

**Supplemental Figure 1.** Complementation of the *cep1* Mutant by *CEP1* Genomic DNA.

**(A)** Pollen germination rate of 55 independent complementation transgenic lines. Bars represent standard deviations. **(B)** Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) of *CEP1* expression in buds tissues at stages 7–10 in  $T_1$  transgenic plants. The *ACTIN1* gene (AT2G37620) was used as a reference for normalization. **(C)** and **(D)** Cracking anthers and released pollen grains in the complementation transgenic lines **(C)** and the *cep1* mutant **(D)**; Bars = 50  $\mu$ m. **(E)** and **(F)** Pollen germination of the complementation transgenic lines **(E)** and *cep1* mutant **(F)**; Bars = 50  $\mu$ m. **(G–L)** Transmission electron micrographs of the tapetum and microspore development in the complementation transgenic lines at stage 8 [(G) and (H)], stage 10 [(I) and (J)], and early stage 11[(K) and (L)]. Bars = 2  $\mu$ m [(G–L)].

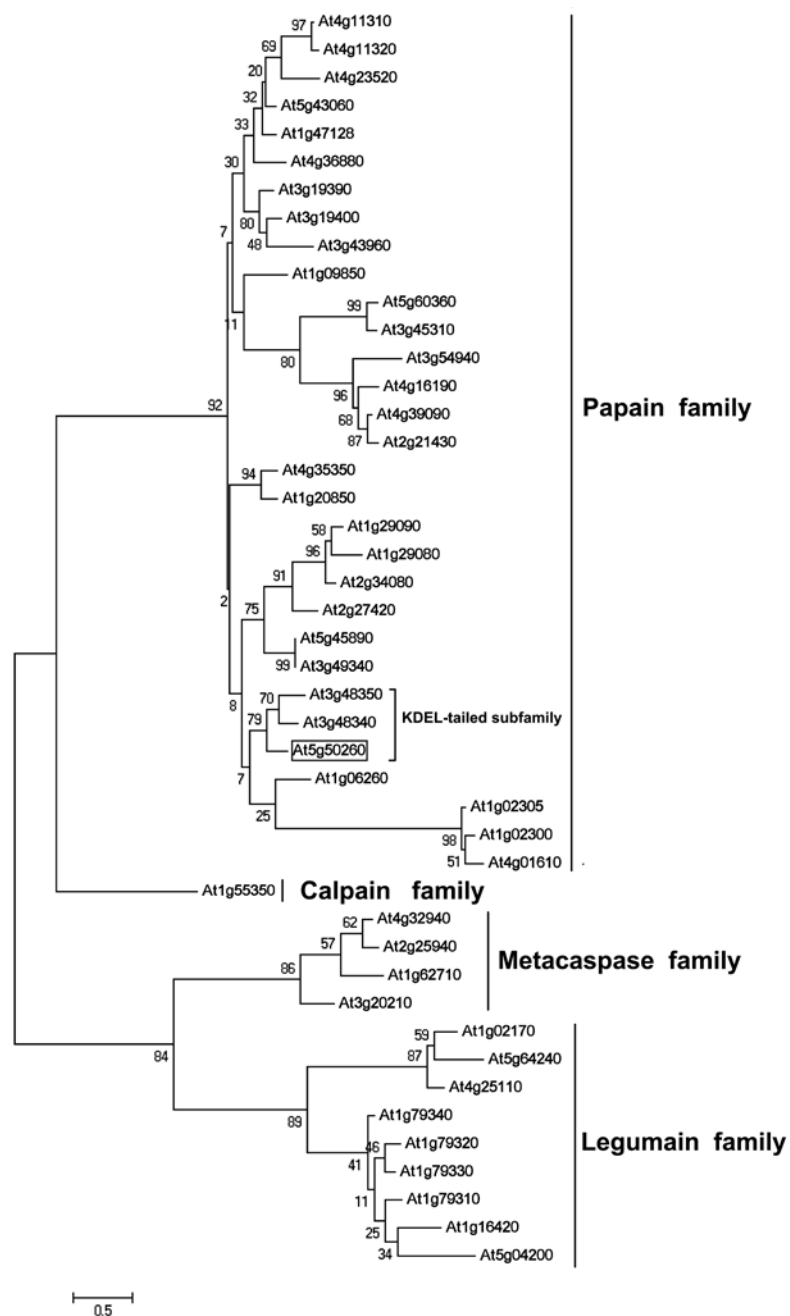
```

ATGAAGCGATTATTGTTCTTGCCTTTGCATGCTTATGGTCTTGAAACCACAAAGGGTTTAGATTCCATAACAAAGATGTGGAATCAGAGAATAGCTTGTGGGAGCT 110
M K R F I V L A L C M L M V L E T T K G L D F H N K D V E S E N S L W E L
-----
ATACGAACCGTGGAGGAGTCACCACACTGTGCTAGGAGCCTTGAAGAGAAAGCTAAGAGGTTCAACGTGTTCAAGCACACGTGAAGCACATCCATGAGACCAACAAGA 220
Y E R W R S H H T V A R S L E E K A K R F N V F K H N V K H I H E T N K
AGGATAAATCTTACAAGCTCAAGCTAAACAAGTTTGGTGACATGACTAGTGAAGAATTTAGGAGAACATACGCTGGTTCAACATCAAACATCATAGGATGTTTCAAGGC 330
K D K S Y K L K L N K F G D M T S E E F R R T Y A G S N I K H H R M F Q G
GAAAAAAGCCACAAAGAGCTTCATGTATGCAAAATGTAATACTCTCCCAACCTCCGTTGATTGGAGAAAGAATGGAGCCGTCACCTCCTCAAAAACCAAGGCCAATG 440
E K K A T K S F M Y A N V N T L P T S V D W R K N G A V T P V K N C G Q C
CGGGAGTTGTTGGGCGTTTTCAACAGTTGTTGCGGTGGAAGGGATCAACCAAAATAAGAACTAAAAAGCTGACATCACTCTCAGAGCAAGAGCTAGTAGATTGTGATACAA 550
G S C W A F S T V V A V E G I N Q I R T K K L T S L S E Q E L V D C D T
ACCAGAACCAGGATGCAATGGAGGCTAATGGACCTTGCTTTTGAGTTCATCAAGGAGAAAGGAGACTCACAAGTGAGCTAGTGTACCCTTACAAGGCTTGTGATGAA 660
N Q N Q G C N G G L M D L A F E F I K E K G G L T S E L V Y P Y K A S D E
ACTTGTGACACAACAAAGAAAATGCTCCGAGTATTTCAATCGATGGACAGAAAGATGTCCTTAAGAACAGCGAGGATGATCTAATGAAAGCTGTGCTAATCAGCCGTG 770
T C D T N K E N A P V V S I D G H E D V P K N S E D D L M K A V A N Q P V
TTCTGTGCTATTGATGCTGGAGGCTCAGACTTTCAATTTACTCCGAGGGAGTGTGTTACCGGGAGATGCGGAACAGAGCTAAACCATGGAGTTGCGGTAGTAGGGTATG 880
S V A I D A G G S D F Q F Y S E G V F T G R C G T E L N H G V A V V G Y
GAACAACGATAGACGGAACAAAGTATTGGATTGTTAAGAATTCATGGGGAGAGGAATGGGGAGAGAAAGGATACATAAGAATGCAGAGAGGGATTGTCATAAAGAAGGA 990
G T T I D G T K Y W I V K N S W G E E W G E K G Y I R M Q R G I R H K E G
CTTTGTGATGCAATGGAAGCTTCTTATCCTCTCAAGAACTCCAACACTAACCTTCTAGACTTTCCTTGGACTCGCTTAAGGATGAACCTAA 1086
L C G I A M E A S Y P L K N S N T N P S R L S L D S L K D E L *

```

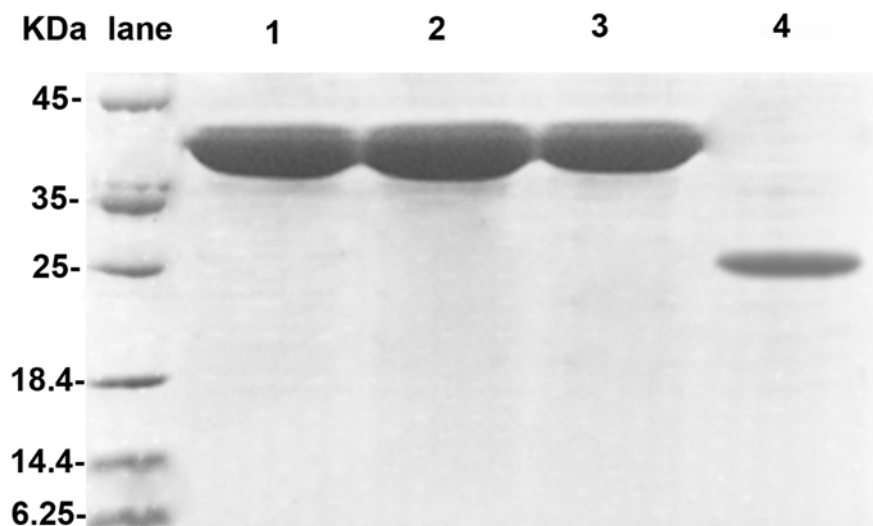
**Supplemental Figure 2.** Nucleotide and Deduced Amino Acid Sequence of *At5g50260*.

The signal peptide (M1-G20) is shown as a dashed underline. The catalytic triad Cys, His, and Asn, and the Glu active site residue are circled. The GCNNGG motif is double-underlined. The ERFNIN motif (E53-N72) is shown in a rectangular box, and the KDEL (K358-L361) motif is underlined. The arrow indicates the self-hydrolysis site. Numbers in the right margins represent nucleotide sequences.



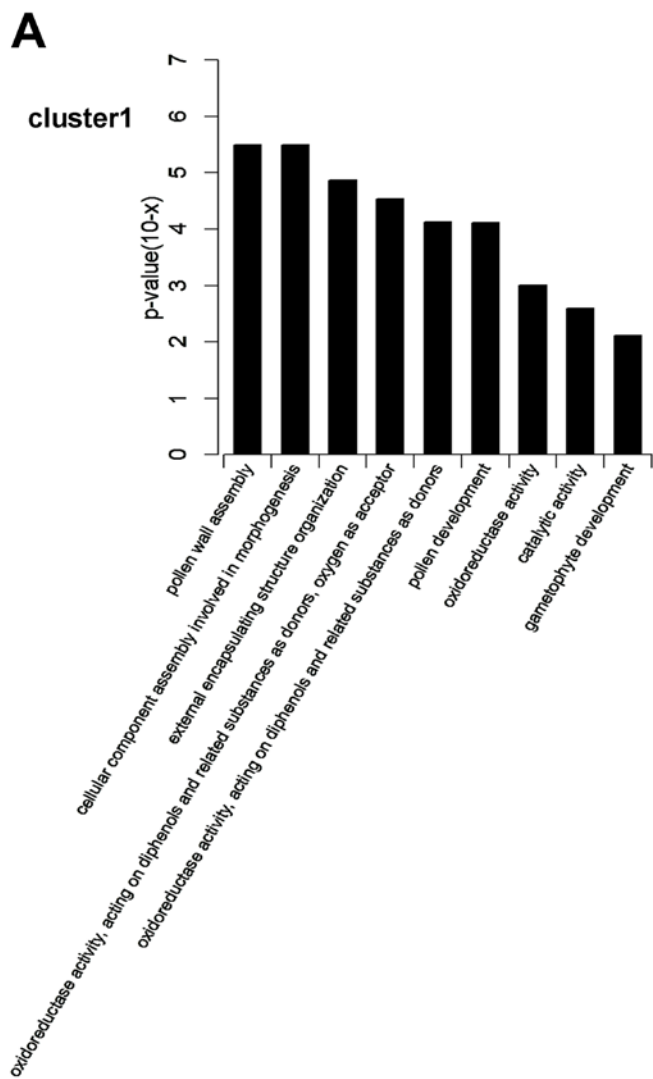
**Supplemental Figure 3.** A Phylogenetic Tree of Papain-like Cysteine Proteases in *Arabidopsis*.

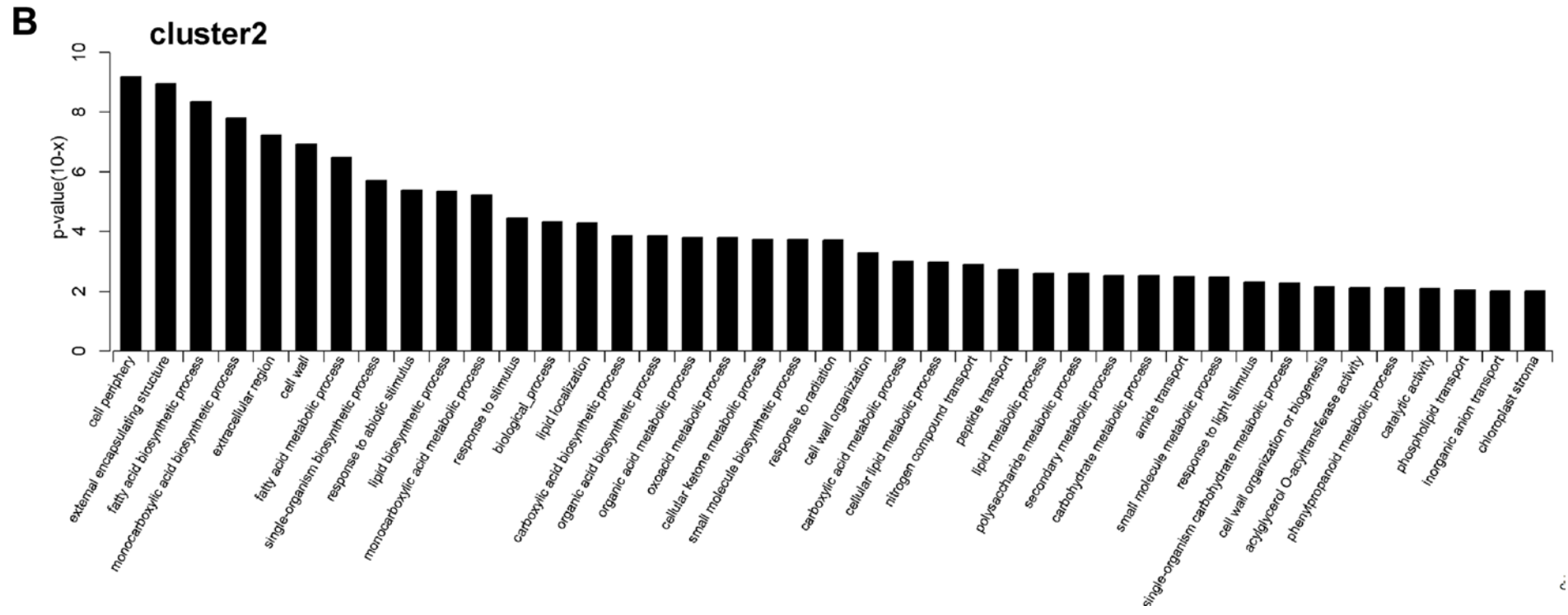
A phylogenetic analysis was conducted using MEGA 5.0. All 51 papain-like cysteine proteases in *Arabidopsis* were selected to establish an unrooted maximum likelihood tree. Numbers at the nodes represent percentage bootstrap values based on 1000 replications. Bootstrap values are shown as percentages. The rectangular box shows the position of At5g50260.

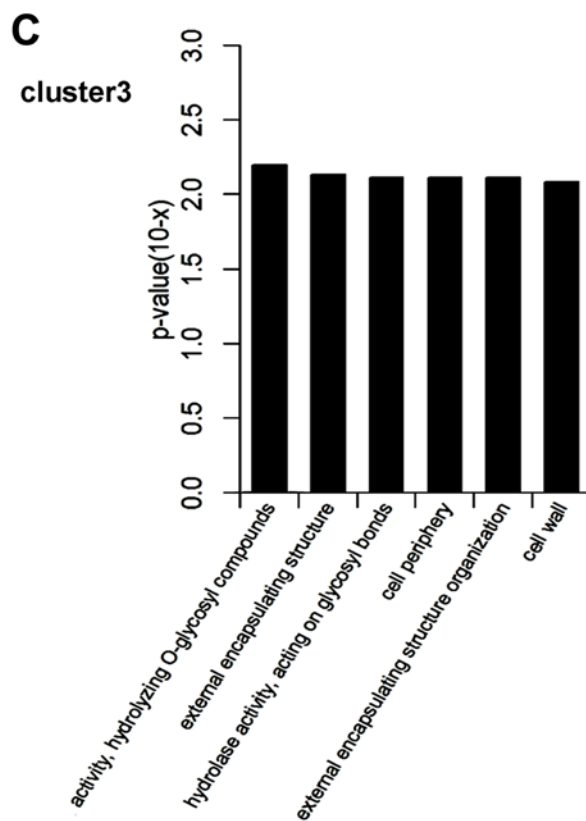


**Supplemental Figure 4.** Purified Recombinant CEP1.

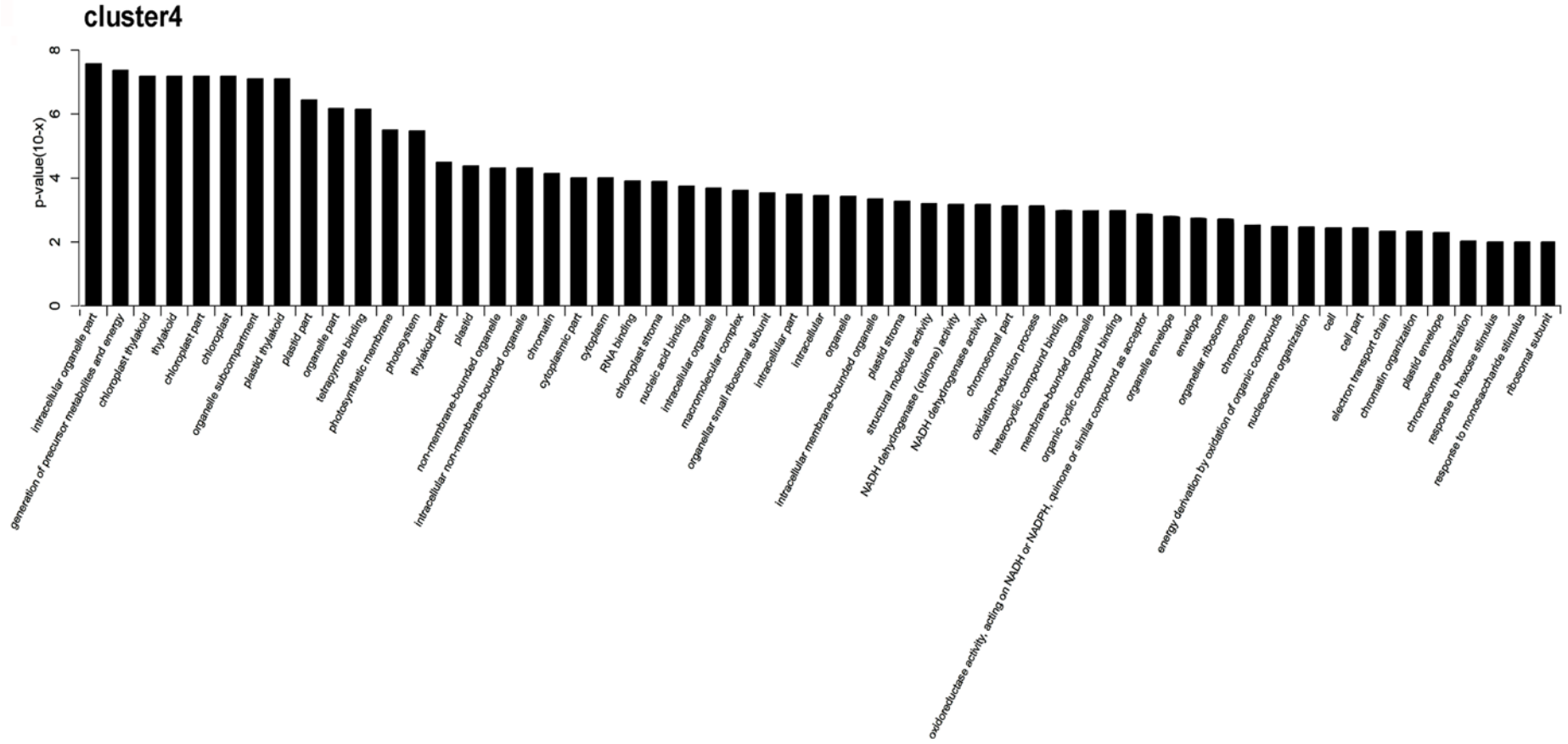
Purified recombinant CEP1 enzyme (lanes 1–3); self-hydrolysis occurred within 12 min at pH 3.0 (lane 4).



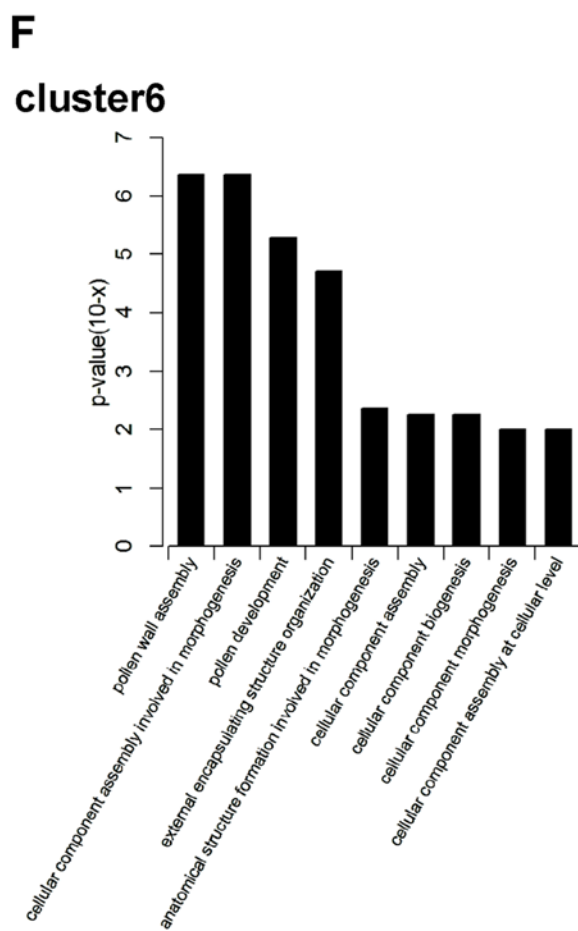
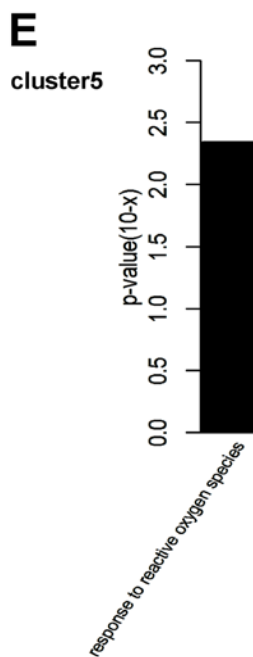




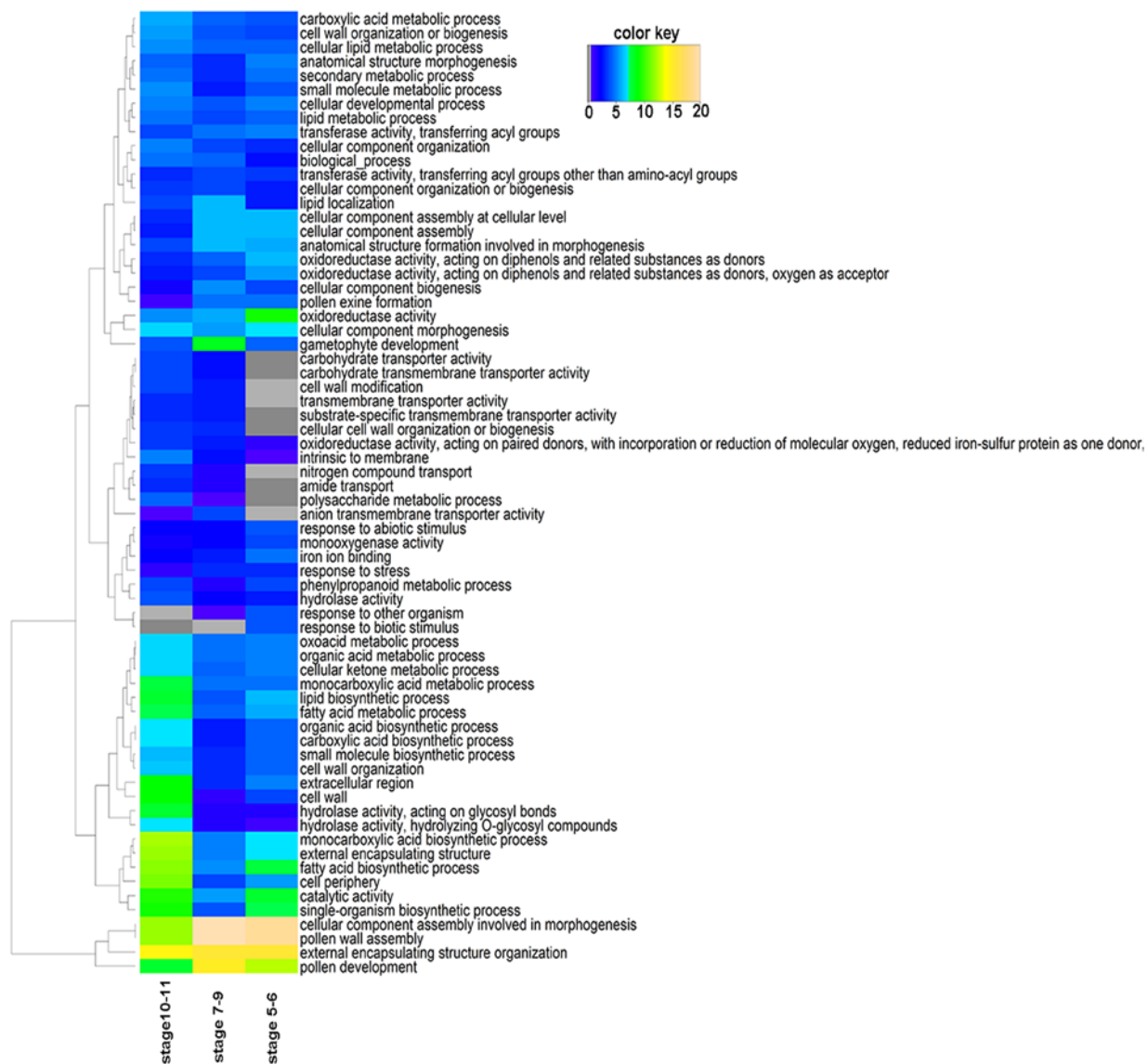
D



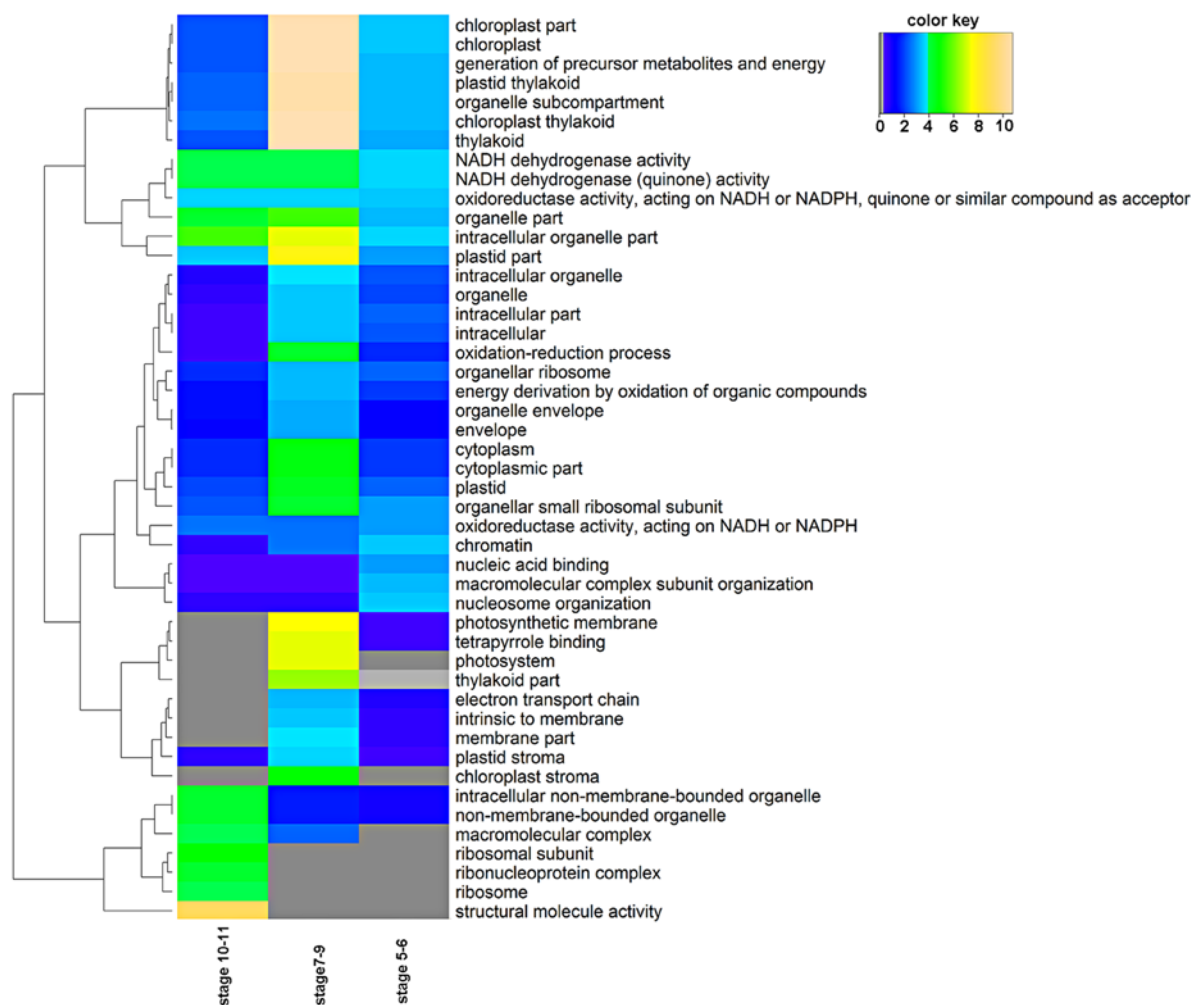




**G**

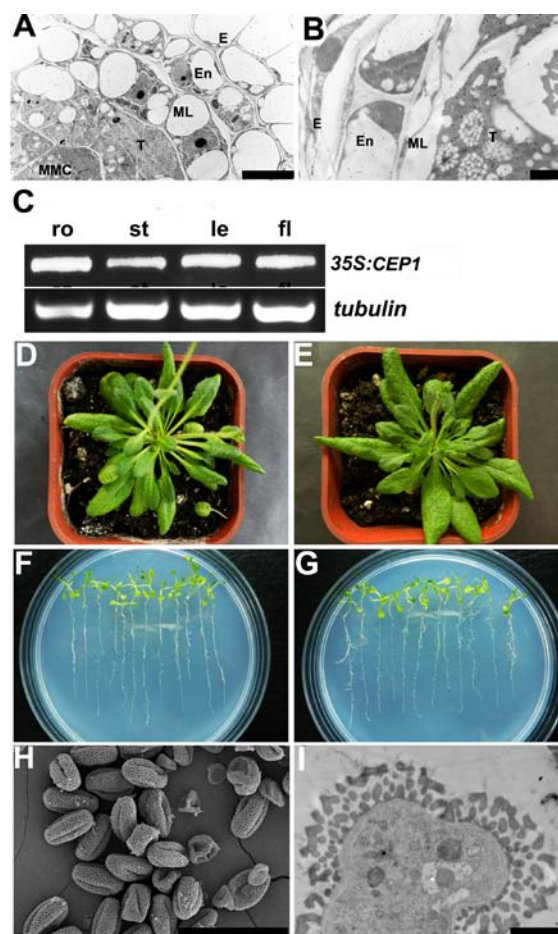


**H**



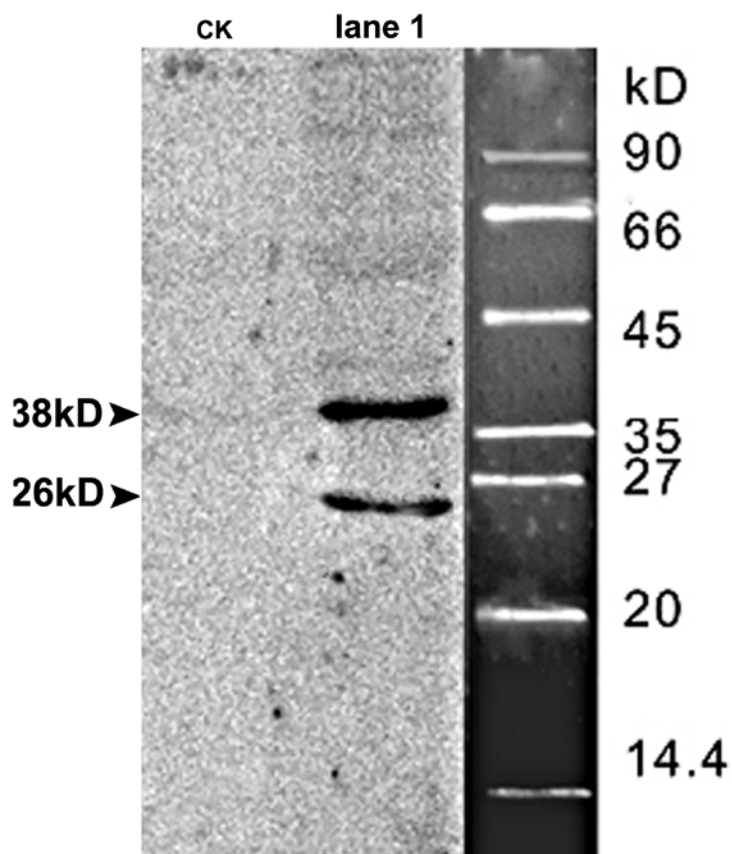
**Supplemental Figure 5.** Gene Ontology (GO) Term Enrichment of Differentially Expressed Genes in the *cep1* Mutant.

(A–F) GO terms that are significantly enriched ( $P$ -value  $< 0.01$ ) in cluster 1 (A), cluster 2 (B), cluster 3 (C), cluster 4 (D), cluster 5 (E), cluster 6 (F). GO terms were sorted based on the  $P$ -value. (G) and (H) GO enrichment of the significantly upregulated (G) and downregulated (H) genes (*cep1* mutant vs. the wild type;  $P < 0.005$ , fold-change  $> 2$  or  $< -2$ ) in at least one of the three stages (stages 5–6, stages 7–9, and stages 10–11). GO terms were sorted based on the  $P$ -value.



**Supplemental Figure 6.** *CEP1* Overexpression Phenotype Using the CaMV35S and *ProCEP1* Promoters.

(A) and (B) Immunogold localization of CEP1 in the anther layers of the *35S:CEP1*-positive transgenic line at stages 5 (A) and 9 (B). E, epidermis; En, endothecium; ML, middle layer; MMC, microspore mother cell; T, tapetum. Bars = 5  $\mu$ m (A), 2  $\mu$ m (B). (C) *CEP1* spatial and temporal expression analyses by reverse transcription-polymerase chain reaction (RT-PCR) in the *35S:CEP1* transgenic lines. ro, roots; st, stems; le, leaves; and fl, flower. (D) and (E). Comparison of the *35S:CEP1* overexpressing transgenic plants (D) and a wild-type plant (E) at the vegetative stage. (F) and (G) Comparison of *35S:CEP1* overexpressing transgenic lines (F) and wild-type plants (G) at the seedling stage. (H) and (I) The *ProCEP1:CEP1* overexpressing transgenic line pollen grains. Bars = 50  $\mu$ m (H) and 2  $\mu$ m (I).



**Supplemental Figure 7** Specificity Analysis of the CEP1 antibodies by immunoblot

CK, the negative control, the rabbit serum without immunity hybridized with the protein extracted from the buds of *Arabidopsis*.; Lane1, The protein extracted from the buds of wild-type *Arabidopsis*.

**Supplemental Table 1.** Quantitative Reverse Transcription-Polymerase Chain Reaction Corroboration of Differentially Expressed Genes.

Gene ID	Fold-change	
	Transcriptomic data	q-RT-PCR
<b>Stage 5-6</b>		
At1G01280.1	2.62	2.92
At1G20130.1	4.74	4.29
At1G50310.1	3.64	3.48
At3G11980.1	2.04	2.16
At3G20670.1	-1.10	-0.93
At3G26125.1	3.32	4.15
At4G33790.1	1.41	1.44
At3G45930.1	-1.12	-0.82
At5G07430.1	4.77	5.01
At5G14980.1	2.02	2.13
<b>Stage7-9</b>		
At1G01280.1	2.54	2.52
At1G20130.1	4.03	4.01
At1G50310.1	3.71	3.68
At4G13230.1	2.89	2.87
At3G05610.1	-1.13	-1.03
At3G05960.1	4.50	4.55
At3G11980.1	1.87	1.74
At3G20670.1	-1.13	-1.47
At3G26125.1	2.38	3.19
At3G45930.1	-1.19	-0.85
At4G33790.1	1.09	1.03
At5G14980.1	3.02	3.11
At5G48140.1	-1.53	-1.02
<b>Stage10-11</b>		
At1G01280.1	5.59	3.68
At1G20130.1	5.06	2.64
At1G26820.1	3.32	2.03
At1G50310.1	3.07	2.23
At3G05960.1	4.45	2.13
At3G11980.1	3.11	2.07
At3G20670.1	-1.05	-2.28
At3G45930.1	-1.16	-2.12
At4G13230.1	3.11	3.11
At4G33790.1	1.94	1.32
At5G07430.1	0.62	1.68
At5G45890.1	7.42	5.57
<b>Pearson correlation coefficient</b>		0.93

Fold-change: Log<sub>2</sub> (*cep1*/wild type).

**Supplemental Table 2.** Partial List of Genes Differentially Expressed in the *cep1* Mutant

Gene ID	Log <sub>2</sub> Ratio ( <i>cep1</i> /WT)			Gene Name	Cluster
	Stage 5-6	Stage 7-9	Stage 10-11		
<b>Chloroplast or plastids formation</b>					
ATCG00020	-2.34	-1.67	0	PSBA (photosystem II reaction center protein A)	4
ATCG00070	0	-2.03	0	PSBK (photosystem II reaction center protein K precursor)	4
ATCG00300	0	-2.31	0	YCF9 (photosystem II reaction center protein Z)	4
ATCG00680	0	-1.41	0	PSBB (photosystem II reaction center protein B)	4
ATCG00220	-1.80	-1.35	-1.31	PSBM (photosystem II reaction center protein M)	4
AT1G30380	-0.53	-1.00	-0.51	PSAK (photosystem I reaction center subunit K)	4
ATCG00340	-1.41	-1.65	0	PSAB (photosystem I reaction center subunit D1)	4
ATCG01050	-1.67	-2.29	0	NDHD (NADH dehydrogenase)	4
ATCG00150	-1.91	-2.24	-1.62	ATPI (ATPase complex CF0 subunit)	4
ATCG00160	-1.54	-2.23	-2.52	RPS2 (chloroplast ribosomal protein S2)	4
ATCG00380	-1.86	-2.07	-1.75	RPS4 (chloroplast ribosomal protein S4)	4
ATCG00650	-2.04	-2.10	-2.45	RPS18 (chloroplast ribosomal protein S18)	4
ATCG00330	-2.62	-2.66	-2.07	RPS14 (chloroplast ribosomal protein S14)	4
<b>Cell wall</b>					
AT1G21310	0.62	1.32	0.78	EXT3 (extensin 3)	2
AT5G25550	0	3.53	2.96	leucine-rich repeat extensin-like protein 7	1
AT2G43150	1.08	1.39	0	Proline-rich extensin-like family protein	5
AT4G38770	3.37	2.58	2.98	PRP4 (proline-rich protein 4)	1
AT1G33590	1.07	0.88	1.47	Leucine-rich repeat (LRR) family protein	2
AT3G25050	3.42	2.38	2.85	XTH3 (xyloglucan endotransglucosylase /hydrolase 3)	1
AT2G38080	3.82	2.89	2.89	LAC4 (laccase-like multicopper oxidase 4)	1
AT4G33220	1.15	0.97	1.33	PME44 (pectin methylesterase 44)	2
<b>Pollen / Pollen wall assembly</b>					
AT2G13680	0	2.24	2.62	CALS5 (callose synthase 5)	6
AT1G62940	1.42	2.29	0	ACOS5 (acyl-CoA synthetase 5)	6
AT1G01280	2.62	2.54	5.58	CYP703A2 (cytochrome P450, family 703, subfamily A, polypeptide 2)	1
AT1G69500	1.39	1.42	3.52	CYP704B1(cytochrome P450, family 704, subfamily B, polypeptide 1)	6
AT1G02050	1.57	1.66	0	PKSA (polyketide synthase A)	5

AT4G34850	1.42	2.20	4.32	PSKB (polyketide synthase B)	6
AT4G35420	1.06	1.31	0	TKPR1 (tetraketide alpha-pyrone reductase 1)	5
AT3G13220	1.44	2.08	0	ABCG26 (ATP-binding cassette transporter G26)	6
AT3G11980	2.04	1.87	3.11	MS2 (fatty acyl-CoA reductase 2)	6
<b>Fatty acid and lipid biosynthetic process</b>					
AT1G01120	1.23	1.32	1.68	KCS1 (3-ketoacyl-CoA synthase 1)	2
AT1G01610	0.81	0.72	1.10	GPAT4 (Bifunctional sn-glycerol-3-phosphate 2-O-acyltransferase/phosphatase)	2
AT1G17420	1.26	1.21	1.56	LOX3 (Lipoxygenase)	2
AT1G49430	1.03	0.59	1.16	LACS2 (long chain acyl-CoA synthetase)	2
AT1G55020	1.31	0	1.35	LOX1 (lipoxygenase)	2
AT1G67560	0	0	1.39	LOX6 (lipoxygenase family protein)	2
AT2G38110	1.00	1.01	2.03	GPAT6 (Bifunctional sn-glycerol-3-phosphate 2-O-acyltransferase/phosphatase)	2
AT3G45140	1.39	1.1	1.49	LOX2 (Chloroplast lipoxygenase)	2
AT4G33790	1.41	0	1.94	CER4 (alcohol-forming fatty acyl-CoA reductase)	2
AT4G34510	0	0	2.05	KCS17 (3-ketoacyl-CoA synthase family)	2
AT4G37760	1.13	0	1.38	SQE3 (squalene epoxidase 3)	2
AT5G37300	0.91	0	1.09	WSD1 (wax ester synthase (WS) and diacylglycerol acyltransferase (DGAT))	2
AT5G57800	1.13	0.72	1.36	CER3 (similarity to the sterol desaturase family)	2
<b>Cell death, Senescence-associated genes</b>					
AT5G45890	0	0	7.34	SAG12 (senescence-associated protein 12)	1
AT1G06260	3.91	3.75	0	Cysteine proteinases superfamily protein	1
AT1G20850	2.26	1.79	1.32	XCP2 (xylem cysteine peptidase 2)	5
AT4G30040	3.61	0	0	aspartyl protease family protein	1
AT4G04460	0	4.56	0	Saposin-like aspartyl protease family protein	1
AT3G52500	0.83	1.14	1.43	aspartyl protease family protein	2



**Supplemental Table 3.** Primer Information Used for Molecular Cloning and Plasmid Construction.

<b>Primer name</b>	<b>Sequence(5'-3')</b>
<b>CEP1M3-F</b>	CAAATCTTAGTTTCGACGATGG
<b>CEP1M3-R</b>	GATTGGAGAAAGAATGGAGCC
<b>CEP1M4-F</b>	GCTTACCTTTCGCTGTTGCTA
<b>CEP1M4-R</b>	AAGCTGTTGCTAATCAGCCTG
<b>CEP1M3/ M4-B</b>	ATTTTGCCGATTTCCGGAAC
<b>CEP1-C-F</b>	TACCCGGGATGAAGCGATTTATTG
<b>CEP1-C-R</b>	AGGAGCTCAATTTAGAGTTCATCCT
<b>CEP1-CM-F</b>	GGCGGTACCTTAGATTTCCATAAC
<b>CEP1-CM-R</b>	GGGGAGCTCTTAGAGTTCATCCTTA
<b>CEP1-SH-F</b>	GGCCTGCAGCAGAAAGCTAAGA
<b>CEP1-SH-R</b>	GGGGAGCTCTAAATTCTTCACTA
<b>ProCEP1-F</b>	GCTCTAGAATTTTTCTGATTTTGGAG
<b>ProCEP1-R</b>	GCCCCCGGGATTTTACTTTTCTTATTC
<b>Pro-CEP1-F</b>	CGCGAGCTCATTTTTCTGATTTTG
<b>Pro-CEP1-R</b>	CGCCTCGAGACATAAAACCATT
<b>35SOEx-F</b>	GGCCCCGGGAATATGAAGCGATTTA
<b>35SOEx-R</b>	GGGGAGCTCGGGAATTTAGAG
<b>An-CEP1-F</b>	GTGGATCCGAGCTAGTGTACCCTT
<b>An-CEP1-R</b>	GTGAATTCTTAGAGTTCATCCTTA
<b>GFP-F</b>	GCTCTAGAATGAGTAAAGGAGA
<b>GFP-R</b>	GCCCCGGGTTTGTATAGTTCAT

**Supplemental Table 4.** Primer Information Used for Reverse Transcription-Polymerase Chain Reaction (RT-PCR)/Quantitative RT-PCR.

<b>Primer name</b>	<b>Sequence(5'-3')</b>
RT-CEP1-F	TATACGAACGGTGGAGGAGTCACC
RT-CEP1-R	TCCGCATCTCCCGGTAAACACTCC
tubulin-F	GCGGACGACCAACTACATAC
tubulin-R	GAGACTATCGCTTCCATCTGCTT
qRT-CEP1-F	CTATTGATGCTGGAGGCTCAGACT
qRT-CEP1-R	GAATCCCTCTCTGCATTCTTATGT
qRT-Actin-F	CGTATGAGCAAGGAGATCAC
qRT-Actin-R	CACATCTGTTGGAAGGTGCT
AT5G45890F	CAACGTCGAACGCATTGA
AT5G45890R	GTTTGGCTTTGGCTAGATAA
AT1G26820F	TTACGGGTTTCTGATCTGAT
AT1G26820R	GCATGAAGGAGATTAGCTTT
AT4G33790F	CTCAATGTCTTGAACCTTGCC
AT4G33790R	TCAAGTTTCTCCTGAACCAAT
AT4G13230F	CACAGCCAAGACCGTTGAGA
AT4G13230R	CATATCATAACCCGATTT
AT3G11980F	GTGAATGGACAAAGACAAG
AT3G11980R	CCTTTTCTAGCAGCTTCAA
AT3G26125F	GCTTAAGGACCAAGGAACATT
AT3G26125R	GGGTGTTTTGTTATCAACCAG
AT1G20130F	TGCAACTGGTAGATTCAGCAA
AT1G20130R	CTGCTTCGGATGTAGTAGGAT
AT1G01280F	GAACAATGTGACTAGGATGCT
AT1G01280R	CACGTCCCTCATTTCTTTCTC
AT5G48140F	TCATCAATTCATTTATCTCCA
AT5G48140R	GATGCCACTAACGTTTTCTC
AT1G50310F	GAAGTACGGACGCAAAATCTC
AT1G50310R	ACCCGATGTTAAGTGCTCCTC
AT3G05960F	ATCATCGGAATGCTTCTCCAG
AT3G05960R	ACGAGGTAAATCCCAACGAAC
AT3G05610F	CGGCAACAAAATTACAAGGA
AT3G05610R	TGTAGAGCGTGTCTTGGTAGC
AT5G07430F	GACAAGCTCTAGCAATGAGGA
AT5G07430R	GTGCAATTGCGTATTCAAGTA
AT3G20670F	GTGAGAAACGATGAGGAGCTG
AT3G20670R	ACAACCACAACACAAATCCCT
AT3G45930F	GTAGAGGTGGAGTCAAGCGTA
AT3G45930R	CCGAATCCGTAGAGAGTTCTT
AT5G14980F	TCAGTCCTCTTCGCAAAATCC
AT5G14980R	GGAGAAAGGAGGAAGCGTGAC