1 Comprehensive analysis of *Arabidopsis thaliana* DNA polymerase epsilon
2 catalytic subunit A and B mutants – an insight into differentially expressed genes
3 and protein-protein interactions

4 Anushka M. Wickramasuriya^{1*}, Thulani M. Hewavithana¹, Kithmee K. de Silva¹,
 5 Ihsan Ullah², Jim M. Dunwell²

6

7 ¹ Department of Plant Sciences, Faculty of Science, University of Colombo, Colombo
 8 03, Sri Lanka

9 ² School of Agriculture, Policy and Development, University of Reading, UK

10

11 *Corresponding author

12 Email: anushka@pts.cmb.ac.lk

13

14 Abstract

15 One of the main replicative enzymes in most eukaryotes, DNA polymerase ε (POLE),
 16 is composed of four subunits, namely a single catalytic and three regulatory subunits.
 17 In *Arabidopsis*, the catalytic subunit of POLE is encoded by two genes: *Arabidopsis*
 18 *thaliana* DNA POLYMERASE EPSILON CATALYTIC SUBUNIT A (*AtPOL2A*) and B
 19 (*AtPOL2B*). Although studies have shown *AtPOL2A* to be involved in various
 20 biological processes, the role of *AtPOL2B* is unclear. Here, we investigated the
 21 transcriptomes of both *atpol2a* and *atpol2b* mutants, and the promoter sequences to
 22 provide a better insight into the targets of *AtPOL2s* at the molecular level. In the present
 23 study, leaf cDNA libraries of four *AtPOL2* mutants (*atpol2a-1* and *atpol2b-1*, -2 and -
 24 3) were sequenced using the Illumina platform. Analysis of gene expression profiles
 25 identified a total of 198, 76, 141 and 67 differentially expressed genes in *atpol2a-1*,

26 *atpol2b-1*, *atpol2b-2* and *atpol2b-3*, respectively; the majority of pericentromeric
 27 transposable elements were transcriptionally active in *atpol2a-1* as compared to *atpol2b*
 28 mutants and wild type. Protein-protein interaction network analysis and molecular
 29 docking identified three (CER1, RPA1E and AT5G60250) and two (PR1 and
 30 AT5G48490) proteins as potential interactors (*cluster size* > 60 and *balanced score* <
 31 -900) of AtPOL2A and AtPOL2B, respectively; Interestingly, these five proteins also
 32 showed a significant interaction between POLE catalytic subunit of *Saccharomyces*
 33 *cerevisiae*. Our *in silico* promoter analysis showed that the *AtPOL2A* promoter
 34 sequence is overrepresented with *cis*-acting regulatory elements (CREs) associate with
 35 cell cycle regulation, meristematic/reproductive tissue-specific pattern of expression
 36 and MYB protein recognition, whereas the *AtPOL2B* promoter sequence was mainly
 37 enriched with stress-responsive elements. The information provided here has led to the
 38 identification of targets of AtPOL2s at the molecular level and CREs putatively
 39 associated with the regulation of *AtPOL2s*. To our knowledge, this study provides the
 40 first comparative transcriptome profiling of single-gene mutants of *AtPOL2s*.

41

42 Short title: Transcriptome analysis of Arabidopsis DNA polymerase ϵ catalytic subunit
 43 mutants

44

45 **Introduction**

46 In all living organisms, DNA polymerases (DNA Pols) play a central role in the
 47 regulation of DNA replication and repair to ensure faithful transmission of genetic
 48 information. Eukaryotic DNA Pols have been arranged into five well characterized
 49 families namely A, B, C, D and X [1–3]; of these, B, C and D family members are the
 50 main regulators of nuclear DNA replication [2]. In most eukaryotes, the members of

51 the family B DNA Pols – α , δ and ϵ are the main replicative DNA Pols; knock out
52 mutants of the main catalytic subunit of DNA polymerase ϵ (POLE), for example, *tilted*
53 (*til*) 1-1 to -3 in Arabidopsis are embryo lethal [4,5]. The present study focuses on the
54 catalytic subunit of the Arabidopsis POLE complex that has been shown to have diverse
55 regulatory roles, for example in DNA replication and repair, and epigenetic gene
56 silencing.

57
58 The POLE catalytic subunit is known to be involved in both replisome assembly and
59 leading strand synthesis in budding yeast [6–8]. It is characterized by the presence of a
60 catalytically active N-terminal and an inactive C-terminal domain. The active sites for
61 polymerase and exonuclease activities are found within the N-terminal domain, the
62 domain which is highly conserved among other family B polymerases i.e. DNA
63 polymerase α , DNA polymerase β and bacteriophage DNA polymerases (T4 and RB69)
64 [9,10]. Recently it has been found that POLE maintains the functional integrity of the
65 replication fork through phosphorylation of the serine 430 residue on the catalytic
66 subunit [11].

67
68 Although a larger portion of the POLE catalytic subunit accounts for the C-terminal
69 domain [10], with the exception of the DIE motif, this domain lacks catalytic residues
70 for both polymerase and exonuclease activities [12]. Tahirov and co-workers have
71 reported that the presence of catalytically active and inactive domains in POLE is due
72 to a more complex evolutionary scenario than a simple tandem duplication [12]. Based
73 on their findings, one possible hypothesis for the existence of this large catalytically
74 inactive domain would be the presence of at least two family B DNA polymerases and
75 a DNA polymerase δ in the archaeal ancestor of the family B DNA polymerases.

However, during evolution, selection pressure may have favored the existence of the catalytically inactive large C-terminal domain of POLE as a means of ensuring the structural integrity of the POLE complex [12].

In *Arabidopsis thaliana*, the catalytic subunit of POLE is encoded by two genes [4,13], *AtPOL2A* (also known as *POLE1*, *TIL1*, *ABSCISIC ACID OVERSENSITIVE 4 (ABO4)*, or *EARLY IN SHORT DAYS 7 (ESD7)*; TAIR locus id: AT1G08260) and *AtPOL2B* (also called *POLE1B*; 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 [4,13–15]. 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 [4]. 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 [13]. Over the past, four hypomorphic alleles of *AtPOL2A* (designated as *till-4*, *abo 4-1* and *abo4-2*, and *esd7-1*) have been identified and characterized. The *till-4* mutant lines, which carry two G to A mutations in exon 12 and intron 14 have exhibited longer cell cycles during embryo development, delayed development and larger cells [4]; both *abo4-1*, which has a point mutation that substitutes Glycine to Arginine (at the 534 residue) and *abo4-2*, which harbors a T-DNA insert in the 12th exon, have shown sensitivity to abscisic acid (ABA), enhanced

homologous recombination, sensitivity to DNA damage agents and constitutive expression of DNA repair genes [14]; the *esd7-1*, which has a Glycine to Arginine in amino acid position 992, has displayed early flowering and altered growth [15]. Although the molecular function of *AtPOL2B* is not well documented, 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 [4] and vegetative to floral transition [15].

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. 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* genes on whole-transcription to gain insight into potential downstream target genes of both *AtPOL2A* and *B* genes and their interactions at a molecular level.

Results

Molecular confirmation of *atpol2* mutants

The homozygous/heterozygous nature of T-DNA mutant lines and the orientation of T-DNA insert(s) were confirmed through a PCR-based screening method. We found that the homozygous, viable *atpol2a-1* line harbours at least two T-DNA inserts in opposite orientations within the *AtPOL2A* locus (S1 Fig). It was also found that one T-DNA insert is positioned within chromosome 1, after the 2,594,915th base (5'—[left border (LB) — right border (RB)]—3' orientation) and the second insert is positioned just

before the 2,594,924th base in the 5'—[RB — LB]—3' orientation. Although the Salk Institute Genomic Analysis Laboratory (SIGnAL) T-DNA express: *Arabidopsis* gene mapping tool (<http://signal.salk.edu>) supports the presence of an insert at the 12th intron in 5'—[LB — RB]—3' orientation in the *atpol2a-1* mutant, the additional insert found within intron 12 of the *AtPOL2A* has not been reported previously.

Similarly, the individuals of each mutant line of *AtPOL2B* used in the present study were (*atpool2b-1* to -3) confirmed by PCR amplification (S2 Fig). Consistent with the mutant information provided in the *Arabidopsis* genome browser and SIGnAL T-DNA express: *Arabidopsis* gene mapping tools, we found that *atpol2b-1* homozygous mutant line harbours a T-DNA insert in the second intron (307 bp from the transcription start site (TSS)) in the 5'—[RB — LB]—3' orientation. Furthermore, in the *atpol2b-3* line, the insertion was in exon 34 (8930 bp from the TSS). However, the *atpol2b-2* line exhibited the presence of at least two T-DNA inserts within *AtPOL2B*; one insert was oriented in the 5'—[LB — RB]—3' direction (4,408 bp from the TSS) and the second insert was detected after 4,429 bp from the TSS in a 5'—[RB — LB]—3' orientation within intron 15.

Morphology of *atpol2* mutant lines

The homozygous *atpol2a-1* line showed an apparent mutant phenotype as compared to WT plants. These mutants were early flowering when grown under a 16 h photoperiod (Fig 1A). Additionally, these individuals produced considerably smaller siliques (0.85 ± 0.02 cm; Fig 1B) with a reduced number of seeds (16 ± 1.63 seeds/ silique) as compared to WT plants (58 ± 1.45 seeds/ silique; (Fig 1C)). However, none of the

atpol2b mutants examined here showed any visible phenotypic variation as compared to WT plants, under our experimental conditions (Fig 1A-B).

Fig 1. Morphological characters of WT and homozygous *atpol2* mutant plants when grown under a 16h photoperiod.

(A) 26 d old plants. (B) mature siliques. (C) 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

In the present study, we used the Illumina platform to sequence cDNA libraries constructed from leaf tissues of WT and homozygous insertional lines of *AtPOL2* genes. A detailed analysis of transcriptome data derived from this study is given in the latter sections. Only the expression levels of the *AtPOL2A* and *B* genes in their respective mutant lines are considered here.

Visualization of the reads mapped to the *AtPOL2A* locus in the *atpol2a-1* derived transcriptome clearly showed increased transcription from the regions up and downstream of the T-DNA insertion site as compared to the WT. Furthermore, analysis of the reads mapped to the *AtPOL2B* locus in transcriptomes derived from *atpol2b-1* to *-3* mutant lines showed that more reads are mapped to regions downstream of the insertion; there were none mapped upstream of the insertion site.

To further confirm the expression patterns detected for *AtPOL2A* and *B* in their 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 wild-type *AtPOL2A* transcripts from regions both up and downstream of the T-DNA insertion site (Fig 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 (Fig 2B).

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, *atpol2b-1* to -3.

Transcriptome analysis of homozygous *atpol2* mutants

Genome-wide transcriptome analysis using RNA-sequencing (RNA-Seq) was performed on homozygous *atpol2a-1* and *atpol2b-1* to -3 mutant lines to identify potential downstream targets of *AtPOL2* genes. 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. Approximately, 98% of sequence reads were aligned to the *A. thaliana* reference genome (TAIR 10).

In total, the present study detected transcript abundance of 27,409 annotated transcripts in at least one of the sequencing libraries examined; the majority of transcripts were

protein-coding. In addition, a considerable fraction of transposable elements (TEs) was transcriptionally active in the *atpol2a-1* mutant (Fig 3).

Fig 3. Different gene types detected in each sequencing library.

Only the results obtained through the Cufflinks package are shown here.

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). As expected there was a variation in the DEGs detected through the three computational methods (Fig 4); DEGs commonly identified through all three analysis tools (Cuffdiff, DESeq and edgeR) 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 (S1 Table). Further comparison of the DEGs detected in *atpol2a-1* in the present study and DEGs reported previously for *AtPOL2A* mutants [16,17] showed that at least 11 genes are commonly differentially expressed in all *AtPOL2A* gene mutants (Fig 5; Table 1).

Fig 4. Volcano plots showing the differential gene expression. DEGs detected in *atpol2a-1* mutants as compared to WT are shown here. Genes with an adjusted *p-value* of ≤ 0.05 and $-1 \geq \log_2 \text{fold-change} \geq 1$ are shown as red dots.

Fig 5. Venn diagram depicting the DEGs detected in different *AtPOL2A* mutants

222 **Table 1. List of genes differentially expressed in *AtPOL2A* gene mutants (*atpol2a-***
 223 ***1* and *abo4*).**

Gene ID	Gene Name	Description
AT1G02205	<i>ECERIFERUM</i> (<i>CER1</i>)	Expression of the <i>CER1</i> gene associated with the production of stem epicuticular wax and pollen fertility. CER1 protein is an aldehyde decarbonylase, but the exact molecular function of this protein remains to be determined.
AT2G18193		P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT3G27060	<i>TSO2</i>	<i>TSO2</i> transcription occurs predominantly at the S-phase of the cell cycle and its expression pattern is consistent with its role in dNDP biosynthesis during DNA replication in actively dividing cells. Critical for cell cycle progression, DNA damage repair and plant development.
AT3G27630	<i>SIAMESE-RELATED 7</i> (<i>SMR7</i>)	SMR7 is a member of the SIAMESE-RELATED Cyclin-dependent kinase inhibitor family. It is induced by ROS/oxidative stress.

AT4G02390	<i>POLY (ADP-RIBOSE) POLYMERASE 2 (PARP2)</i>	Encodes a DNA-dependent nuclear poly (ADP-ribose) polymerase, thought to be involved in post-translational modification.
AT4G19130	<i>REPLICATION PROTEIN A 1E (RPA1E)</i>	Replication factor-A protein 1-like protein
AT4G22960		FAM63A-like protein (DUF544)
AT4G34510	<i>3-KETOACYL-CO A SYNTHASE 1 (KCS17)</i>	A member of the 3-ketoacyl-CoA synthase family involved in the biosynthesis of very long chain fatty acids
AT5G04150	<i>BHLH101</i>	A member of the basic helix-loop-helix transcription factor family protein
AT5G24280	<i>GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1 (GMII)</i>	Structural-maintenance-of- chromosomes-hinge domain-containing protein. Involved in somatic homologous recombination
AT5G60250		Zinc finger (C3HC4-type RING finger) family protein

224

225 Gene Ontology (GO) enrichment analysis of the 198 DEGs detected in *atpol2a-1*
 226 showed significant enrichment (P-value ≤ 0.05) for GO terms such as response to
 227 defense, oxidation-reduction process, response to UV-A, response to light stimulus,
 228 response to water deprivation, response to ABA and response to jasmonic acid; the list
 229 of enriched GO terms is shown in Fig 6. Analysis of the promoter sequences (1000 bp

upstream genomic region from the TSS) 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 possibly function 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 SUPPRESSOR OF OVEREXPRESSION OF CO1 (SOC1)/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).

Fig 6. GO enrichment analysis of DEGs detected in *atpol2a-1*.

Only the GO terms related to biological processes at $P\text{-value} \leq 0.05$ are shown here.

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

Motif sequence and logo	Possible biological role(s) (enriched GO terms at P-value ≤ 0.05)
Motif 1	Transmembrane receptor protein
Consensus sequence:	tyrosine kinase signaling pathway
WVAAAAARAAAAARAAAAAA	Regulation of transcription, DNA-dependent
Circadian rhythm
	Response to jasmonic acid stimulus
	Cytoskeleton organization
	Cytokinin mediated signaling pathway
	Protein amino acid phosphorylation
Motif 2	Circadian rhythm
Consensus sequence:	Regulation of transcription
MYTTTTTTTTTTTTTT

Motif 3	Regulation of transcription
Consensus sequence:	Transmembrane receptor protein
WCMYBYYCTCTYCTYTCYYY	tyrosine kinase signaling pathway
Protein amino acid phosphorylation
	Plastid organization
	Megagametogenesis
	Leaf development
	Root development
	Abaxial cell fate specification
Motif 4	Photosynthesis
Consensus sequence:	Response to ABA stimulus
GATKGACACGTGG	Response to far-red light
Response to wounding
	Response to water deprivation
	Response to blue light
	Circadian rhythm
Motif 5	Mitochondrial transport
Consensus sequence:	
KRYGRADVRRKTVRHVGTGGTGGK	
KGGS

Comparison of DEGs detected in the three *atpol2b* mutants (*atpol2b-1* to -3) showed a substantial difference among the mutants (Fig 7); only 12 genes were commonly

differentially expressed in all three mutants and their descriptions are given in Table 3.

Genes that were differentially expressed in at least two of the *atpol2b* studied here were

considered as DEGs in *AtPOL2B* mutants and used in downstream analyses.

Fig 7. Venn diagram depicting the DEGs detected in *atpol2b1 – 3*

Table 3. List of DEGs commonly detected in *atpol2b* mutants. Bold gene ids designate genes that showed differential expression in both *atpol2a* and *b* mutants.

Gene ID	Gene Name	Description
AT1G54010	GDSL-like lipase 23 (GLL23)	Acylhydrolase superfamily protein
AT1G73260	<i>KUNITZ TRYPSIN INHIBITOR 1 (KTII)</i>	Encodes a trypsin inhibitor involved in modulating programmed cell death in plant- pathogen interactions.
AT2G10940		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT2G14247		
AT2G14610	<i>PATHOGENESIS- RELATED GENE 1 (PRI)</i>	<i>PRI</i> expression is induced in response to a variety of pathogens. It is a useful molecular marker for systemic acquired resistance (SAR) response. Expression is salicylic-acid responsive.

AT2G26010	<i>PLANT DEFENSIN 1.3 (PDF1.3)</i>	Predicted to encode a pathogenesis-related protein. Belongs to the plant defensin family.
AT3G44300	<i>NITRILASE 2 (NIT2)</i>	Encodes an enzyme that catalyzes the hydrolysis of indole-3-acetonitrile (IAN) to indole-3-acetic acid (IAA). Mutants have reduced sensitivity to IAN and are sensitive to IAA. This enzyme likely participates in other non-auxin-related metabolic pathways.
AT4G11650	<i>OSMOTIN 34 (OSM34)</i>	Osmotin-like protein
AT4G19810	<i>Class V chitinase (ChiC)</i>	Glycosyl hydrolase family protein with chitinase insertion domain-containing protein. ChiC encodes a Class V chitinase that is a part of glycoside hydrolase family. ChiC transcript levels are increased in response to ABA, jasmonic acid and osmotic stress.
AT5G48490	<i>DIR1-LIKE</i>	Encodes a protein with similarity to a lipid transfer protein that may contribute to systemic acquired SAR; bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein.
AT5G65080	<i>MADS AFFECTING</i>	MADS-box transcription factor family protein; encodes MADS-domain protein. The gene is

FLOWERING 5 upregulated during vernalization and regulates
(*MAF5*) flowering time.

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 (Fig 8).

Fig 8. GO enrichment analysis of DEGs detected in at least two *atpol2b* mutants.

Only the GO terms related to biological processes at P-value ≤ 0.05 are shown here.

Analysis of the promoter sequences (1000 bp upstream from the TSS) of DEGs identified in at least two *atpol2b* mutants revealed enrichment of five motifs (Table 4). The majority of DEGs contained motifs I (68/72 genes) and IV (66/72 genes) in their promoter regions that may interact with MADS-box transcription factor SOC1/ AGL20.

Table 4. 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	Regulation of transcription	AT2G45660
Consensus sequence:	Response to jasmonic acid stimulus	
RWAAAAAAAAAARA	Cytoskeleton organization	
AAAAAVA

Motif II

Translation

Consensus sequence:

GTGSTGGMDAAACCR

KCCABC

Motif III

Consensus sequence:

AYBSGWGGGAAATRG

RCTCHNGCWTGTGYC

RANCHMWDYSC

Motif IV

Transmembrane receptor protein:

AT2G45660

Consensus sequence:

tyrosine kinase signaling pathway

GAAKDRVAMRAVRA

Protein amino acid phosphorylation

AGARARARADANRRD

Regulation of transcription

Leaf morphogenesis

Flower development

Protein-protein interaction (PPI) analysis

Interestingly, the PPI network constructed for the DEGs identified in *atpol2a-1* using the StringDB showed that genes encoding REPLICATION PROTEIN A 1E (RPA1E; AT4G19130), TSO2 (AT3G27060) and AT3G49160 proteins may interact with AtPOL2A/TIL1 (Fig 9). Further, it was noted that RPA1E interacts with GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1 (GIM1, AT5G24280), and TSO2 interacts with POLY (ADP-RIBOSE) POLYMERASE 2 (PARP2; AT4G02390), and AT3G49160 interacts with FERRETIN 1 (FER1; AT5G01600).

Fig 9. Protein-protein interaction network of DEGs detected in *atpol2a-1*.

This network was constructed using StringDB. Network nodes represent proteins (colored nodes represent query proteins; empty nodes: proteins of unknown structures; filled nodes: some 3D structure is known or predicted.). Edges represent protein-protein interactions.

PPI network analysis of the DEGs detected in *atpol2b* mutants also showed known and predicted interactions among genes (Fig 10). Interestingly, of the 12 genes commonly

differentially expressed in all three *atpol2b* mutants (see Table 3), genes encoding six proteins (KTI 1 (AT1G73260), PR1 (AT2G14610), PDF1.3 (AT2G26010), NIT2 (AT3G44300), OSM34 (AT4G11650) and ChiC (AT4G19810)) showed known/predicted interactions among them (Fig 10). However, none of the proteins were interacting with AtPOL2B/TIL2.

Fig 10. Protein-protein interaction network of DEGs detected in at least two *atpol2b* mutants.

This network was constructed using StringDB. Network nodes represent proteins (colored nodes represent query proteins; empty nodes: proteins of unknown structures; filled nodes: some 3D structure is known or predicted.). Edges represent protein-protein interactions.

Docking sites of potential targets 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 (Fig 11). Docking scores obtained for these proteins are given in Table 5; 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 DEGs detected in the present study.

Table 5. ClusPro docking scores for AtPOL2s and their known interactors.

Gene identifier	Protein name	Subcellular localization	Complex with AtPOL2A		Complex with AtPOL2B	
			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

Fig 11. Protein-protein interactions of AtPOL2 proteins extracted from StringDB.

The common interactors of AtPOL2A and AtPOL2B proteins are highlighted in blue.

(A) Interactors of AtPOL2A (TIL1). (B) Interactors of AtPOL2B (TIL2).

Structural modeling of 22 DEGs (Table 1 and 3) using SWISS-MODEL predicted the 3D structures of 20 proteins, while two genes (AT3G27630 and AT2G14247) failed to locate homologous protein sequences for modeling. Out of the structures predicted, 90% (18) protein structures showed over 85% residues clustered in favoured 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 6). Of these, three (CER1, RPA1E and AT5G60250) and two proteins (PR1 and AT5G48490) indicated notably significant interactions (*cluster size* > 60 and *balanced score* < -900) with AtPOL2A and AtPOL2B proteins, respectively (Fig 12). 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* (S2 Table).

Table 6. ClusPro docking scores for AtPOL2s and their potential targets.

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 decarboxylase CER1	Endoplasmic reticulum membrane	117	-1101.1
	AT5G60250*	Uncharacterized protein	Cytoplasm	94	-1172.4
	AT5G04150	Transcription factor bHLH101	Nucleus	91	-806

AtPOL2B	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	-	-
	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 (ChiC)	Secreted	72	-757.4

AT3G44300*	Nitrilase 2	Cytoplasm	66	-1009.3
AT1G54010	Inactive GDSL esterase/lipase-like protein 23	Secreted	51	-787.5
AT5G65080	Protein MADS AFFECTING FLOWERING 5	Nucleus	41	-548.1
AT2G14247	Uncharacterized protein	Nucleus	-	-

Fig 12. 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 AtPOL2s**

The first insight into the promoter elements of *AtPOL2* genes came from a study conducted by Ronceret and co-workers in 2005 [9]. 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 *AtPOL2s* at the molecular level. 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 (ATG) were selected and analysed using the

PLACE and PlantCARE database 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. For instance, promoter motifs ARR1AT, DOFCOREZM, GATABOX, GT1CONSENSUS, GTGANTG10, MYCCONSSENSUSAT, POLLEN1LELAT52, and WRKY71OS were most abundant in both gene promoters (Fig 13; descriptions of these CREs are included in S3A Table). In addition, CGCGBOXAT element (also known as ‘CGCG box’) involved in Ca²⁺/calmodulin regulation, IBOXCORE element conserved in upstream of light-regulated genes, and MYBCORE element (consensus MYB recognition sequence) were overrepresented in the *AtPOL2A* promoter whereas CURECORECR (a copper-response element), INRNTPSADB (light-responsive initiator element) and RAV1AAT (recognition sequence of DNA-binding protein RAV1) were overrepresented in the *AtPOL2B* promoter.

Furthermore, 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* gene in its promoter. In addition, several homeodomain-leucine zipper (HD-Zip) protein binding sites (ATHB2ATCONSENSUS and ATHB5ATCORE) and MYB-related binding sites (i.e. BOXLCOREDPCAL, CCA1ATLHCB1, IBOX, MYB2AT, MYB2CONSENSUSAT, MYBPLANT, MYBPZM) were detected.

Interestingly, a hexamer motif found in the Arabidopsis histone H4 gene promoter (HEXAMERATH4) was also detected in the *AtPOL2A* promoter. By contact, QELEMENTZM13 (Quantitative elements (Q-element)), which is known to enhance activity in pollen-specific genes; AUXRETGA1GMGH3, an auxin-responsive CRE; LTREIHVBLT49, a low-temperature responsive element; PREATPROD, 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; WBOXPCWRKY1 and WBOXNTCHN48 (pathogen/elicitor responsive elements) were detected only in the *AtPOL2B* promoter. Important CREs detected only in one of the *AtPOL2s* gene promoters and their functions are listed in S3B and S3C Tables.

Fig 13. Frequencies of important CREs detected in promoter regions of both *AtPOL2* genes (2500 bp upstream from the translation start site)

Fig 14. 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.

Analysis of the promoter sequences using PlantCARE also 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 (Fig 14A). 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 (Fig 14B). 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 (Fig 14C).

Discussion

DNA Pols play a central role in maintaining genomic integrity. In the present study, we investigated the potential indirect targets of the two isoforms of the catalytic subunit of Arabidopsis POLE, AtPOL2A and AtPOL2B at the transcriptomic level. We further identified the putative CREs in the promoter regions of *AtPOL2* genes to better understand the potential targets of *AtPOL2* genes.

Transcriptome analysis of *AtPOL2* gene mutants

The publicly available T-DNA insertional mutant line collections have played an important role in functional studies to explore gene functions and gene/protein regulatory networks. In the present study, we employed four homozygous T-DNA insertional lines. All of these mutants, except *atpol2b-3*, contained insertions in introns. However, all mutants showed a change in the expression pattern of the mutated gene as compared to that of in the WT; although Ronceret *et al.* (2005) report no expression of *AtPOL2B* in *atpol2b* mutants (SALK_001413 and SALK_056503), we detected increased transcript accumulation downstream of the insertion site in all *atpol2b* mutants. Detection of transcripts both upstream and downstream of the T-DNA

insertion site is supported by previous literature [18]. Such fusion/truncated transcripts formed due to activation of T-DNA promoter or termination of transcription due to inserted T-DNA fragment could translate either into a functional or a truncated protein [19]. A closer look at the primer sequences used by Ronceret and co-workers [13] to amplify *AtPOL2B* transcripts showed that they have been designed from regions upstream of the T-DNA insertion site. This might be the most likely reason for the failure to detect *AtPOL2B* expression in *atpol2b* mutants in the previous literature.

Consistent with previous findings [14], the *AtPOL2A* mutant examined in the present study (*atpol2a-1*) was also early flowering and exhibited smaller siliques than WT/*AtPOL2B* mutants. This observation, together with the expression pattern of the mutated gene in the T-DNA insertional line, strongly suggests *atpol2a-1* is a weak allele of *AtPOL2A*. Moreover, as reported in previous studies [4,13,15], 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 subsequent transcriptomic analysis to explore potential downstream targets of *AtPOL2B*.

Several different bioinformatics tools have been developed and used in the analysis of expression data generated through next-generation technologies. However, having multiple tools for analysis has made the analysis and interpretation of data challenging for researchers and has led to large variations in outputs [20,21]. Hence, in the present study, two computational tools for transcript assembly (Cufflinks and HTSeq) and three tools for differential gene expression analysis (cuffdiff, edgeR and DESeq) were used; only DEGs resulted from all the tools were considered for the downstream analysis.

467 This resulted in a total of 198 genes in *atpol2a-1*. Han *et al.* (2015) [17] and Pedroza-
468 Garcia *et al.* (2017) [16] report 749 ($-1 \geq \text{Log}_2(\text{FC}) > 1$; $P < 0.01$) and 372 ($-1 \geq \text{Log}_2$
469 $(\text{FC}) \geq 1.5$; $P \leq 0.01$) DEGs in a single-gene mutant of *AtPOL2A*, respectively. The
470 discrepancies observed in DEGs between mutants of *AtPOL2A* are likely to result from
471 the differences in the plant material and growth conditions, sequencing platforms used
472 to generate transcriptomic data, sequencing depth, number of replicates and analysis
473 methods. Despite discrepancies, we identified 11 DEGs in all *AtPOL2A*-deficient
474 mutants. This included several genes involved in the DNA damage and repair response
475 pathways such as *TSO2*, *SMR7*, *RPA1E*, *PARP1* and *GMI1*. Further analysis of these
476 DEGs through PPI network analysis showed that *TSO2*, *RPA1E* and AT3G49160 (a
477 pyruvate kinase-related protein) [22] may associate with *AtPOL2A*; this was further
478 supported by the protein-protein docking using Cluspro. In Arabidopsis, *TSO2* is one
479 of the three members of the ribonucleotide reductase (RNR) small subunit genes, which
480 are transcribed predominantly at the S-phase of the cell cycle [23]. Roa *et al.* (2009)
481 [24] report E2F target genes including *AtPOL2A* and *TSO2* coexpress in the DNA repair
482 network. The *tso2* mutants have also shown increased DNA damage, the release of
483 transcriptional gene silencing and sensitivity to UV-C light as observed in *atpol2a*
484 mutants [23]. Moreover, *tso2-1 rnr21-1* double mutants have shown induced expression
485 of several molecular markers associated with DNA damage and repair i.e. *PARP1*,
486 *PARP2* and *RAD51* [23]. Recently, upregulation of *PARP2*, *RPA1E* and *POL2A* have
487 been observed in loss-of-function mutants of the F-box protein *F-BOX-LIKE17*
488 (*FBL17*); in Arabidopsis, *FBL17* functions in cell-cycle regulation [25], and is required
489 for cell division during pollen development. Our PPI analysis also showed that
490 uncharacterized AT3G49160 interacts with FER1; both of these genes are known to be
491 expressed in response to hydrogen peroxide [26–28].

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 [14,17]. TEs found in the pericentromeric heterochromatin are heavily methylated. 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 [29]. Consistent with previous studies, we also observed that the silencing of the majority of pericentromeric TEs are released in *atpol2a-1* mutant.

Notably, the analysis of the expression profiles of three *atpol2b* mutants showed that the DEGs in mutants are random; remarkably, there was little overlap between the *atpol2b* mutant-derived DEGs. It is reported that POLE may be involved in the repair of replication errors [16]; the *POLE* mutants might not be able to repair errors during DNA replication and such replication errors can occur randomly. Therefore, in different pole mutants, the DEGs could be random.

Interestingly, our transcriptome analysis showed that *AtPOL2B* is more likely to interact with stress-responsive genes; several genes related to biotic stress responses were differentially expressed in all three *atpol2b* mutants examined i.e. *GLL23*, *KTII*, *PR1* (a salicylic acid marker gene) [30,31], *PDF1.3* (a plant defensin) [32], *OSM34*, *DIR1-LIKE* (may contribute to systemic acquired resistance) [33] 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 [34]. Moreover, *KTII* (serine protease inhibitor) is known to be involved in the regulation of programmed cell death during biotic stress tolerance in Arabidopsis [35,36]. In addition to defense-related processes, *OSM34* has shown to be effective against abiotic stresses [37–39]. 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 [40]. 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 [41]. Interestingly, protein-protein docking of these DEGs confirmed significant interaction between AtPOL2B and NIT2, OSM34 and Chic proteins.

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); no significant alteration in floral development genes including *MAF5* has been observed in *AtPOL2A* mutant *esd7-1* [15]. However, Han *et al.* (2015) also report differential expression of *MAF* in the *AtPOL2A* mutant, *abo4-1* [17]. 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 *FLC*, *MAF1*, *MAF4* and *MAF5* [42,43]. These observations suggest *AtPOL2A* is likely to regulate the methylation status of *MAF5* and thereby control the expression of *MAF5*.

Computational analysis of *AtPOL2* promoter regions

As reported in Ronceret *et al.*, 2005 [13], 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 [44]. 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 [45]. 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 [46]. 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 [47].

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 [48], and HEXMOTIFTAH3H4, a hexamer motif (ACGTCA) found in H3 and H4 histone gene promoters [49] were found in *AtPOL2A* and *AtPOL2B* promoters, respectively. The

hexamer motif CCGTCG found in the wheat histone genes has been 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 [48]. Moreover, the conserved hexamer motif ACGTCA found in the plant histone H3 and H4 genes binds specifically with the nuclear protein HBP-1 (histone gene-binding protein-1) [50–53], which is more abundant in actively proliferating cells; it is reported that HBP-1 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) TFs, which act as transcriptional regulators involved in many plant-specific biological processes [54–56]. Engel *et al.* (2005) [57] 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 [58].

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 [59]. 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 [60]. 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 [61]. Moreover, two CWWWWWWWG 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 [62]. Further, a TGTCACACMCUCUMISIN (TGTCACA) was found in the promoter region of *AtPOL2A*; this enhancer element is responsible for the fruit-specific expression of cucumisin gene in melon (*Cucumis melo* L.) [63]. 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 [64]. 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 [64]. 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 HOMOMEBOX 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 [65,66], whereas *AtHB5* appears to be involved in ABA-related responses and/or water deficit conditions [67]. In addition, several MYB protein recognition sequences were found in the promoter region of *AtPOL2A* (BOXLCOREDPCAL, CCA1ATLHCB1 [68], MYB2CONSENSUSAT, MYBPLANT (a binding site for a flower-specific MYB protein [69]), MYBPZM, MYB2AT and IBOX [70]). In plants, MYB transcription factors are involved in regulating many biological processes including hormonal signal transduction and abiotic/biotic stress tolerance [71,72], and more importantly, these TFs are involved in the regulation of the cell cycle at the G2/M transition [73].

We also noted that several light-responsive elements (GATABOX) and stress responsive elements (MYCCONSENSUSAT, ARR-1 binding elements [74,75], WRKY71OS [76] 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, PREATPROD, 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 *AtPOL2B* is more likely to play a role in modulating stress responses in Arabidopsis.

Conclusion

This study, based on several *AtPOL2* mutants, is the first to present a detailed comparison of the transcriptomes of *AtPOL2A* and *AtPOL2B* mutants on a global scale. Most importantly, the findings of the present study begin to unravel the potential targets of *AtPOL2* genes at the molecular level and suggest new avenues for future studies.

Materials and 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°C, 60% relative humidity and 16 h photoperiod (Fitotron plant growth chambers, Weiss Gallenkamp, UK).

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.

Molecular screening of T-DNA mutants

Genomic DNA was extracted from leaves of mutants and WT plants using the DNeasy® Plant Mini Kit (Qiagen, UK) according to the manufacturer's instructions.

Homozygous mutant lines were identified through a PCR-based screening method using the primers designed through the SIGnAL iSct tool (<http://signal.salk.edu/tdnaprimers.2.html>). The list of primers used to confirm T-DNA mutant lines used in the present study is given in S4 Table.

In brief, two main PCR reactions with primer combinations, (1) LB and RB genomic primers; (2) LB/RB primer and a T-DNA specific primer (LBb1.3 or LBa1) were set up to confirm mutant lines. Amplifications were performed in a volume of 25.0 µl, containing 12.5 µl of 2X BioMix PCR master mix (Bioline, UK), 0.75 µl of 0.3 µM forward and reverse primer (Invitrogen, UK) and 1.0 µl of genomic DNA. Each amplification was carried out in a GeneAmp PCR system 9700 (Applied Biosystems) using the following programme: 2 min at 94°C; followed by 30 cycles of 30 s at 94°C, 30 s at 55°C, and 1 min at 72°C; and finally 10 min at 72°C for the final elongation. All the amplified products were confirmed by sequencing at a commercial company (Source Bioscience, UK).

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 quantified through Agilent BioAnalyzer 2100. cDNA libraries were constructed and amplified using random primers according to the manufacturer's instructions (Illumina TruSeq RNA sample preparation guide v2 for Illumina paired-end multiplexed sequencing). 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).

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/>). After clearing for adapter sequence contaminations, TopHat 2.1.1 [77] (<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 [78]. The Cufflinks (2.1.1) was used to assemble reads that were mapped and to normalize transcript levels [79] by means of Fragments Per Kilobase of exon per Million fragments mapped (FPKM) [79]; in addition, HTSeq-count 0.11.1 [80] (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) [81], DESeq [82] (<https://bioconductor.org/packages/release/bioc/html/DESeq.html>) and edgeR [83] (<https://bioconductor.org/packages/release/bioc/html/edgeR.html>) were used for the differential gene expression analysis; of these DESeq and edgeR can run on the dataset without replicates, and cuffdiff can be used to compare the log-ratio of gene expression

(FPKM) between mutants. 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 [84]. GO terms and KEGG pathways were considered enriched if the associated Benjamini-Hochberg adjusted p-value was less than 0.05.

The promoter sequences (1000 bp upstream of TSS) of each DEG 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 (Gene Ontology for Motifs) tool (<https://meme-suite.org/meme/tools/gomo>).

Interactome analysis and docking sites of potential targets 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/>). All proteins encoding genes that were differentially expressed in AtPOL2A/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 (Table 1 and 3) 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

[85]. In addition, the subcellular localization of the proteins was explored using the LocTree3 tool (<https://roslab.org/services/loctree3/>). The ClusPro server 2.0 (<https://cluspro.bu.edu/>), which employs a rigid body protein-protein docking [86], 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. First-strand cDNA was synthesized using SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, UK) using 500 ng of total RNA.

Gene sequences of *AtPOL2A* (*AT1G08260*) and *AtPOL2B* (*AT2G27120*) and *Arabidopsis UBC21* (*AT5G25760*) were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov>). Primers were designed using the NCBI-Primer BLAST tool (<http://www.ncbi.nlm.nih.gov/tools/primer-blast>). One primer of each pair was designed, where possible, to span an exon-exon junction (see S5 Table).

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 *UBC* reference gene. Data were obtained and analyzed using Rotor-Gene 6000 series software Version 1.7.

In silico* promoter analysis of *AtPOL2s

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 *cis*-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.

788 **Abbreviations**

789	ABA	Abscisic acid
790	ABO	ABSCISIC ACID OVERSENSITIVE
791	ChiC	Class V chitinase
792	CRE	<i>Cis</i> -acting regulatory element
793	DEG	Differentially expressed gene
794	DNA Pols	DNA polymerases
795	ESD	EARLY IN SHORT DAYS
796	GO	Gene Ontology
797	HBP-1	Histone gene-binding protein-1
798	LB	Left border
799	NIT2	NITRILASE 2
800	POLE	DNA polymerase ϵ
801	PPI	Protein-protein interaction
802	PR1	PATHOGENESIS-RELATED GENE 1
803	Q-element	Quantitative element
804	RB	Right border
805	RPA1E	REPLICATION PROTEIN A 1E
806	TE	Transposable element
807	TIL	TILTED
808	TSS	Transcription start site
809	WT	Wild type

810

811 **Acknowledgments**

812 Not applicable.

References

1. Burgers PM. Eukaryotic DNA polymerases in DNA replication and DNA repair. *Chromosoma*. 1998;107:218–27.
2. Filee J, Forterre P, Sen-Lin T, Laurent J. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J Mol Evol*. 2002;54:763–73.
3. Sanchez Mde L, Costas C, Sequeira-Mendes J, Gutierrez C. Regulating DNA replication in plants. *Cold Spring Harb Perspect Biol* [Internet]. 2012;4. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23209151>
4. Jenik PD, Jurkuta RE, Barton MK. Interactions between the cell cycle and embryonic patterning in *Arabidopsis* uncovered by a mutation in DNA polymerase epsilon. *Plant Cell*. 2005;17:3362–77.
5. Pedroza-Garcia J-A, De Veylder L, Raynaud C. Plant DNA polymerases. *Int J Mol Sci*. 2019;20:4814.
6. Meng X, Wei L, Devbhandari S, Zhang T, Xiang J, Remus D, et al. DNA polymerase ϵ relies on a unique domain for efficient replisome assembly and strand synthesis. *Nat Commun*. 2020;11:2437.
7. Pursell ZF, Isoz I, Lundstrom E-B, Johansson E, Kunkel TA. Yeast DNA polymerase participates in leading-strand DNA replication. *Science*. 2007;317:127–30.
8. Shinbrot E, Henninger EE, Weinhold N, Covington KR, Göksenin AY, Schultz N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication

- 835 strand specific mutation patterns and human origins of replication. *Genome Res.*
836 2014;24:1740–50.
- 837 9. Hogg M, Johansson E. DNA polymerase epsilon. In: MacNeill S. (eds) *The*
838 *Eukaryotic Replisome: a Guide to Protein Structure and Function. Subcellular*
839 *Biochemistry*, vol 62. Springer, Dordrecht. [https://doi.org/10.1007/978-94-007-](https://doi.org/10.1007/978-94-007-4572-8_13)
840 4572-8_13
- 841 10. Pospiech H, Syvaaja JE. DNA polymerase epsilon - more than a polymerase. *Sci*
842 *World J.* 2003;3:87–104.
- 843 11. Pellicanò G, Al Mamun M, Jurado-Santiago D, Villa-Hernández S, Yin X,
844 Giannattasio M, et al. Checkpoint-mediated DNA polymerase ϵ exonuclease
845 activity curbing counteracts resection-driven fork collapse. *Mol Cell.*
846 2021;S1097276521003117.
- 847 12. Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin EV. Evolution of
848 DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon
849 and a chimeric origin of eukaryotic polymerases from two classes of archaeal
850 ancestors. *Biol Direct.* 2009;4:11.
- 851 13. Ronceret A, Guillemot J, Lincker F, Gadea-Vacas J, Delorme V, Bechtold N, et
852 al. Genetic analysis of two *Arabidopsis* DNA polymerase epsilon subunits during
853 early embryogenesis. *Plant J.* 2005;44:223–36.
- 854 14. Yin H, Zhang X, Liu J, Wang Y, He J, Yang T, et al. Epigenetic regulation,
855 somatic homologous recombination, and abscisic acid signaling are influenced by
856 DNA polymerase epsilon mutation in *Arabidopsis*. *Plant Cell.* 2009;21:386–402.

- 857 15. del Olmo I, Lopez-Gonzalez L, Martin-Trillo MM, Martinez-Zapater JM, Pineiro
858 M, Jarillo JA. *EARLY IN SHORT DAYS 7 (ESD7)* encodes the catalytic subunit of
859 DNA polymerase epsilon and is required for flowering repression through a
860 mechanism involving epigenetic gene silencing. *Plant J.* 2010;61:623–36.
- 861 16. Pedroza-Garcia JA, Mazubert C, Del Olmo I, Bourge M, Domenichini S, Bounon
862 R, et al. Function of the plant DNA polymerase epsilon in replicative stress
863 sensing, a genetic analysis. *Plant Physiol.* 2017;173:1735–49.
- 864 17. Han YF, Huang HW, Li L, Cai T, Chen S, He XJ. The cytosolic iron-sulfur cluster
865 assembly protein MMS19 regulates transcriptional gene silencing, DNA repair,
866 and flowering time in *Arabidopsis*. *PLoS One.* 2015;10:e0129137.
- 867 18. Wang YH. How effective is T-DNA insertional mutagenesis in *Arabidopsis*? *J*
868 *Biochem Tech.* 2008;1:11–20.
- 869 19. Ohtomo I, Ueda H, Shimada T, Nishiyama C, Komoto Y, Hara-Nishimura I, et al.
870 Identification of an allele of *VAM3/SYP22* that confers a semi-dwarf phenotype in
871 *Arabidopsis thaliana*. *Plant Cell Physiol.* 2005;46:1358–65.
- 872 20. Salifu SP, Nyarko HN, Doughan A, Msatsi HK, Mensah I, Bukari A-RA. RNA-
873 seq analyses: benchmarking differential expression analyses tools reveals the
874 effect of higher number of replicates on performance. *bioRxiv.*
875 2020;2020.06.10.144063.
- 876 21. Corchete LA, Rojas EA, Alonso-Lopez D, De Las Rivas J, Gutierrez NC,
877 Burguillo FJ. Systematic comparison and assessment of RNA-seq procedures for
878 gene expression quantitative analysis. *Sci Rep.* 2020;10:19737.

- 879 22. Andre C, Froehlich JE, Moll MR, Benning C. A heteromeric plastidic pyruvate
880 kinase complex involved in seed oil biosynthesis in *Arabidopsis*. Plant Cell.
881 2007;19:2006–22.
- 882 23. Wang C, Liu Z. *Arabidopsis* ribonucleotide reductases are critical for cell cycle
883 progression, DNA damage repair, and plant development. Plant Cell.
884 2006;18:350–65.
- 885 24. Roa H, Lang J, Culligan KM, Keller M, Holec S, Cognat V, et al. Ribonucleotide
886 reductase regulation in response to genotoxic stress in *Arabidopsis*. Plant Physiol.
887 2009;151:461–71.
- 888 25. Gentric N, Masoud K, Journot RP, Cognat V, Chaboute ME, Noir S, et al. The F-
889 box-like protein FBL17 is a regulator of DNA-damage response and colocalizes
890 with RETINOBLASTOMA RELATED1 at DNA lesion sites. Plant Physiol.
891 2020;183:1295–305.
- 892 26. Laloi C, Stachowiak M, Pers-Kamczyc E, Warzych E, Murgia I, Apel K. Cross-
893 talk between singlet oxygen- and hydrogen peroxide-dependent signaling of stress
894 responses in *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 2007;104:672–7.
- 895 27. op den Camp RG, Przybyla D, Ochsenbein C, Laloi C, Kim C, Danon A, et al.
896 Rapid induction of distinct stress responses after the release of singlet oxygen in
897 *Arabidopsis*. Plant Cell. 2003;15:2320–32.
- 898 28. Petit J-M, Briat J-F, Lobréaux S. Structure and differential expression of the four
899 members of the *Arabidopsis thaliana* ferritin gene family. Biochem.
900 2001;359:575–82.

- 901 29. Bourguet P, Lopez-Gonzalez L, Gomez-Zambrano A, Pelissier T, Hesketh A,
902 Potok ME, et al. DNA polymerase epsilon is required for heterochromatin
903 maintenance in *Arabidopsis*. *Genome Biol.* 2020;21:283.
- 904 30. Pecenkova T, Pleskot R, Zarsky V. Subcellular localization of *Arabidopsis*
905 pathogenesis-related 1 (PR1) protein. *Int J Mol Sci* [Internet]. 2017;18. Available
906 from: <https://www.ncbi.nlm.nih.gov/pubmed/28406455>
- 907 31. Van Loon LC, Van Strien EA. The families of pathogenesis-related proteins, their
908 activities, and comparative analysis of PR-1 type proteins. *Physiol Mol Plant*
909 *Pathol.* 1999;55:85–97.
- 910 32. Biedrzycki ML, L V, Bais HP. Transcriptome analysis of *Arabidopsis thaliana*
911 plants in response to kin and stranger recognition. *Plant Signal Behav.*
912 2011;6:1515–24.
- 913 33. Champigny MJ, Isaacs M, Carella P, Faubert J, Fobert PR, Cameron RK. Long
914 distance movement of DIR1 and investigation of the role of DIR1-like during
915 systemic acquired resistance in *Arabidopsis*. *Front Plant, Sci* 2013;4:230.
- 916 34. Lai CP, Huang LM, Chen LO, Chan MT, Shaw JF. Genome-wide analysis of
917 GDSL-type esterases/lipases in *Arabidopsis*. *Plant Mol Biol.* 2017;95:181–97.
- 918 35. Arnaiz A, Talavera-Mateo L, Gonzalez-Melendi P, Martinez M, Diaz I,
919 Santamaria ME. *Arabidopsis* kunitz trypsin inhibitors in defense against spider
920 mites. *Front Plant Sci.* 2018;9:986.

- 921 36. Li J, Brader G, Palva ET. Kunitz trypsin inhibitor: an antagonist of cell death
922 triggered by phytopathogens and fumonisin b1 in *Arabidopsis*. Mol Plant.
923 2008;1:482–95.
- 924 37. Capelli N, Diogon T, Greppin H, Simon P. Isolation and characterization of a
925 cDNA clone encoding an osmotin-like protein from *Arabidopsis thaliana*. Gene.
926 1997;191:51–6.
- 927 38. Bashir MA, Silvestri C, Ahmad T, Hafiz IA, Abbasi NA, Manzoor A, et al.
928 Osmotin: a cationic protein leads to improve biotic and abiotic stress tolerance in
929 plants. Plants. 2020;9:992.
- 930 39. Hao J, Wu W, Wang Y, Yang Z, Liu Y, Lv Y, et al. *Arabidopsis thaliana* defense
931 response to the ochratoxin A-producing strain (*Aspergillus ochraceus* 3.4412).
932 Plant Cell Rep. 2015;34:705–19.
- 933 40. Ohnuma T, Numata T, Osawa T, Mizuhara M, Lampela O, Juffer AH, et al. A
934 class V chitinase from *Arabidopsis thaliana*: gene responses, enzymatic
935 properties, and crystallographic analysis. Planta. 2011;234:123–37.
- 936 41. van der Woude L, Piotrowski M, Klaasse G, Paulus JK, Krahn D, Ninck S, et al.
937 The chemical compound “Heatin” stimulates hypocotyl elongation and interferes
938 with the *Arabidopsis* NIT1-subfamily of nitrilases. Plant J [Internet].
939 2021;Available from: <https://www.ncbi.nlm.nih.gov/pubmed/33768644>
- 940 42. Cao Y, Dai Y, Cui S, Ma L. Histone H2B monoubiquitination in the chromatin of
941 *FLOWERING LOCUS C* regulates flowering time in *Arabidopsis*. Plant Cell.
942 2008;20:2586–602.

- 943 43. Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, Steinmetz A, et al. Di-
944 and tri- but not monomethylation on histone H3 lysine 36 marks active
945 transcription of genes involved in flowering time regulation and other processes
946 in *Arabidopsis thaliana*. Mol Cell Biol. 2008;28:1348–60.
- 947 44. de Jager SM, Menges M, Bauer UM, Murra JA. *Arabidopsis* E2F1 binds a
948 sequence present in the promoter of S-phase-regulated gene *AtCDC6* and is a
949 member of a multigene family with differential activities. Plant Mol Biol.
950 2001;47:555–68.
- 951 45. Kosugi S, Ohashi Y. E2F sites that can interact with E2F proteins cloned from
952 rice are required for meristematic tissue-specific expression of rice and tobacco
953 proliferating cell nuclear antigen promoters. Plant J. 2002;29:45–59.
- 954 46. Planchais S, Perennes C, Glab N, Mironov V, Inze D, Bergounioux C.
955 Characterization of *cis*-acting element involved in cell cycle phase-independent
956 activation of *Arabidopsis* *CycB1;1* transcription and identification of putative regulatory
957 proteins. Plant Mol Biol. 2002;50:111–27.
- 958 47. Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Grissem W, et
959 al. Genome-wide identification of potential plant E2F target genes. Plant Physiol.
960 2005;139:316–28.
- 961 48. Chaubet N, Flenet M, Clement B, Brignon P, Gigot C. Identification of *cis*-
962 elements regulating the expression of an *Arabidopsis* histone H4 gene. Plant J.
963 1996;10:425–35.

- 964 49. Mikami K, Tabata T, Kawata T, Nakayama T, Iwabuchi M. Nuclear protein(s)
965 binding to the conserved DNA hexameric sequence postulated to regulate
966 transcription of wheat histone genes. FEBS Lett. 1987;223:273–8.
- 967 50. Mikami K, Sakamoto A, Takase H, Tabata T, Iwabuchi M. Wheat nuclear protein
968 HBP-1 binds to the hexameric sequence in the promoter of various plant genes.
969 Nucleic Acids Res. 1989;17:9707–17.
- 970 51. Terada R, Nakayama T, Iwabuchi M, Shimamoto K. A type I element composed
971 of the hexamer (ACGTCA) and octamer (CGCGGATC) motifs plays a role(s) in
972 meristematic expression of a wheat histone H3 gene in transgenic rice plants. Plant
973 Mol Biol. 1995;27:17–26.
- 974 52. Tabata T. Leucine-zipper type transcription factors of wheat. Bot Mag Tokyo.
975 1991;104:171–81.
- 976 53. Mikami K, Nakayama T, Takefumi, Kawata, Tabata T, Iwabuchi M. Specific
977 interaction of nuclear protein HBP-1 with the conserved hexameric sequence
978 ACGTCA in the regulatory region of wheat histone genes. Plant Cell Physiol.
979 1989;30:107–19.
- 980 54. Yanagisawa S, Schmidt RJ. Diversity and similarity among recognition sequences
981 of Dof transcription factors. Plant J. 1999;17:209–14.
- 982 55. Yanagisawa S. Dof1 and Dof2 transcription factors are associated with expression
983 of multiple genes involved in carbon metabolism in maize. Plant J. 2000;21:281–
984 8.

- 985 56. Yanagisawa S. Dof domain proteins: plant-specific transcription factors
986 associated with diverse phenomena unique to plants. *Plant Cell Physiol.*
987 2004;45:386–91.
- 988 57. Engel ML, Holmes-Davis R, McCormick S. Green sperm. identification of male
989 gamete promoters in *Arabidopsis*. *Plant Physiol.* 2005;138:2124–33.
- 990 58. Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, et al. Functional
991 analysis of *cis*-regulatory elements within the promoter of the tobacco late pollen
992 gene *g10*. *Plant Mol Biol.* 2001;45:577–85.
- 993 59. Bate N, Twell D. Functional architecture of a late pollen promoter: pollen-specific
994 transcription is developmentally regulated by multiple stage-specific and co-
995 dependent activator elements. *Plant Mol Biol.* 1998;37:859–69.
- 996 60. Mitsuda N, Hisabori T, Takeyasu K, Sato MH. VOZ; isolation and
997 characterization of novel vascular plant transcription factors with a one-zinc finger
998 from *Arabidopsis thaliana*. *Plant Cell Physiol.* 2004;45:845–54.
- 999 61. Mitsuda N, Takeyasu K, Sato MH. Pollen-specific regulation of vacuolar H⁺-
1000 PPase expression by multiple *cis*-acting elements. *Plant Mol Biol.* 2001;46:185–
1001 92.
- 1002 62. Tang W, Perry SE. Binding site selection for the plant MADS domain protein
1003 AGL15: an *in vitro* and *in vivo* study. *J Biol Chem.* 2003;278:28154–9.
- 1004 63. Yamagata H, Yonesu K, Hirata A, Aizono Y. TGTCACA motif is a novel *cis*-
1005 regulatory enhancer element involved in fruit-specific expression of the *cucumisin*
1006 gene. *J Biol Chem.* 2002;277:11582–90.

- 1007 64. Hamilton DA, Schwarz YH, Mascarenhas JP. A monocot pollen-specific
1008 promoter contains separable pollen-specific and quantitative elements. Plant Mol
1009 Biol. 1998;38:663–9.
- 1010 65. Carabelli M, Turchi L, Ruzza V, Morelli G, Ruberti I. Homeodomain-Leucine
1011 zipper II family of transcription factors to the limelight: central regulators of plant
1012 development. Plant Signal Behav. 2013;8:e25447.
- 1013 66. Carabelli M, Sessa G, Baima S, Morelli G, Ruberti I. The *Arabidopsis Athb-2* and
1014 *-4* genes are strongly induced by far-red-rich light. Plant J. 1993;4:469–79.
- 1015 67. Henriksson E, Olsson AS, Johannesson H, Johansson H, Hanson J, Engstrom P,
1016 et al. Homeodomain leucine zipper class I genes in Arabidopsis. Expression
1017 patterns and phylogenetic relationships. Plant Physiol. 2005;139:509–18.
- 1018 68. Wang ZY, Kenigsbuch D, Sun L, Harel E, Ong MS, Tobin EM. A Myb-related
1019 transcription factor is involved in the phytochrome regulation of an Arabidopsis
1020 *Lhcb* gene. Plant Cell. 1997;9:491–507.
- 1021 69. Sablowski RW, Moyano E, Culianez-Macia FA, Schuch W, Martin C, Bevan M.
1022 A flower-specific Myb protein activates transcription of phenylpropanoid
1023 biosynthetic genes. EMBO J. 1994;13:128–37.
- 1024 70. Rose A, Meier I, Wienand U. The tomato I-box binding factor LeMYBI is a
1025 member of a novel class of myb-like proteins. Plant J. 1999;20:641–52.
- 1026 71. Urao T, Yamaguchi-Shinozaki K, Urao S, Shinozaki K. An Arabidopsis *myb*
1027 homolog is induced by dehydration stress and its gene product binds to the
1028 conserved MYB recognition sequence. Plant Cell. 1993;5:1529–39.

- 1029 72. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. *Arabidopsis*
1030 *AtMYC2* (bHLH) and *AtMYB2* (MYB) function as transcriptional activators in
1031 abscisic acid signaling. *Plant Cell*. 2003;15:63–78.
- 1032 73. Cominelli E, Tonelli C. A new role for plant R2R3-MYB transcription factors in
1033 cell cycle regulation. *Cell Res*. 2009;19:1231–2.
- 1034 74. Sakai H, Aoyama T, Oka A. *Arabidopsis* ARR1 and ARR2 response regulators
1035 operate as transcriptional activators. *Plant J*. 2000;24:703–11.
- 1036 75. Ross EJ, Stone JM, Elowsky CG, Arredondo-Peter R, Klucas RV, Sarath G.
1037 Activation of the *Oryza sativa* non-symbiotic haemoglobin-2 promoter by the
1038 cytokinin-regulated transcription factor, ARR1. *J Exp Bot*. 2004;55:1721–31.
- 1039 76. Xie Z, Zhang ZL, Zou X, Huang J, Ruas P, Thompson D, et al. Annotations and
1040 functional analyses of the rice *WRKY* gene superfamily reveal positive and
1041 negative regulators of abscisic acid signaling in aleurone cells. *Plant Cell Physiol*.
1042 2005;137:176–89.
- 1043 77. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with
1044 RNA-Seq. *Bioinformatics*. 2009;25:1105–11.
- 1045 78. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential
1046 gene and transcript expression analysis of RNA-seq experiments with TopHat and
1047 Cufflinks. *Nat Protoc*. 2012;7:562–78.
- 1048 79. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al.
1049 Transcript assembly and quantification by RNA-Seq reveals unannotated

- transcripts and isoform switching during cell differentiation. Nat Biotechnol.
2010;28:511–5.
80. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
81. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31:46–53.
82. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.
83. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
84. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007;8:R183.
85. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 2018;27:293–315.
86. Kozakov D, Hall DR, Xia B, Porter KA, Padhorney D, Yueh C, et al. The ClusPro web server for protein–protein docking. Nat Protoc. 2017;12:255–78.

Supporting information

1071 **S1 Figure. Molecular confirmation of the *atpol2a-1* mutant.**

1072 **S2 Figure. Molecular confirmation of *atpol2b* mutants.**

1073 **S1 Table List of DEGs detected in *atpol2a* and *atpol2b* mutants as compared to the**
1074 **WT.** This file lists the DEGs detected in *atpol2a* (S1A Table), *atpol2b-1* (S1B Table),
1075 *atpol2b-2* (S1C Table) and *atpol2b-3* (S1D Table) mutants as compared to the WT.

1076 **S2 Table. ClusPro docking scores for DEGs in *atpol2a* and *atpol2b* mutants with**
1077 **POL2 protein of *S. cerevisiae*.**

1078 **S3 Table. List of important CREs detected in genomic regions (2500 bp upstream**
1079 **of the translation start site) of *AtPOL2A* and *AtPOL2B*.** This file lists the CREs
1080 detected in both *AtPOL2A* and *AtPOL2B* (S3A Table), only in *AtPOL2A* (S3B Table)
1081 and only in *AtPOL2B* (S3C Table).

1082 **S4 Table. Primer sequences used in mutant screening**

1083 **S5 Table. Primer sequences used to amplify *AtPOL2A* and *AtPOL2B* cDNA by RT-**
1084 **qPCR.**

(A)

<https://doi.org/10.1101/2022.02.14.480442>

CC-BY 4.0 International license

(C)

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8

Figure 11

(A)

(B)

(C)

(D)

(E)

Figure 12

Figure 13

Figure 14