOL2A* is the main catalytic subunit encoding gene (Ronceret et al. 2005). Over the past, four hypomorphic alleles of *AtPOL2A* (designated as *til1-4*, *abo4-1* and *abo4-2*, and *esd7-1*) have been identified and characterized. The *til1-4* mutant line, which carries two G to A mutations in exon 12 and intron 14 have exhibited longer cell cycles during embryo development, delayed development and larger cells (Jenik et al. 2005). Both *abo4-1*, which has a point mutation that substitutes Glycine to Arginine (at the 534 residue) and *abo4-2*, which harbors a transfer DNA (T-DNA) insert in the exon 12, have shown sensitivity to abscisic acid (ABA), enhanced homologous

recombination, sensitivity to DNA damage agents and constitutive expression of DNA repair genes (Yin et al. 2009). The *esd7-1*, which has a Glycine to Arginine in amino acid position 992, has displayed early flowering and altered growth (del Olmo et al. 2010). Although the molecular function of *AtPOL2B* is not well documented, the detection of developmental defects in embryos and early flowering phenotypes in double mutants of *AtPOL2* genes have led to the discovery of partial functional redundancy of these genes during embryogenesis (Jenik et al. 2005) and vegetative to floral transition (del Olmo et al. 2010).

Transcriptome analysis on *abo4-1* has shown that several genes related to DNA replication and repair, cell cycle control and flower development are differentially expressed in the mutant as compared to that of wild-type (WT) plants (Han et al. 2015; Pedroza-Garcia et al. 2017). Although *AtPOL2A* has been functionally characterized, functional characterization of its isoform is hampered due to the functional redundancy of the two genes. Hence, in the present study, we explore the effect of disrupting *AtPOL2A* and *AtPOL2B* on the transcriptome to gain insight into potential downstream target genes of both *AtPOL2A* and *B* genes and their interactions at a molecular level.

Results

Morphology of *atpol2* mutant lines

The homozygous *atpol2a-1* line showed an early flowering phenotype compared to WT plants when grown under a 16 h photoperiod (Figure 1a). Additionally, these individuals produced considerably smaller siliques (0.85 ± 0.02 cm) with a reduced number of seeds (16 ± 1.63 seeds/ silique) as compared to WT plants (58 ± 1.45 seeds/ silique; Figure 1b). However, none of the *atpol2b* mutants examined here showed any significant phenotypic variation as compared to WT plants, under our experimental conditions.

Fig. 1 Morphological characters of WT and homozygous *atpol2* mutant plants when grown under a 16h photoperiod. (a) 26 d old plants; (b) opened siliques of a WT plant having normal green seeds and a homozygous *atpol2a-1* mutant with aborted seeds (black arrows). The opened siliques of all three *atpol2b* lines were similar to those of WT plants and therefore are not shown here.

Expression of *AtPOL2A* and *B* genes in *atpol2a-1* and *atpol2b-1* to *-3* mutant lines

A total number of 3572 sequence reads were mapped to the *AtPOL2A* locus in the leaf transcriptome of WT, of which 2813 (78.75%) were mapped to the region upstream (chromosome 1: 2588854 to 2594915) of the insertion site of the *atpol2a-1*. Further visualization of the reads mapped to the *AtPOL2A* locus in the *atpol2a-1* mutant derived leaf transcriptome clearly showed that a higher number of reads were mapped in the region downstream of the T-DNA insertion site as compared to the WT (Table 1). Furthermore, analysis of the reads mapped to the *AtPOL2B* locus in transcriptomes derived from *atpol2b-1* to *-3* mutant lines showed that more reads were mapped to regions downstream of the insertion site as compared to the upstream of the insertion site (Table 1); no reads were mapped to the *AtPOL2B* locus in the WT transcriptome. To further confirm the expression patterns detected for *AtPOL2A* and *B* in their respective mutant lines, RT-qPCR analysis was performed using two sets of primers for each gene: one primer pair was designed from a region upstream to the insertion site and another pair from a region downstream of the insertion. This further confirmed that the *atpol2a-1* mutant line produces transcripts from regions both up and downstream of the T-DNA insertion site (Figure 2a). Additionally, all the

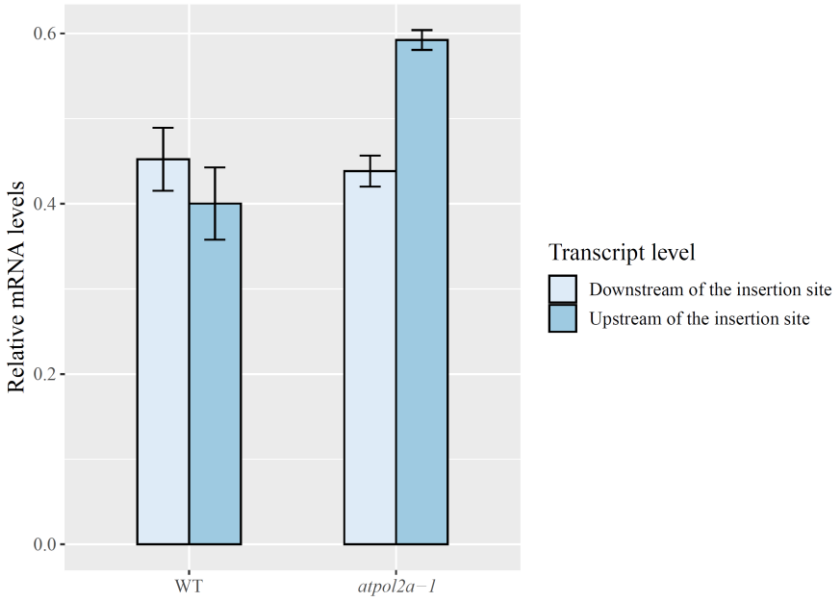
AtPOL2B mutant lines studied here showed notable expression of *AtPOL2B* transcripts downstream of the T-DNA insertion as compared to the WT (Figure 2b).

Further analysis of *AtPOL2* expression in flower tissues of the respective mutant line(s) showed that *AtPOL2A* is expressed at a relatively low level in the *atpol2a-1* mutant line as compared to the WT (Figure 3). Similarly, *AtPOL2B* was also expressed at a relatively low level in all three *AtPOL2B* mutant lines as compared to the WT (Figure 3). It was also evident that both *AtPOL2A* and *B* are expressed at a higher level in flower tissues as compared to leaf tissues of WT *Arabidopsis*.

Table 1 Number of reads mapped to the up and downstream regions of the insertion sites in the *atpol2* mutants

Mutant	**Region upstream of the insertion site**	**Region downstream of the insertion site**	**Total number of mapped reads**	
			Region upstream of the insertion site	Region downstream of the insertion site
atpol2a-1	Chromosome 1: 2588854 to 2594915	Chromosome 1: 2594916 to 2606892	2892 (46.93%)	3271 (53.07%)
atpol2b-1	Chromosome 2: 11581214 to 11581529	Chromosome 2: 11581530 to 11594397	0 (0%)	2636 (100%)
atpol2b-2	Chromosome 2: 11581214 to 11585630	Chromosome 2: 11585631 to 11594397	2 (0.59%)	337 (99.41%)
atpol2b-3	Chromosome 2: 11581214 to 11590152	Chromosome 2: 11590153 to 11594397	19 (5.34%)	337 (94.66%)

(a)



(b)

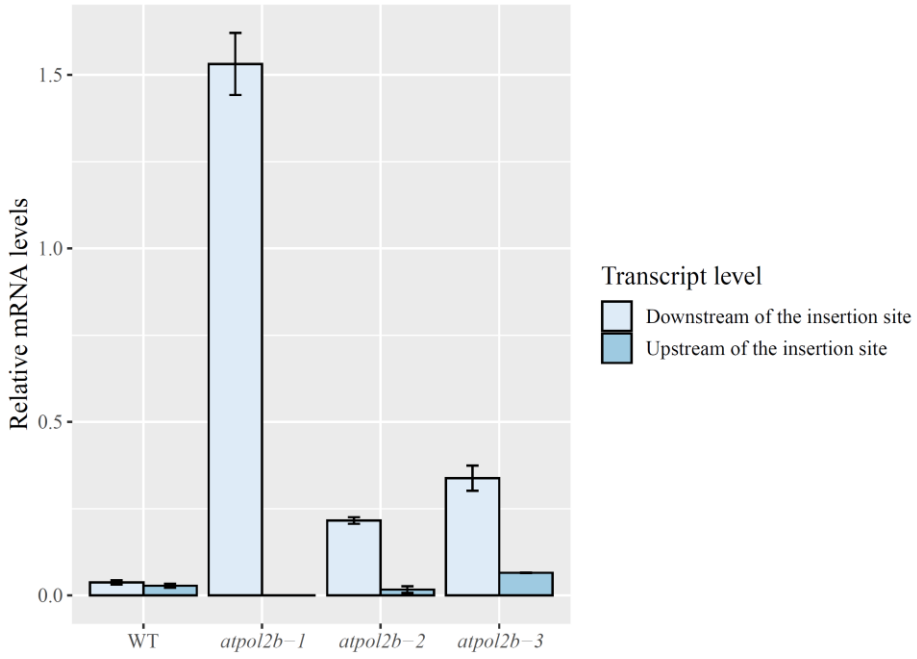


Fig. 2 Expression levels of *AtPOL2A* and *B* transcripts in WT and their mutant lines detected through RT-qPCR. (a) Expression of *AtPOL2A* in leaf tissues of WT and *atpol2a-1*; (b) Expression of *AtPOL2B* in leaf tissues of WT and *atpol2b-1* to *-3*.

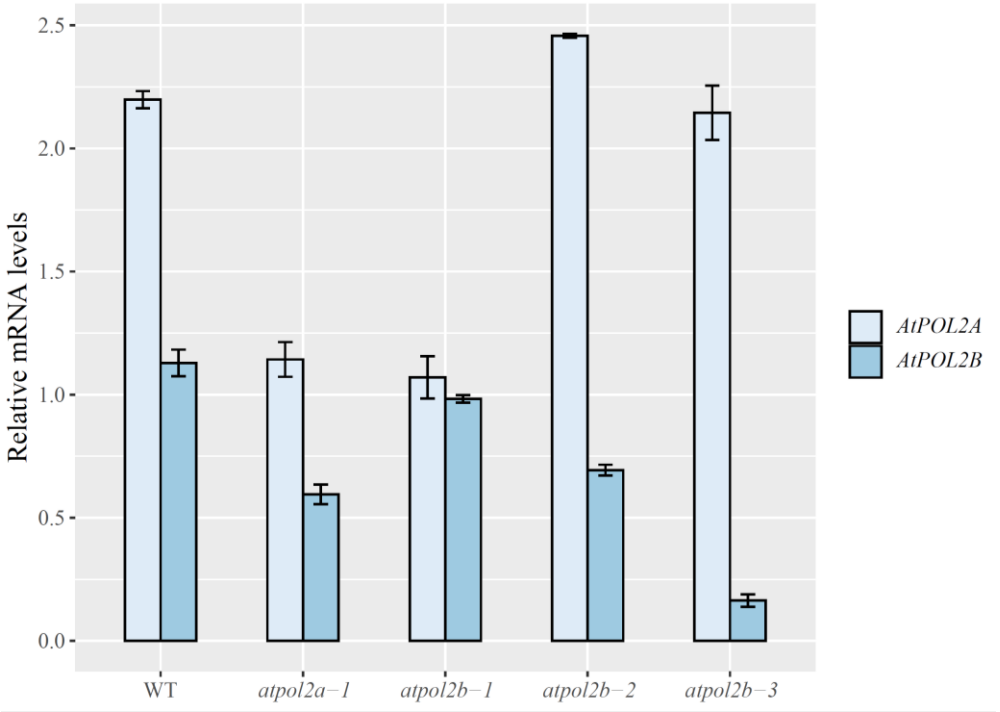


Fig. 3 Expression levels of *AtPOL2*s in flower tissues of WT and mutant lines detected through RT-qPCR

Transcriptome analysis of homozygous *atpol2* mutants

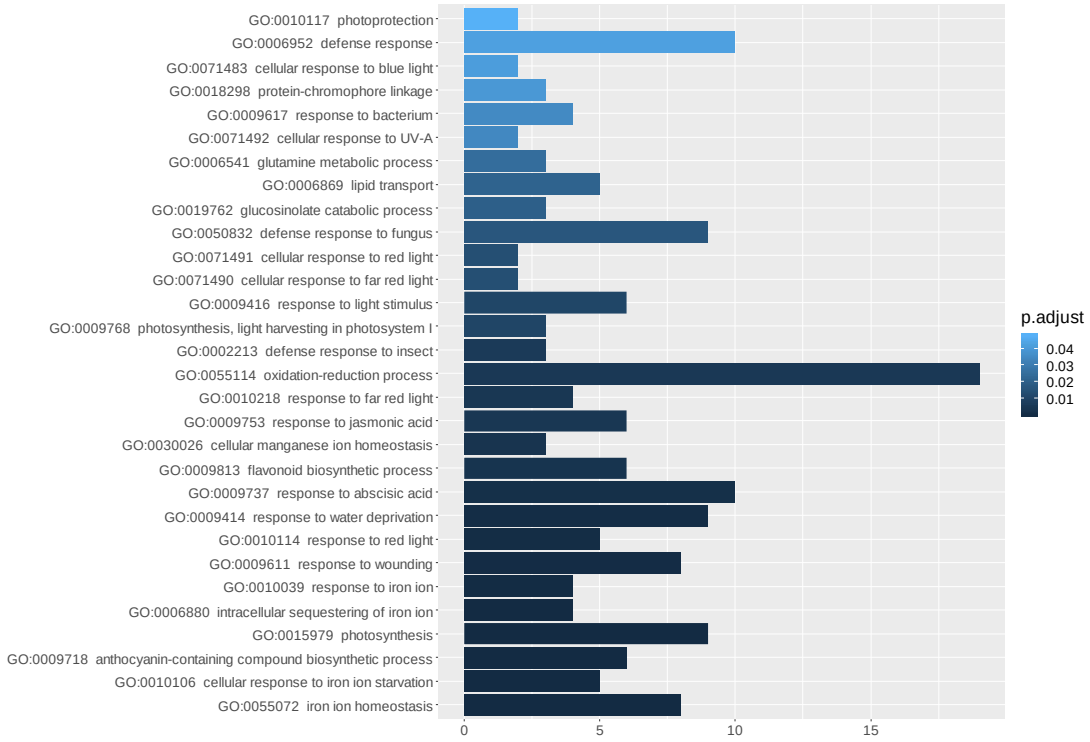
In total, 179025732, 167015310, 180266730, 158718412 and 223844382 reads were obtained from the leaf cDNA libraries of the WT, *atpol2a-1*, *atpol2b-1*, *atpol2b-2* and *atpol2b-3*, respectively. The abundance of 27409 annotated transcripts were detected in at least one of the sequencing libraries examined. The majority of the transcripts were protein-coding. In addition, a considerable fraction of transcriptionally active transposable elements (TEs) was detected in the *atpol2a-1* mutant as compared to *atpol2b* mutants and WT.

Differential gene expression analysis was performed using three computational tools based on the negative binomial distribution to generate a high-confidence list of differentially expressed genes (DEGs). DEGs common to all three analysis tools for each mutant line were considered for the downstream functional analysis. It was evident that at least 198 genes in *atpol2a-1*, 76 genes in *atpol2b-1*, 141 genes in *atpol2b-2* and 67 genes in *atpol2b-3* are differentially expressed as compared to WT (Online Resource 1). Further comparison of the DEGs detected in *atpol2a-1* in the present study and DEGs reported previously for *AtPOL2A* mutants (Han et al. 2015; Pedroza-Garcia et al. 2017) showed

that at least 11 DEGs (AT1G02205 (*CER1*), AT2G18193, AT3G27060 (*TSO2*), AT3G27630 (*SIAMESE-RELATED 7* (*SMR7*)), AT4G02390 (*POLY (ADP-RIBOSE) POLYMERASE 2* (*PARP2*)), AT4G19130 (*Replication Protein A 1E* (*RPA1E*)), AT4G22960, AT4G34510 (*3-ketoacyl-CoA synthase 17*/*KCS17*), AT5G04150 (*BHLH101*), AT5G24280 (*GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1* (*GMI1*)) and AT5G60250) are common to all *AtPOL2A* gene mutants.

Gene Ontology (GO) enrichment analysis of the 198 DEGs detected in *atpol2a-1* showed significant enrichment (p-value ≤ 0.05) for GO terms such as response to defense, oxidation-reduction process, response to UV-A, response to light stimulus, response to water deprivation, response to ABA and response to jasmonic acid (Figure 4a). According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs identified in *atpol2a-1*, 'biosynthesis of secondary metabolites (ath01110)', 'metabolic pathways (ath01100)', 'MAPK signaling pathway (ath04016)', 'tryptophan metabolism (ath00380)', 'flavonoid biosynthesis (ath00941)', 'anthocyanin biosynthesis (ath00942)', 'anthocyanin biosynthesis (ath00942)' and 'photosynthesis – antenna proteins (ath00196)' were significantly enriched.

(a)



(b)

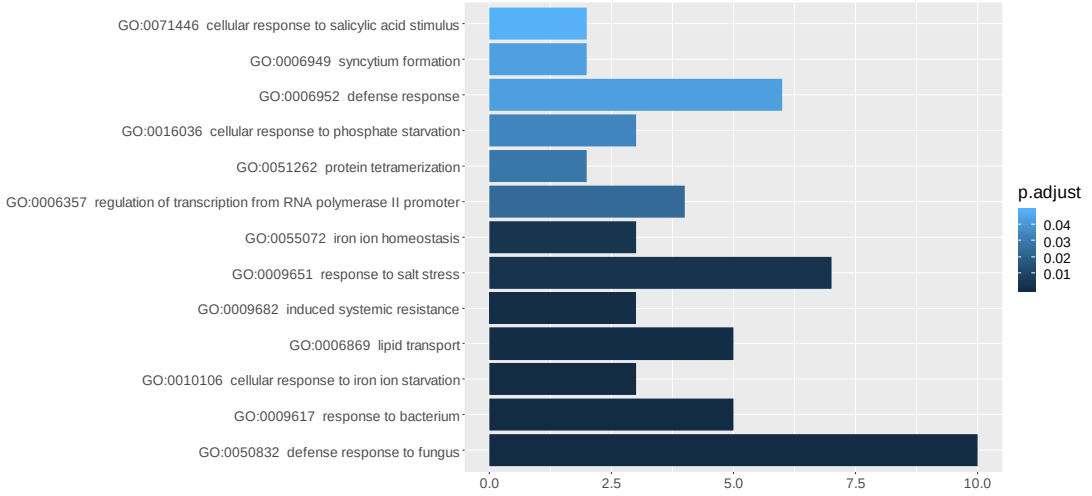
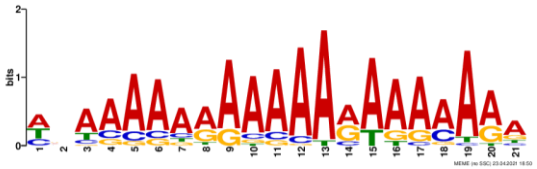
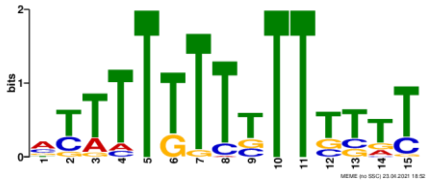
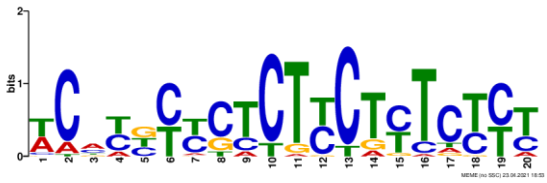
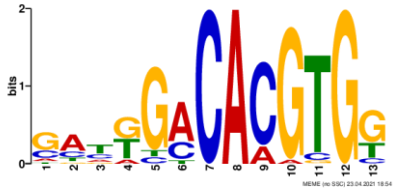


Fig. 4 GO enrichment analysis of DEGs detected in *atpol2* mutants. (a) *atpol2a-1*; (b) *atpol2b* mutants. Only the GO terms related to biological processes at p-value ≤ 0.05 are shown here.

Moreover, analysis of the promoter sequences of all DEGs detected in *atpol2a-1* using the Multiple Em for Motif Elicitation (MEME) suite showed five motifs are significantly overrepresented in DEGs ($p < 0.05$) (Table 2). The possible biological roles of these motifs were studied using Gene Ontology for Motifs (GOMo) tool. It was noted that several motifs were possibly involved in regulating circadian rhythm, and responses to ABA and jasmonic acid. Interestingly, a motif (motif-3: WCMYBYYCTCTYCTYTCYYY) that may function in regulating megagametogenesis was discovered in promoter regions of 28 DEGs. Further analysis of these discovered motifs using PlantPAN 3.0 showed that motif-1 and 2 may interact with MADS-box transcription factor AGAMOUS-LIKE 20 (AGL 20; AT2G45660), whereas motif-4 may interact with ABA INSENSITIVE 5 (ABI5; AT2G36270) and/or PHYTOCHROME INTERACTING FACTOR 1 (PIF 1; AT2G20180).

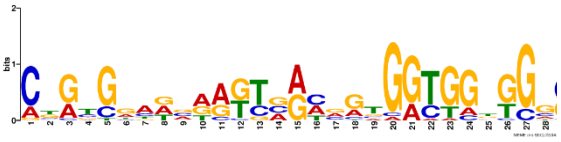
Table 2 Motifs discovered in promoter regions of DEGs detected in *atpol2a-1* mutant using MEME Suite

Motif sequence and logo	**Possible biological role(s) (enriched GO terms at p-value ≤ 0.05)**
Motif 1 Consensus sequence: WVAAAAARAAAAARAAAAAA 	Transmembrane receptor protein tyrosine kinase signaling pathway; regulation of transcription, DNA-dependent; circadian rhythm; response to jasmonic acid stimulus; cytoskeleton organization; cytokinin mediated signaling pathway; protein amino acid phosphorylation
Motif 2 Consensus sequence: MYTTTTTTTTTTTTT 	Circadian rhythm; regulation of transcription
Motif 3 Consensus sequence: WCMYBYYCTCTYCTYTCYYY 	Regulation of transcription; transmembrane receptor protein tyrosine kinase signaling pathway; protein amino acid phosphorylation; plastid organization; megagametogenesis; leaf development; root development; abaxial cell fate specification
Motif 4 Consensus sequence: GATKGACACGTGG 	Photosynthesis; response to ABA stimulus; response to far-red light; response to wounding; response to water deprivation; response to blue light; circadian rhythm

Motif 5	Mitochondrial transport

Consensus sequence:

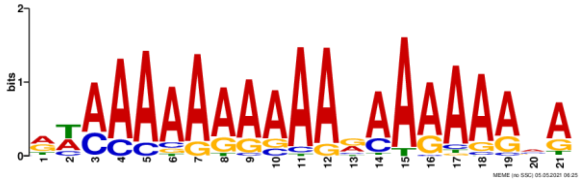
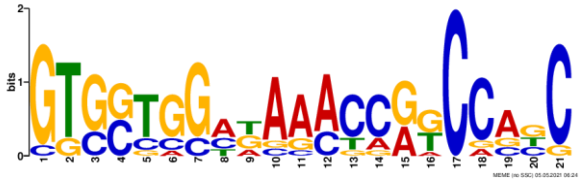
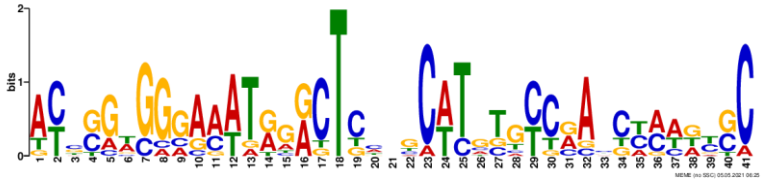
KRYGRADVRRKTVRHVGTGGTGGKKGGS



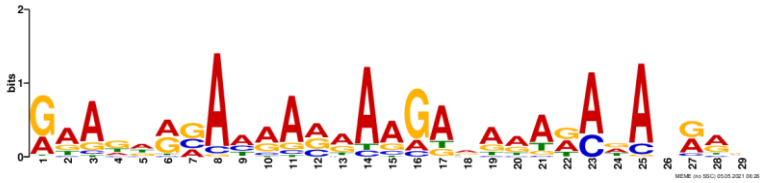
Comparison of DEGs detected in the three *atpol2b* mutants (*atpol2b-1* to *-3*) showed a substantial difference among the mutants; only 11 [AT1G54010 (*GDSL-like lipase 23* (*GLL23*)), AT1G73260 (*KUNITZ TRYPSIN INHIBITOR 1* (*KTI1*)), AT2G10940, AT2G14247, AT2G14610 (*Pathogenesis-related gene 1* (*PR1*)), AT2G26010 (*Plant defensin 1.3* (*PDF1.3*)), AT3G44300 (*NITRILASE 2* (*NIT2*)), AT4G11650 (O*smotin 34* (*OSM34*)), AT4G19810 (*Class V chitinase* (*ChiC*)), AT5G48490 and AT5G65080 (*MADS AFFECTING FLOWERING 5* (*MAF5*))] were common to all three mutants. Of these, AT2G10940, AT2G14247, AT2G14610, AT2G26010 and AT5G65080 showed differential expression in both *atpol2a* and *b* mutants. GO enrichment analysis of DEGs detected in *atpol2b* mutants revealed a significant enrichment (p-value ≤ 0.05) for GO terms relating to response to defense, induced systemic resistance, response to salt stress, lipid transport and iron homeostasis (Figure 4b). The KEGG pathway enrichment analysis of DEGs identified in *atpol2b* mutants showed that the pathway entries 'MAPK signaling pathway (ath04016)', 'biosynthesis of secondary metabolites (ath01110)', 'metabolic pathways (ath01100)', 'tryptophan metabolism (ath00380)' were significantly enriched.

Analysis of the promoter sequences of DEGs identified in *atpol2b* mutants revealed enrichment of five motifs (Table 3). The majority of DEGs contained motifs I (68/72 genes) and IV (66/72 genes) in their promoter regions which may interact with MADS-box transcription factor AGL 20 (AT2G45660).

Table 3 Motifs discovered in the promoter regions of DEGs detected in *atpol2b* mutants using MEME Suite

Motif sequence and logo	**Possible biological role(s)** (enriched GO terms at p-value ≤ 0.05)	**Corresponding transcription factor**
Motif I Consensus sequence: RWAAAAAAAAAARAAAA AAVA 	Regulation of transcription; response to jasmonic acid stimulus; cytoskeleton organization	AT2G45660
Motif II Consensus sequence: GTGSTGGMDAAACCRKCC ABC 	Translation	
Motif III Consensus sequence: AYBSGWGGGAAATRGRCT CHNGCWTGTGYCRANCHM WDYSC 		

Motif IV Consensus sequence: GAAKDRVAMRAVRAAGAR ARARADANRRD	Transmembrane receptor protein: tyrosine kinase signaling pathway; protein amino acid phosphorylation; regulation of transcription; leaf morphogenesis; flower development; polarity specification of adaxial/abaxial axis; DNA endoreduplication; circadian rhythm	AT2G45660



Protein-protein interaction (PPI) analysis

Interestingly, the PPI network constructed for the proteins encoded by the predicted DEGs in *atpol2a-1* using the StringDB showed that genes encoding RPA1E (AT4G19130), TSO2 (AT3G27060) and PARP2 (AT4G02390) may interact with AtPOL2A (Figure 5). Further, it was noted that RPA1E interacts with GIM1 (AT5G24280), and AT4G22960, and PARP2 interacts with AT4G22960, AT3G27630 and AT1G07500. TSO2 was shown to interact with AT3G49160. (Figure 5).

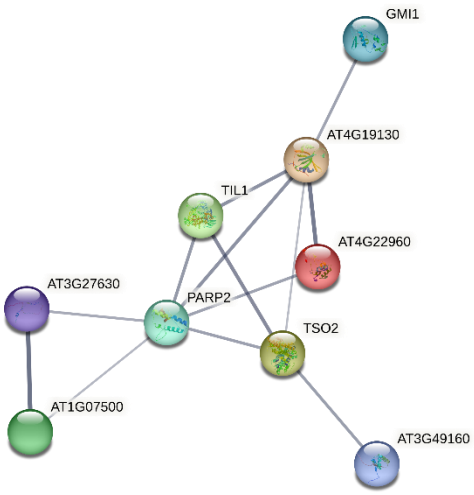


Fig. 5 A simplified PPI network of proteins encoded by DEGs detected in *atpol2a-1* showing only the important interactions. This network was constructed using StringDB. Network nodes represent proteins (colored nodes represent query proteins; empty nodes represent proteins of unknown structures and filled

nodes represent some 3D structure that is known or predicted). The thickness of network edges represents the strength of data support (low (0.150), medium (0.400), high (0.700) and highest (0.900)). Complete PPI network is provided in Online Resource 2.

Similarly, PPI network analysis of the proteins encoded by the predicted DEGs in *atpol2b* mutants also showed known and predicted interactions among genes (Online Resource 3). Interestingly, of the 11 DEGs in all three *atpol2b*, genes encoding seven proteins (KTI1 (AT1G73260), PR1 (AT2G14610), PDF1.3 (AT2G26010), NIT2 (AT3G44300), OSM34 (AT4G11650) ChiC (AT4G19810) and AT2G10940) showed known/predicted interactions among them. However, none of the proteins encoded by the DEGs were interacting with AtPOL2B/TIL2 based on interactome data available in StringDB. This formed the basis for further investigation of potential interactions between AtPOL2s and the proteins encoded by the DEGs in *atpol2* mutants through protein-protein docking.

Docking sites of potential interactors of AtPOL2s

Analysis of the AtPOL2A and B interacting proteins through the StringDB showed that DNA polymerase epsilon subunit B2 (DPB2), DNA polymerase alpha subunit B (POLA2), DNA polymerase alpha primase subunit (POLA3), DNA replication licensing factor MCM2, DNA primase large subunit (EMB2813) and DNA polymerase delta small subunit (POLD2) are common interactors AtPOL2s. Docking scores obtained for these proteins are given in Table 4. The docking results demonstrated that the cluster size ranged from 38 to 152, while the balanced score ranged from -808.9 to -1174.5. Therefore, this scale was used to validate the interactions between AtPOL2s and several proteins encoded by the DEGs detected in the present study.

Table 4 ClusPro docking scores for AtPOL2s and their known interactors

			Complex with AtPOL2A		Complex with AtPOL2B	
Gene identifier	**Protein name**	**Subcellular localization**	**Cluster size**	**Balanced score**	**Cluster size**	**Balanced score**
AT5G22110	DNA polymerase epsilon subunit B2	Nucleus	152	-1010.3	93	-1009.2
AT5G41880	DNA primase	Cytoplasm	103	-901.1	61	-808.9
AT1G67630	DNA polymerase alpha subunit B	Nucleus	61	-910.1	43	-950.5
AT1G67320	DNA primase large subunit	Nucleus	60	-1044.9	39	-940.2
AT1G44900	DNA replication licensing factor MCM2	Nucleus	44	-1080.8	72	-1174.5
AT2G42120	DNA polymerase delta small subunit	Nucleus	38	-919.5	53	-840.8

Structural modeling of 22 proteins encoded by the DEGs of *atpol2a* and *b* mutants using SWISS-MODEL predicted the 3D structures of 20 proteins, while two proteins (AT3G27630 and AT2G14247) failed to locate template protein sequences for modeling. Out of the structures predicted, 90% (18) protein structures showed over 85% residues clustered in favored regions in the Ramachandran plot, indicating stable conformations. Further analysis of these proteins revealed seven complexes with AtPOL2A and six complexes with AtPOL2B (*cluster size* > *38* and *balanced score* less than -808.9; Table 5). Of these, three (CER1, RPA1E and AT5G60250) and two (PR1 and AT5G48490) proteins indicated notably significant interactions with AtPOL2A and AtPOL2B proteins, respectively (Figure 6). Furthermore, these proteins also showed *cluster size* > *60* and *balanced score* < *-900* when docked with the POLE catalytic modules (6HV8 and 6HV9) of *S. cerevisiae* (Online Resource 4).

Table 5 ClusPro docking scores for AtPOL2s and their potential interactors (significant protein interactions are highlighted with an asterisk (*cluster size > 60* and *balanced score < -900*)

	Gene identifier	Protein name	Subcellular localization	Cluster size	Balanced score
AtPOL2A	AT1G02205*	Very-long-chain aldehyde decarbonylase CER1	Endoplasmic reticulum membrane	117	-1101.1
	AT5G60250*	Uncharacterized protein	Cytoplasm	94	-1172.4
	AT5G04150	Transcription factor bHLH101	Nucleus	91	-806
	AT5G24280*	Structural maintenance of chromosomes flexible hinge domain-containing protein GMI1	Cytoplasm	85	-941.4
	AT4G19130*	Replication protein A 70 kDa DNA-binding subunit E	Nucleus	78	-910.8
	AT4G02390	Poly (ADP-ribose) polymerase 2	Nucleus	74	-810.5
	AT4G34510*	3-ketoacyl-CoA synthase 17	Cytoplasm	71	-929.2
	AT3G27060	Ribonucleoside-diphosphate reductase small chain C	Cytoplasm	56	-795.1
	AT4G22960	Uncharacterized protein	Cytoplasm	51	-1150.1
	AT2G18193	AAA-ATPase At2g18193	Mitochondrion membrane	34	-919
	AT3G27630	Cyclin-dependent protein kinase inhibitor SMR7	Nucleus	-	-
AtPOL2B	AT2G26010	Defensin-like protein 14	Secreted	211	-783.9
	AT1G73260	Kunitz trypsin inhibitor 1	Secreted	134	-876.5
	AT5G48490*	Uncharacterized protein	Secreted	128	-956.7
	AT2G14610*	Pathogenesis-related protein 1	Secreted	124	-932.3

	AT4G11650	Osmotin-like protein OSM34	Vacuole	115	-857.2
	AT2G10940	Uncharacterized protein	Cytoplasm	105	-871.3
	AT4G19810	Class V chitinase	Secreted	72	-757.4
	AT3G44300*	Nitrilase 2	Cytoplasm	66	-1009.3
	AT1G54010	Inactive GDSL esterase/lipase-like protein 23	Secreted	51	-787.5
	AT5G65080	MADS AFFECTING FLOWERING 5	Nucleus	41	-548.1
	AT2G14247	Uncharacterized protein	Nucleus	-	-

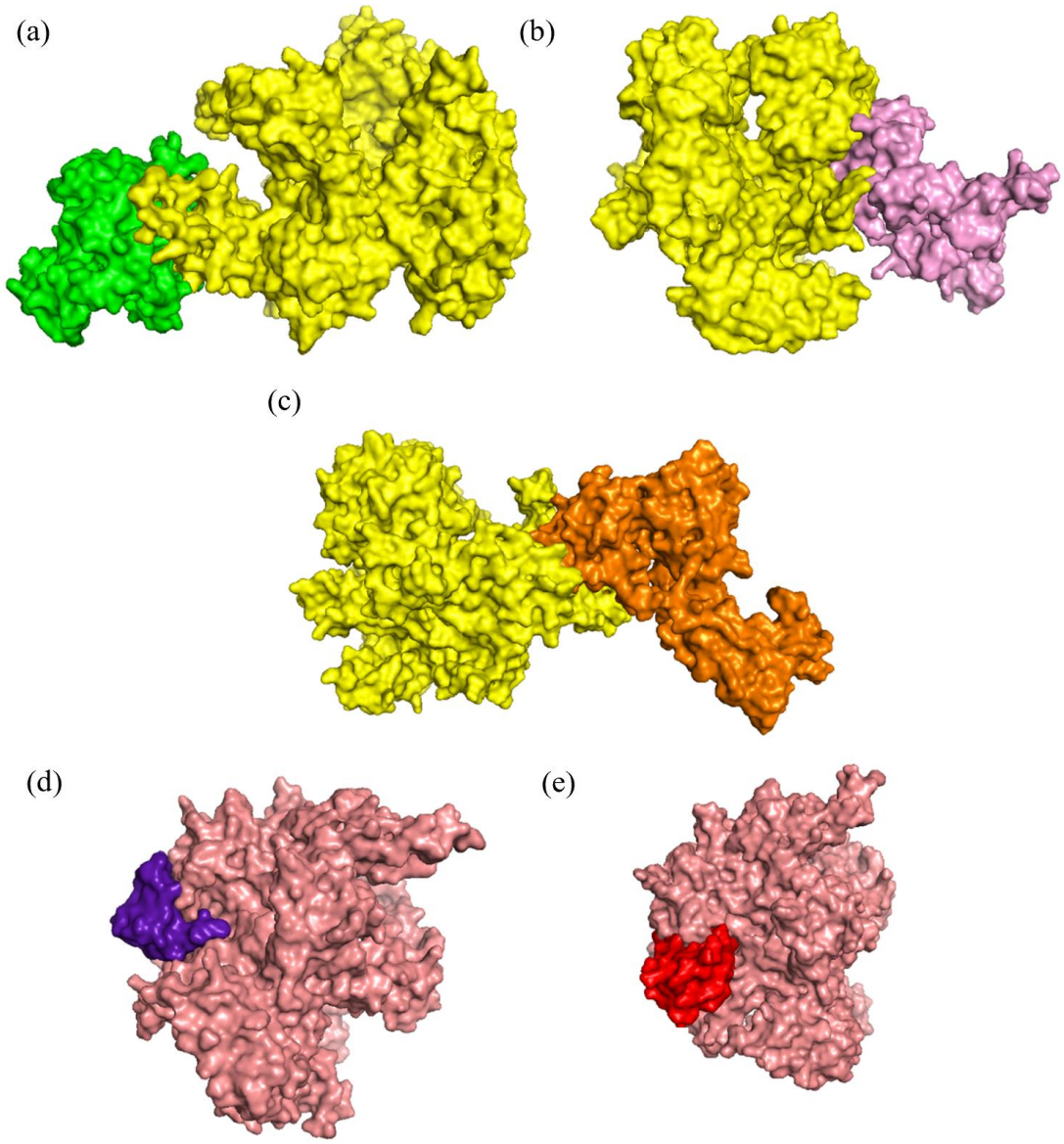


Fig. 6 The 3-dimensional surface structures of potential AtPOL2A and AtPOL2B interactions. AtPOL2A and AtPOL2B are shown in yellow and salmon colours, respectively. The interaction sites of these protein docking complexes were visualized by PyMol. (a) AtPOL2A-CER1 interaction; (b) AtPOL2A-AT5G60250 interaction; (c) AtPOL2A- RPA1E interaction; (d) AtPOL2B-PR1 interaction; (e) AtPOL2A-AT5G48490 interaction.

***In silico* analysis of *cis*-acting regulatory elements (CREs) of promoter sequences of *AtPOL2*s**

The *AtPOL2A* gene contains a 2091 bp long 5′ untranslated region. Hence, 2500 bp upstream regions of both *AtPOL2* genes, from the translation start site were selected and analyzed using the PLACE and PlantCARE databases to determine the putative CREs. The results obtained through the PLACE tool showed that a total of 113 and 110 putative CREs are distributed in *AtPOL2A* and *AtPOL2B* promoters, respectively. Of these, 80 elements were found in both gene promoters (descriptions of these CREs are provided in Online Resource 5). For instance, the *AtPOL2A* gene promoter contained several CREs contributing to tissue-specific expression including pollen-specific elements (POLLEN2LELAT52 and VOZATVPP), embryo-specific elements (CARGCW8GAT), fruit-specific elements (TGTCACACMCUCUMISIN) and meristem-specific elements (E2F1OSPCNA). This gene also contained E2FANTRNR, an element related to up-regulation of the promoter at G1/S transition, and MYBCOREATCYCB1, a core CRE in *CYCLIN B1;1* in its promoter. In addition, several homeodomain-leucine zipper (HD-Zip) protein binding sites (ATHB2ATCONSENSUS and ATHB5ATCORE) and MYB-related binding sites (i.e. BOXLCOREDCPAL, CCA1ATLHCB1, IBOX, MYB2AT, MYB2CONSENSUSAT, MYBPLANT, MYBPZM) were detected. Interestingly, a hexamer motif found in the *Arabidopsis* histone H4 promoter (HEXAMERATH4) was also detected in the *AtPOL2A* promoter. By contact, QELEMENTZMZM13 (Quantitative elements (Q-element)), which is known to enhance activity in pollen-specific genes; AUXRETGA1GMGH3, an auxin-responsive CRE; LTRE1HVBLT49, a low-temperature responsive element; PREATPRODH, a hypoosmolarity-responsive element; DRE2COREZMRAB17, SBOXATRBCS and MYBATRD22 (drought and/or ABA-responsive elements); SORLIP5AT (a light-responsive element); HEXMOTIFTAH3H4, a conserved hexamer motif known to interact with histone DNA binding proteins; WBBOXPCWRKY1 and WBOXNTCHN48 (pathogen/elicitor responsive elements) were detected only in the *AtPOL2B* promoter.

Similarly, the analysis of the promoter sequences using PlantCARE showed that both *AtPOL2* promoters contained hormone-responsive elements such as auxin-responsive, gibberellin responsive and methyl jasmonate responsive. However, ABA-responsive and salicylic acid responsive elements were found only in the *AtPOL2A* promoter sequence (Figure 7a). Among the environmental stress-responsive elements, drought-responsive elements were only detected in the *AtPOL2A* promoter, and defense-responsive and low-temperature responsive elements were only detected in the *AtPOL2B* promoter

(Figure 7b). Further, it was noted that both *AtPOL2* genes contained a considerable number of light-responsive elements. As expected, an element responsible for the regulation of the cell cycle was only detected in the *AtPOL2A* promoter sequence (Figure 7c).

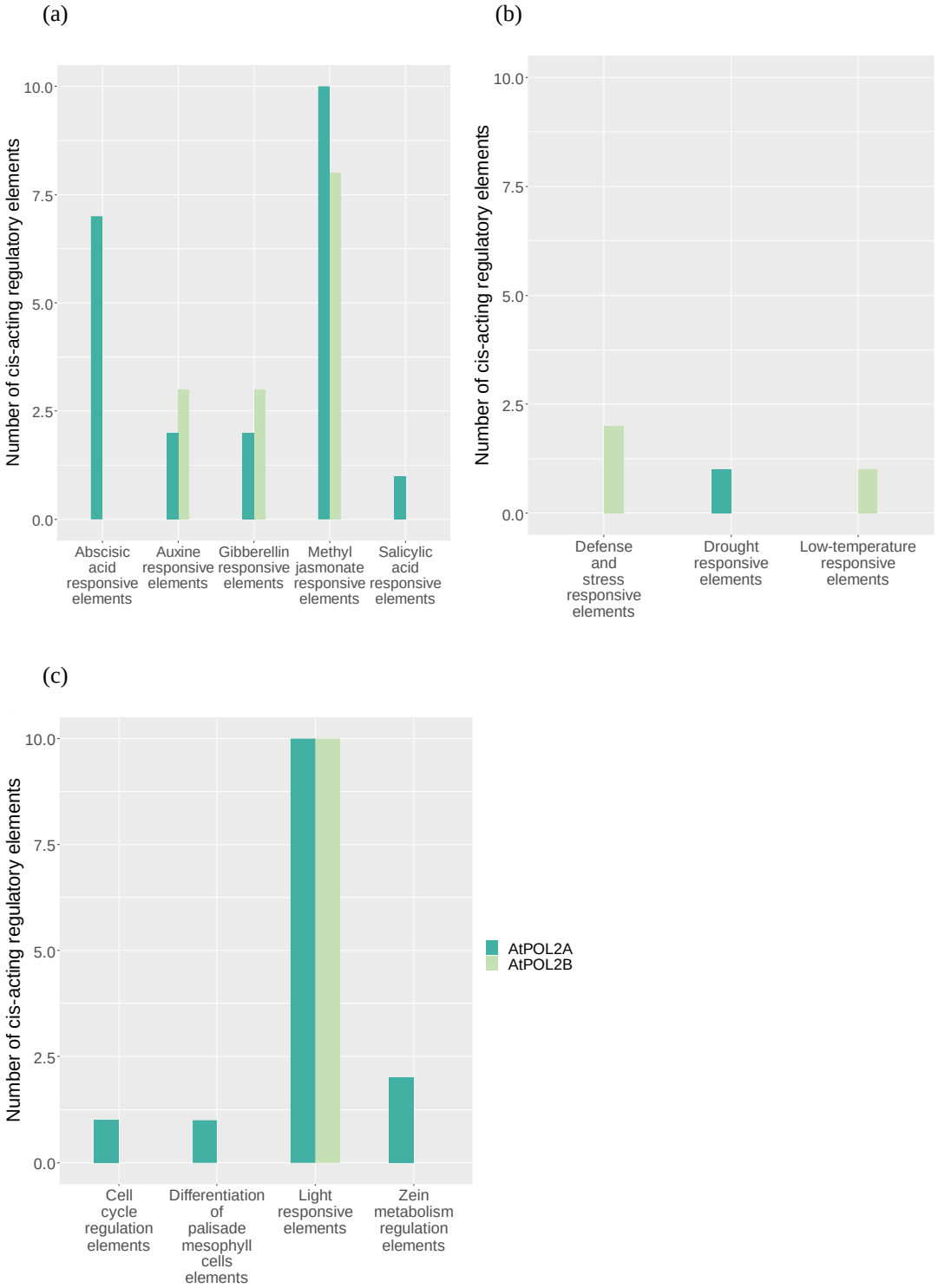


Fig. 7 Abundance of CREs identified using PlantCARE in the 2500 bp upstream region of each *AtPOL2* gene. (a) Phytohormone-responsive CREs; (b) Environmental stress-responsive CREs; (c) Plant growth and developmental processes related CREs.

Discussion

The publicly available T-DNA insertional mutant line collections have played an important role in functional studies to explore gene functions and regulatory networks. All the mutant lines examined in the present study showed a change in the expression pattern of the mutated gene as compared to that of the WT. Consistent with previous findings (Yin et al. 2009), the *AtPOL2A* mutant examined in the present study (*atpol2a-1*) was also early flowering and exhibited smaller siliques than WT or *atpol2b* mutants. Moreover, as reported in previous studies (Jenik et al. 2005; Ronceret et al. 2005; del Olmo et al. 2010), none of the homozygous *AtPOL2B* mutant lines examined here exhibited visible phenotypic variation as compared to the WT. However, as all of them showed a considerable change in the expression pattern of the mutated gene, they were used in the subsequent transcriptomic analysis.

KEGG pathway and GO enrichment analyses of DEGs identified in *atpol2* mutants showed that no pathway/GO term associated with DNA replication and repair was significantly enriched. However, further analysis of the DEGs detected in *atpol2a-1* identified several important genes involved in DNA replication including *TSO2*, *SMR7*, *RPA1E*, *PARP1*, *GMI1* and AT3G49160 (a pyruvate kinase-related protein) (Andre et al. 2007). In *Arabidopsis*, *TSO2* is one of the three members of the ribonucleotide reductase (RNR) small subunit genes, which transcribes predominantly at the S-phase of the cell cycle (Wang and Liu, 2006). Roa et al. (2009) report E2F target genes including *AtPOL2A* and *TSO2* coexpress in the DNA repair network. Moreover, *tso2-1 rnr21-1* double mutants have shown induced expression of several molecular markers associated with DNA damage and repair i.e. *PARP1*, *PARP2* and *ATRAD51* (Wang and Liu, 2006). Recently, upregulation of *PARP2*, *RPA1E* and *POL2A* have been observed in loss-of-function mutants of the F-box protein FBL17. In *Arabidopsis*, *FBL17* functions in cell-cycle regulation (Gentric et al. 2020) and is required for cell division during pollen development. Our PPI analysis also showed that uncharacterized AT3G49160 interacts with FER1; both of these genes are known to be expressed in response to hydrogen peroxide (Petit et al. 2001; op den Camp et al. 2003; Laloi et al. 2007).

Mutations in several genes related to DNA/histone assembly and replication have been reported to release the transcriptional gene silencing of pericentromeric *Athila* retrotransposon at the *transcriptional silent*

information (*TSI*) i.e. *abo4-1* and *abo4-2* mutants (Yin et al. 2009; Han et al. 2015). Consistent with previous studies, we observed that the silencing of the majority of pericentromeric TEs is released in the *atpol2a-1* mutant. In a recent study, Bourguet and his colleagues have noted that the significant release of heterochromatin silencing in *AtPOL2A* gene mutants is due to increased levels of DNA methylation and histone H3 lysine 9 methylation, and thus *AtPOL2A* is important to maintain heterochromatin structure and function (Bourguet et al. 2020).

Interestingly, our transcriptome analysis showed that *AtPOL2B* is more likely to coexpress with stress-responsive genes; several genes related to biotic stress responses were differentially expressed in all three *atpol2b* mutants examined. These include *GLL23*, *KTI1*, *PR1* (a salicylic acid marker gene) (Van Loon and Van Strien, 1999; Pecenkova et al. 2017), *PDF1.3* (a plant defensin) (Biedrzycki et al. 2011), *OSM34* and *ChiC*. Phylogenetic analysis of GDSL-type esterases/lipases has identified four clades (I –IV), and GLL 23 has been placed in clade IIIa. T-DNA knockout lines of many genes in this clade have shown increased resistance to biotic stresses (Lai et al. 2017). Moreover, KTI1 (serine protease inhibitor) is known to be involved in the regulation of programmed cell death during biotic stress tolerance in *Arabidopsis* (Li et al. 2008; Arnaiz et al. 2018). In addition to defense-related processes, OSM34 has shown to be effective against abiotic stresses (Capelli et al. 1997; Hao et al. 2015; Bashir et al. 2020). Furthermore, the *ChiC* gene which was also differentially expressed in all *atpol2b* mutants is known to be induced by the plant stress-related hormones (ABA, jasmonic acid) and by the stress resulting from the elicitor flagellin, NaCl and osmosis (Ohnuma et al. 2011). In addition to these genes, a member of a NITRILASE 1-subfamily, *NIT2* was also differentially expressed in all *atpol2b* mutants. NIT2 regulates hypocotyl elongation in response to high temperatures (van der Woude et al. 2021).

Under our experimental conditions, *MAF5* was differentially expressed in both *atpol2a* and *b* mutants compared to WT. This finding is contradictory to the results obtained by del Olmo et al. (2010) where no significant alteration in floral development genes including *MAF5* has been observed in *AtPOL2A* mutant *esd7-1* (del Olmo et al. 2010). However, Han et al. (2015) also report differential expression of *MAF* in the *AtPOL2A* mutant, *abo4-1* (Han et al. 2015). Previous studies have shown that early flowering mutants of *Arabidopsis* exhibit a reduction in levels of H3 methylation (H3K4me3 and H3K36me2) at the flowering locus genes *FLOWERING LOCUS C* (*FLC*), *MAF1*, *MAF4* and *MAF5* (Cao et al. 2008;

Xu et al. 2008). Therefore, it is interesting to examine the role of *AtPOL2A* in controlling the methylation status of *MAF5* and thereby controlling the expression of *MAF5*.

The first insight into the promoter elements of *AtPOL2* genes came from a study conducted by Ronceret and co-workers in 2005. However, since then, no comprehensive analysis of the promoter sequences of these two genes has been performed. Therefore, an *in silico* promoter analysis was conducted to explore the potential interactors of *AtPOL2*s at the molecular level. As reported in Ronceret et al. (2005), one E2FANTRNR element (also known as E2Fa element), which is a binding site for E2F-like proteins, was detected upstream of the translation start site of *AtPOL2A*. In addition, E2F1OSPCNA, which is also an E2F transcription factor binding site, was identified in the promoter region of *AtPOL2A*. In *Arabidopsis*, consensus E2F-binding sites have been identified in the promoter regions of several regulators of the cell cycle (e.g. D-type cyclin *CYCD3* and *CDC6*) suggesting a potential role of E2F and E2F-like proteins in G1-to-S phase transition of the cell cycle (de Jager et al. 2001). Furthermore, Kosugi and Ohashi (2002) report, E2F binding sites are required for meristematic tissue-specific activity of proliferating cell nuclear antigen promoters (PCNA) in rice and tobacco (Kosugi and Ohashi, 2002). Neither E2FANTRNR nor E2F1OSPCNA elements were found upstream of the translation start site of *AtPOL2B*. However, E2FCONSENSUS (an E2F consensus sequence) and Myb-binding core (AACGG) motifs were detected in both promoter regions. Myb-binding core, AACGG is responsible for cell cycle phase-independent transcriptional activation of B-cyclin *CYCB1;1* in *Arabidopsis* (Planchais et al. 2002). Analysis of the microarray data has further confirmed that *Arabidopsis* E2F target genes are expressed during G1 and S phases and most of the proteins encoded by these genes function in cell cycle regulation, DNA replication and chromatin dynamics (Vandepoele et al. 2005).

Moreover, both *AtPOL2A* and *B* promoters contained motifs reported to be conserved in the promoter regions of plant and animal histone genes. HEXAMERATH4, a hexamer motif (CCGTCG) found in histone H4 promotes (Chaubet et al. 1996), and HEXMOTIFTAH3H4, a hexamer motif (ACGTCA) found in H3 and H4 histone gene promoters (Mikami et al. 1987) were found in *AtPOL2A* and *AtPOL2B* promoters, respectively. The hexamer motif CCGTCG found in the wheat histone genes has shown to be essential for a meristematic-tissue specific pattern of gene expression i.e. expression in organs containing meristematic tissues such as roots and flower buds than in fully expanded leaves and therefore, promoters

having CCGTCG motifs could exhibit a cell-division inducible activity (Chaubet et al. 1996). Moreover, the conserved hexamer motif ACGTCA found in the plant histone H3 and H4 genes bind specifically with the nuclear protein HISTONE GENE-BINDING PROTEIN 1 (HBP1) (Mikami et al. 1989a, b; Tabata, 1991; Terada et al. 1995), which is more abundant in actively proliferating cells. It is reported that HBP1 may be involved in cell-cycle dependent expression of histone genes in plants. Detection of an ACGTCA motif upstream of the translation initiation codon of *AtPOL2B* hence suggests cell-cycle dependent transcription of *AtPOL2B*.

Notably, DOFCOREZM and GTGANTG10 were abundantly found in both *AtPOL2* promoter sequences. DOFCOREZM elements specifically recognize plant-specific Dof (DNA-binding with one finger) transcription factors (TFs), which act as transcriptional regulators involved in many plant-specific biological processes (Yanagisawa and Schmidt, 1999; Yanagisawa, 2000, 2004). Engel et al. (2005) report the presence of core binding site for Dof transcription factors in promoter sequences of two sperm-specific genes *AtGEX1* (AT5G55490) and *AtGEX2* (AT5G49150). Furthermore, GTGANTG10 is a GATA motif found within the promoter of the tobacco late pollen gene g10; Rogers et al. (2001) have demonstrated that this gene is expressed preferentially and maximally in mature pollen.

Interestingly, two co-dependent CREs, POLLEN1LELAT52 (AGAAA) and POLLEN2LELAT52 (TCCACCATA) were found in the *AtPOL2A* promoter. Both of these CREs are required for the tissue-specific regulation of late pollen genes during pollen development (Bate and Twell, 1998). Also, two VOS-binding motifs were found upstream of the translation initiation codon of *AtPOL2A*. *Arabidopsis* VOS proteins (AtVOZ1 and AtVOZ2) function as transcriptional activators (Mitsuda et al. 2004). The pollen-specific transcriptional activation of the *Arabidopsis AVP1* gene that encodes vacuolar H+-pyrophosphatase (V-PPase) is controlled by binding of AtVOZ1 and AtVOZ2 transcription factors to a 38 bp pollen-specific CRE (Mitsuda et al. 2001). Moreover, two CWWWWWWWWG motifs (a variant of CArG motif) that interact with AGAMOUS-like 15 (AGL15) were found in the *AtPOL2A* promoter. The MADS domain family transcription factor AGL15 preferentially accumulates during embryo development (Tang and Perry, 2003). Further, TGTCACACMCUCUMISIN (TGTCACA) motif was found in the promoter region of *AtPOL2A*. This enhancer element is responsible for the fruit-specific expression of the *cucumisin* gene in melon (*Cucumis melo* L.) (Yamagata et al. 2002). Detection of

AGL15 binding sites and TGTCACACMCUCUMISIN motif in the *AtPOL2A* gene further confirms the function of *AtPOL2A* in *Arabidopsis* embryo development. By contrast, three QELEMENTZMZM13 (Q-element) and 16 POLLEN1LELAT52 motifs were identified within the regulatory region of *AtPOL2B*. The Q-element has been identified in a promoter region of a pollen-specific gene (*ZM13*) in maize (Hamilton et al. 1998). However, the pollen-specific expression of *ZM13* is known to be regulated by two motifs (pollen-specific and Q-element) in the promoter; the presence of only the Q-element in the promoter has no ability to cause the expression of *ZM13* in pollen (Hamilton et al. 1998). These findings further confirm *AtPOL2B* is more likely to exhibit no or little expression in reproductive tissues.

The present study also found two binding sites for *Arabidopsis* HOMEOBOX 2 (ATHB2) and two binding sites for AtHB5 in the *AtPOL2A* promoter. Of these HD-Zip class proteins, ATHB2, which expresses during both vegetative and reproductive stages of plant growth, is found to be strongly induced in response to Red to Far-Red light ratio (Carabelli et al. 1993, 2013), whereas AtHB5 appears to be involved in ABA-related responses and/or water deficit conditions (Henriksson et al. 2005). In addition, several MYB protein recognition sequences were found in the promoter region of *AtPOL2A* (BOXLCOREDCPAL, CCA1ATLHCB1 (Wang et al. 1997), MYB2CONSENSUSAT, MYBPLANT (a binding site for a flower-specific MYB protein (Sablowski et al. 1994), MYBPZM, MYB2AT and IBOX (Rose et al. 1999). In plants, MYB transcription factors are involved in regulating many biological processes including hormonal signal transduction and abiotic/biotic stress tolerance (Urao et al. 1993; Abe et al. 2003), and more importantly, these TFs are involved in the regulation of the cell cycle at the G2/M transition (Cominelli and Tonelli, 2009).

We also noted that several light-responsive elements (GATABOX) and stress-responsive elements (MYCCONSENSUSAT, ARR-1 binding elements (Sakai et al. 2000; Ross et al. 2004), WRKY71OS (Xie et al. 2005) are distributed within both *AtPOL2* promoters. However, as compared to the *AtPOL2A* promoter sequence, many stress-responsive elements were found in the *AtPOL2B* promoter. i.e. abiotic stress-responsive elements (LTRE1HVBLT49, PREATPRODH, DRE2COREZMRAB17, SBOXATRBCS and MYBATRD22) and biotic stress-responsive (pathogen/elicitor responsive) elements (WBBOXPCWRKY1 and WBOXNTCHN48). Transcriptome analysis of *atpol2b* mutants

together with the promoter analysis confirm that *AtPOL2B* is more likely to play a role in modulating stress responses in *Arabidopsis*.

Methods

Planting materials and growth conditions

Seeds of *A. thaliana* (ecotype Columbia-0) T-DNA insertion lines of *AtPOL2A* [SALK_096314C (*atpol2a-1*)] and *AtPOL2B* [N859810 (*atpol2b-1*), SALK_001413 (*atpol2b-2*), SALK_056503 (*atpol2b-3*)]; WT (N70000) were obtained from the Nottingham *Arabidopsis* Stock Centre, UK. Seeds were germinated and grown under controlled environmental conditions at 25^{o}C, 60% relative humidity and 16 h photoperiod.

Morphological characterization

The average silique length and the number of days to flower were recorded for mutants and WT plants. Green siliques were harvested 12 d after flowering from 4-6 week old plants and opened using fine needles under a dissecting microscope (Leica MZ9.5, Germany). Images of opened siliques were captured using a Leica DFC290 (Germany) camera and edited in Adobe Photoshop.

Preparation of cDNA libraries and whole transcriptome sequencing

Two fully expanded leaves were harvested from each confirmed homozygous mutant line and WT plants at the time of development of the first flower bud. Total RNA was extracted using the RNeasy® Plant Mini Kit (Qiagen, UK) according to the manufacturer's instructions and on-column DNase digestion was performed using RNase-Free DNase I (1500 Kunitz units). Extracted RNA was quantified through Agilent BioAnalyzer 2100. cDNA libraries were constructed and amplified using random primers. Figure 8 summarizes the main steps involved in the construction of Illumina sequencing libraries. Adapter sequences used for the library preparation and amplification are provided in the Online Resource 6. For sequencing, two samples were pooled and loaded into each flow cell and sequenced using Illumina HiSeq 2000 platform (2 x 50 bp read length).

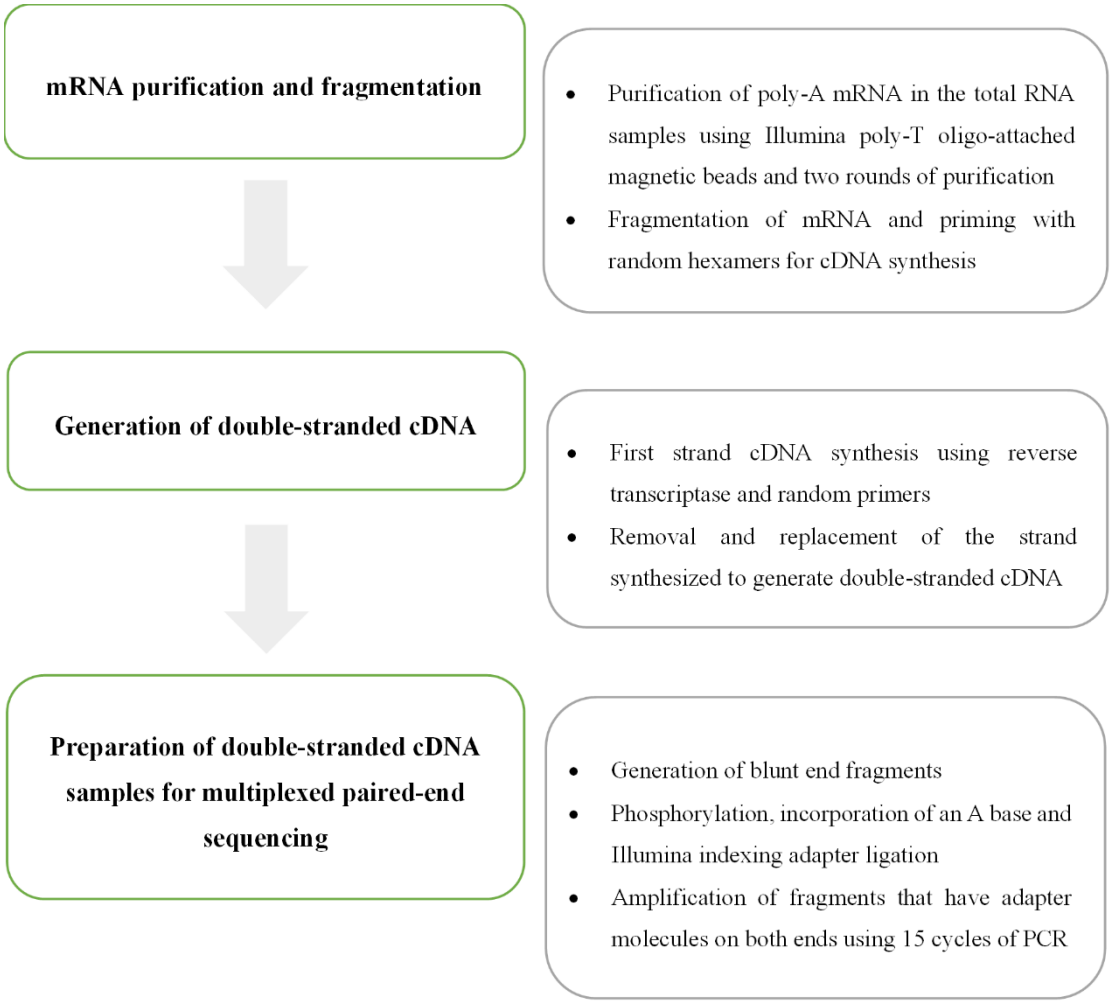


Fig. 8 Preparation of Illumina sequencing libraries

Aligning sequence reads to the *A. thaliana* reference genome

The quality of raw sequence reads generated through sequencing was assessed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were quality and adapter trimmed using Trimmomatic (v0.39) (Bolger et al. 2014). Only reads that passed the Phred quality score (Q-score > = Q30) were included in the mapping. TopHat 2.1.1 (Trapnell et al. 2009) (http://tophat.cbcb.umd.edu) with Bowtie2 2.3.4.1 (http://bowtie-bio.sourceforge.net/index.shtml) was used to align sequences to the *Arabidopsis* reference genome (TAIR10), allowing only 2 base mismatches per read alignment (Trapnell et al. 2012). The Cufflinks (2.1.1) was used to assemble reads that were mapped and to normalize transcript levels by means of Fragments Per Kilobase of exon per Million fragments mapped (FPKM) (Trapnell et al. 2010). In addition, HTSeq-count 0.11.1 (Anders et al. 2015) (https://htseq.readthedocs.io/en/release_0.11.1/count.html) program with the intersection 'union' option

was also employed to estimate gene-level expression of read alignments sorted by 'name' using SAMtools 1.11 (http://www.htslib.org/doc/1.11/samtools.html).

Identification of DEGs and downstream analysis

Three tools based on the negative binomial distribution, cuffdiff (a part of the Cufflinks package) (Trapnell et al. 2013), DESeq (Anders and Huber, 2010) (https://bioconductor.org/packages/release/bioc/html/DESeq.html) and edgeR (Robinson et al. 2010) (https://bioconductor.org/packages/release/bioc/html/edgeR.html) were used for the differential gene expression analysis. The fold change (FC) ratio of transcripts between samples was used to filter DEGs; the genes that fell within $-1 \geq \text{Log}_2 \text{ FC} \geq 1$ with an adjusted p-value of ≤ 0.05 were considered as differentially expressed. Functional enrichment analyses were performed using the DAVID v6.8 (Huang et al. 2007). GO terms and KEGG pathways were considered enriched if the associated Benjamini-Hochberg adjusted p-value was less than 0.05.

Interactome analysis and docking sites of potential interactors of AtPOL2s

In order to identify proteins that interact with AtPOL2A and B, an interactome analysis was done using STRING database (https://string-db.org/). Proteins encoding genes that were differentially expressed in *atpol2a* and *atpol2b* mutants were used as the set of potential interactors and default parameters were used to construct the network (interaction score > 0.400).

The protein sequences of the 22 DEGs were retrieved from the TAIR database using 'Bulk data retrieval' tool. The 3D structures of the proteins were generated using SWISS-MODEL (https://swissmodel.expasy.org/). Each predicted structure was confirmed using Ramachandran plots using the MOLProbity program (Williams et al. 2018). In addition, the subcellular localization of the proteins was explored using the LocTree3 tool (https://rostlab.org/services/loctree3/). The ClusPro server 2.0 (https://cluspro.bu.edu/), which employs a rigid body protein-protein docking (Kozakov et al. 2017), was used to predict the docking sites for the AtPOL2 proteins with their target proteins. The best docking structure was selected based on the largest cluster size and minimum balanced score and visualized through PyMOL molecular graphics system (version 2.0). In addition, proteins interacting with AtPOL2A and AtPOL2B were identified using StringDB, structurally modeled through SWISS-

MODEL. Further, the existing 3D structures of POL2 proteins in *S. cerevisiae* were retrieved from PDB (PDB accession numbers: 6HV8 and 6HV9; https://www.rcsb.org/). These protein structures were docked with the predicted structures of the putative targets of AtPOL2 using ClusPro 2.0.

Confirmation of *AtPOL2* expression through RT-qPCR

Total RNA was extracted from leaf tissues of mutant and WT plants using the RNeasy® Plant Mini Kit (Qiagen, UK) according to the manufacturer's instructions and on-column DNase digestion was performed using RNase-Free DNase I (1500 Kunitz units). First-strand cDNA was synthesized using SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, UK) using 500 ng of total RNA. mRNA sequences of *AtPOL2A* (AT1G08260) and *AtPOL2B* (AT2G27120) and *Arabidopsis* ubiquitinating enzyme *UBC21* (AT5G25760) were retrieved from GenBank (http://www.ncbi.nlm.nih.gov). To detect the expression level of *AtPOL2A* in the *atpol2a-1* mutant, the mRNA (NM_001331765) region upstream (2249-3534 nt) and downstream (6695-8206 nt) of the insertion sites were selected to design forward/reverse primer pairs. Similarly, to detect the expression level of *AtPOL2B* in *atpol2b* mutants, one primer pair was designed from mRNA (NM_001336102) region upstream (1-211 nt) of the insertion site in *atpol2b-1* and another primer was designed from the downstream region (5266-6065 nt) of the insertion site in *atpol2b-3*. The primer sets were designed using the NCBI-Primer BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast). Each primer pair was designed in a way that either both primers were separated by at least one intron on the corresponding genomic DNA or one primer spanned exon-exon junction. Primer sequences used in RT-qPCR analysis are given in Online Resource 6.

RT-qPCR was performed in 20.0 µl final reaction volume containing 10.0 µl of 2X SensiMix SYBR® No-ROX master mix (Bioline, UK), 0.3 µM of forward and reverse primers and 1.0 µl of cDNA. PCR was carried out on a Rotor-Gene 6000 System (Corbett Life Sciences, UK). Thermal cycling conditions consisted of one cycle of 95°C for 10 min; then 40 cycles of denaturing at 95°C for 15 s and annealing/elongation at 60°C for 1 min, followed by melting curve analysis from 60 to 95°C at the rate of 0.5°C per s. Each sample was repeated thrice. The fluorescence was recorded during the annealing/elongation step in each cycle. Primer specificity was verified by the melting curve analysis, and by DNA sequencing of the PCR products. The relative standard curve method was employed for

gene expression analysis. Five-fold serial dilutions of cDNA were used to construct standard curves. The slope of the standard curve was used to estimate the PCR amplification efficiency. The expression of the target *AtPOL2* genes was normalized to the *UBC21* reference gene. Data were obtained and analyzed using Rotor-Gene 6000 series software Version 1.7.

In silico promoter analysis

Promoter sequences (2500 bp upstream of the translation start site) of each *AtPOL2* gene were retrieved from the TAIR (https://www.arabidopsis.org/). PLACE (database of plant c*is*-acting regulatory DNA elements; http://www.dna.affrc.go.jp/PLACE/) and PlantCARE databases were employed for screening of putative CREs (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Additionally, novel motifs were discovered using MEME.

In addition, the promoter sequences (1000 bp upstream of transcription start site) of each DEG identified in *atpol2a* and *atpol2b* mutants were retrieved from the TAIR (https://www.arabidopsis.org/). The motif-based sequence analysis tool, MEME Suite (https://meme-suite.org/meme/tools/meme) was used to discover overrepresented motifs in DEGs; biological roles of predicted motifs were determined using the GOMo tool (https://meme-suite.org/meme/tools/gomo).

Declarations

Ethics approval

Not Applicable

Consent for publication

Not Applicable

Data availability

The datasets generated during the current study are available in the European Bioinformatics Institute ArrayExpress repository, E-MTAB-2465 (https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-2465?query=E-MTAB-2465%20), and all data generated during the study are included in this published article and its supplementary information files.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Funding

Not Applicable

Authors' contributions

A.M.W. conceived and designed the study, performed mutant screening and expression studies, analyzed and interpreted transcriptome data, and drafted the manuscript. T.M.H. analyzed and interpreted transcriptome data and helped to draft the manuscript. K. K. D. conducted the protein-protein docking study and helped to draft the manuscript. I. U. carried out RT-qPCR analysis. J.M.D. conceived and designed the study and helped to draft the manuscript. All authors read and approved the final manuscript.

References

1. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63–78.
2. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106.
3. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–9.
4. Andre C, Froehlich JE, Moll MR, Benning C (2007) A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in *Arabidopsis*. Plant Cell 19:2006–22.
5. Arnaiz A, Talavera-Mateo L, Gonzalez-Melendi P, Martinez M, Diaz I, Santamaria ME (2018) Arabidopsis kunitz trypsin inhibitors in defense against spider mites. Front Plant Sci 9:986.
6. Bashir MA, Silvestri C, Ahmad T, Hafiz IA, Abbasi NA, Manzoor A, et al (2020) Osmotin: a cationic protein leads to improve biotic and abiotic stress tolerance in plants. Plants 9:992.
7. Bate N, Twell D (1998) Functional architecture of a late pollen promoter: pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. Plant Mol Biol 37:859–69.
8. Biedrzycki ML, Venkatachalam L, Bais HP (2011) Transcriptome analysis of *Arabidopsis thaliana* plants in response to kin and stranger recognition. Plant Signal Behav 6:1515–24.
9. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.

10. Bourguet P, Lopez-Gonzalez L, Gomez-Zambrano A, Pelissier T, Hesketh A, Potok ME, et al. (2020) DNA polymerase epsilon is required for heterochromatin maintenance in *Arabidopsis*. Genome Biol 21:283.
11. Cao Y, Dai Y, Cui S, Ma L (2008) Histone H2B monoubiquitination in the chromatin of *FLOWERING LOCUS C* regulates flowering time in *Arabidopsis*. Plant Cell 20:2586–602.
12. Capelli N, Diogon T, Greppin H, Simon P (1997) Isolation and characterization of a cDNA clone encoding an osmotin-like protein from *Arabidopsis thaliana*. Gene 191:51–6.
13. Carabelli M, Sessa G, Baima S, Morelli G, Ruberti I (1993) The *Arabidopsis Athb-2* and *-4* genes are strongly induced by far-red-rich light. Plant J 4:469–79.
14. Carabelli M, Turchi L, Ruzza V, Morelli G, Ruberti I (2013) Homeodomain-Leucine zipper II family of transcription factors to the limelight: central regulators of plant development. Plant Signal Behav 8:e25447.
15. Chaubet N, Flenet M, Clement B, Brignon P, Gigot C (1996) Identification of *cis*-elements regulating the expression of an *Arabidopsis* histone H4 gene. Plant J 10:425–35.
16. Cominelli E, Tonelli C (2009) A new role for plant R2R3-MYB transcription factors in cell cycle regulation. Cell Res 19:1231–2.
17. de Jager SM, Menges M, Bauer UM, Murra JA (2001) Arabidopsis E2F1 binds a sequence present in the promoter of S-phase-regulated gene AtCDC6 and is a member of a multigene family with differential activities. Plant Mol Biol 47:555–68.
18. del Olmo I, Lopez-Gonzalez L, Martin-Trillo MM, Martinez-Zapater JM, Pineiro M, Jarillo JA (2010) *EARLY IN SHORT DAYS 7* (*ESD7*) encodes the catalytic subunit of DNA polymerase epsilon and is required for flowering repression through a mechanism involving epigenetic gene silencing. Plant J 61:623–36.
19. Engel ML, Holmes-Davis R, McCormick S (2005) Green sperm. Identification of male gamete promoters in Arabidopsis. Plant Physiol 138:2124–33.
20. Gentric N, Masoud K, Journot RP, Cognat V, Chaboute ME, Noir S, et al. (2020) The F-box-like protein FBL17 is a regulator of DNA-damage response and colocalizes with RETINOBLASTOMA RELATED1 at DNA lesion sites. Plant Physiol 183:1295–305.
21. Hamilton DA, Schwarz YH, Mascarenhas JP (1998) A monocot pollen-specific promoter contains separable pollen-specific and quantitative elements. Plant Mol Biol 38:663–9.

22. Han YF, Huang HW, Li L, Cai T, Chen S, He XJ (2015)The cytosolic iron-sulfur cluster assembly protein MMS19 regulates transcriptional gene silencing, DNA repair, and flowering time in *Arabidopsis*. PLoS One 10:e0129137.

23. Hao J, Wu W, Wang Y, Yang Z, Liu Y, Lv Y, et al. (2015) *Arabidopsis thaliana* defense response to the ochratoxin A-producing strain (*Aspergillus ochraceus* 3.4412). Plant Cell Rep 34:705–19.

24. Henriksson E, Olsson AS, Johannesson H, Johansson H, Hanson J, Engstrom P, et al. (2005) Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships. Plant Physiol 139:509–18.

25. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. (2007) The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8:R183.

26. Jenik PD, Jurkuta RE, Barton MK (2005) Interactions between the cell cycle and embryonic patterning in *Arabidopsis* uncovered by a mutation in DNA polymerase epsilon. Plant Cell 17:3362–77.

27. Kosugi S, Ohashi Y (2002) E2F sites that can interact with E2F proteins cloned from rice are required for meristematic tissue-specific expression of rice and tobacco proliferating cell nuclear antigen promoters. Plant J 29:45–59.

28. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. (2017) The ClusPro web server for protein–protein docking. Nat Protoc 12:255–78.

29. Lai CP, Huang LM, Chen LO, Chan MT, Shaw JF (2017) Genome-wide analysis of GDSL-type esterases/lipases in *Arabidopsis*. Plant Mol Biol 95:181–97.

30. Laloi C, Stachowiak M, Pers-Kamczyc E, Warzych E, Murgia I, Apel K (2007) Cross-talk between singlet oxygen- and hydrogen peroxide-dependent signaling of stress responses in *Arabidopsis thaliana*. Proc Natl Acad Sci 104:672–7.

31. Li J, Brader G, Palva ET (2008) Kunitz trypsin inhibitor: an antagonist of cell death triggered by phytopathogens and fumonisin b1 in *Arabidopsis*. Mol Plant 1:482–95.

32. Mikami K, Nakayama T, Takefumi, Kawata, Tabata T, Iwabuchi M (1989a) Specific interaction of nuclear protein HBP-1 with the conserved hexameric sequence ACGTCA in the regulatory region of wheat histone genes. Plant Cell Physiol 30:107–19.

33. Mikami K, Sakamoto A, Takase H, Tabata T, Iwabuchi M (1989b) Wheat nuclear protein HBP-1 binds to the hexameric sequence in the promoter of various plant genes. Nucleic Acids Res 17:9707–17.

34. Mikami K, Tabata T, Kawata T, Nakayama T, Iwabuchi M (1987) Nuclear protein(s) binding to the conserved DNA hexameric sequence postulated to regulate transcription of wheat histone genes. FEBS Lett 223:273–8.

35. Mitsuda N, Hisabori T, Takeyasu K, Sato MH (2004) VOZ; isolation and characterization of novel vascular plant transcription factors with a one-zinc finger from *Arabidopsis thaliana*. Plant Cell Physiol 45:845–54.

36. Mitsuda N, Takeyasu K, Sato MH (2001) Pollen-specific regulation of vacuolar H+-PPase expression by multiple *cis*-acting elements. Plant Mol Biol 46:185–92.

37. Ohnuma T, Numata T, Osawa T, Mizuhara M, Lampela O, Juffer AH, et al. (2011) A class V chitinase from *Arabidopsis thaliana*: gene responses, enzymatic properties, and crystallographic analysis. Planta 234:123–37.

38. op den Camp RG, Przybyla D, Ochsenbein C, Laloi C, Kim C, Danon A, et al. (2003) Rapid induction of distinct stress responses after the release of singlet oxygen in Arabidopsis. Plant Cell 15:2320–32.

39. Pecenkova T, Pleskot R, Zarsky V (2017) Subcellular localization of *Arabidopsis* pathogenesis-related 1 (PR1) protein. Int J Mol Sci 18(4):825.

40. Pedroza-Garcia JA, De Veylder L, Raynaud C (2019) Plant DNA polymerases. Int J Mol Sci 20:4814.

41. Pedroza-Garcia JA, Mazubert C, Del Olmo I, Bourge M, Domenichini S, Bounon R, et al. (2017) Function of the plant DNA polymerase epsilon in replicative stress sensing, a genetic analysis. Plant Physiol 173:1735–49.

42. Petit JM, Briat JF, Lobréaux S (2001) Structure and differential expression of the four members of the *Arabidopsis thaliana* ferritin gene family. Biochem 359:575–82.

43. Planchais S, Perennes C, Glab N, Mironov V, Inze D, Bergounioux C (2002) Characterization of *cis*-acting element involved in cell cycle phase-independent activation of Arath;*CycB1*;1 transcription and identification of putative regulatory proteins. Plant Mol Biol 50:111–27.

44. Roa H, Lang J, Culligan KM, Keller M, Holec S, Cognat V, et al. (2009) Ribonucleotide reductase regulation in response to genotoxic stress in Arabidopsis. Plant Physiol 151:461–71.

45. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–40.

46. Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, et al. (2001) Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. Plant Mol Biol 45:577–85.

47. Ronceret A, Guilleminot J, Lincker F, Gadea-Vacas J, Delorme V, Bechtold N, et al. (2005) Genetic analysis of two Arabidopsis DNA polymerase epsilon subunits during early embryogenesis. Plant J 44:223–36.

48. Rose A, Meier I, Wienand U (1999) The tomato I-box binding factor LeMYBI is a member of a novel class of Myb-like proteins. Plant J 20:641–52.

49. Ross EJ, Stone JM, Elowsky CG, Arredondo-Peter R, Klucas RV, Sarath G (2004) Activation of the *Oryza sativa* non-symbiotic haemoglobin-2 promoter by the cytokinin-regulated transcription factor, ARR1. J Exp Bot 55:1721–31.

50. Sablowski RW, Moyano E, Culianez-Macia FA, Schuch W, Martin C, Bevan M (1994) A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. EMBO J 13:128–37.

51. Sakai H, Aoyama T, Oka A (2000) *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. Plant J 24:703–11.

52. Tabata T (1991) Leucine-zipper type transcription factors of wheat. Bot Mag Tokyo 104:171–81.

53. Tang W, Perry SE (2003) Binding site selection for the plant MADS domain protein AGL15: *an in vitro* and *in vivo* study. J Biol Chem 278:28154–9.

54. Terada R, Nakayama T, Iwabuchi M, Shimamoto K (1995) A type I element composed of the hexamer (ACGTCA) and octamer (CGCGGATC) motifs plays a role(s) in meristematic expression of a wheat histone H3 gene in transgenic rice plants. Plant Mol Biol 27:17–26.

55. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31:46–53.

56. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–11.

57. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–78.

58. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–5.

59. Urao T, Yamaguchi-Shinozaki K, Urao S, Shinozaki K (1993) An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell 5:1529–39.

60. van der Woude L, Piotrowski M, Klaasse G, Paulus JK, Krahn D, Ninck S, et al. (2021) The chemical compound "Heatin" stimulates hypocotyl elongation and interferes with the Arabidopsis NIT1-subfamily of nitrilases. Plant J 106(6):1523–40

61. Van Loon LC, Van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol 55:85–97.

62. Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Gruissem W, et al. (2005) Genome-wide identification of potential plant E2F target genes. Plant Physiol 139:316–28.

63. Wang C, Liu Z (2006) *Arabidopsis* ribonucleotide reductases are critical for cell cycle progression, DNA damage repair, and plant development. Plant Cell 18:350–65.

64. Wang ZY, Kenigsbuch D, Sun L, Harel E, Ong MS, Tobin EM (1997) A Myb-related transcription factor is involved in the phytochrome regulation of an Arabidopsis *Lhcb* gene. Plant Cell 9:491–507.

65. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. (2018) MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci 27:293–315.

66. Xie Z, Zhang ZL, Zou X, Huang J, Ruas P, Thompson D, et al. (2005) Annotations and functional analyses of the rice *WRKY* gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Cell Physiol 137:176–89.

67. Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, Steinmetz A, et al. (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. Mol Cell Biol 28:1348–60.

68. Yamagata H, Yonesu K, Hirata A, Aizono Y (2002) TGTCACA motif is a novel *cis*-regulatory enhancer element involved in fruit-specific expression of the *cucumisin* gene. J Biol Chem 277:11582–90.

69. Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. Plant J 17:209–14.

70. Yanagisawa S (2004) Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. Plant Cell Physiol 45:386–91.

71. Yanagisawa S (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. Plant J 21:281–8.

72. Yin H, Zhang X, Liu J, Wang Y, He J, Yang T, et al. (2009) Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase epsilon mutation in *Arabidopsis*. Plant Cell 21:386–402.