The Plant Journal - Supporting Information

A tale of two morphs: developmental patterns and mechanisms of seed coat differentiation in the dimorphic diaspore model *Aethionema arabicum* (Brassicaceae) [CC-BY]

Waheed Arshad[1], Teresa Lenser[2], Per KI Wilhelmsson[3], Jake O Chandler[1], Tina Steinbrecher[1], Federica Marone[4], Marta Pérez[1], Margaret E Collinson[5], Wolfgang Stuppy[6,7], Stefan A Rensing[3], Günter Theißen[2], Gerhard Leubner-Metzger[1,8,*]

Author affiliations

[1] Department of Biological Sciences, Royal Holloway University of London, Egham, TW20 0EX, United Kingdom; Web: "The Seed Biology Place" – *www.seedbiology.eu*

[2] Matthias Schleiden Institute / Genetics, Friedrich Schiller University Jena, D-07743 Jena, Germany

[3] Plant Cell Biology, Department of Biology, University of Marburg, D-35043 Marburg, Germany

[4] Swiss Light Source, Paul Scherrer Institute, CH-5232 Villigen, Switzerland

[5] Department of Earth Sciences, Royal Holloway University of London, Egham, TW20 0EX, United Kingdom

[6] Botanischer Garten der Ruhr-Universität Bochum, Universitätsstraße 150, D-44780, Bochum, Germany

[7] The Royal Botanic Gardens, Kew, Wellcome Trust Millennium Building, Wakehurst Place, Ardingly, West Sussex RH17 6TN, United Kingdom

[8] Laboratory of Growth Regulators, Palacký University and Institute of Experimental Botany, Czech Academy of Sciences, CZ-78371 Olomouc, Czech Republic

[*] For correspondence (e-mail gerhard.leubner@rhul.ac.uk)

Supporting Information - Detail Experimental Procedures

Whole seed staining and developmental analysis of seed coat differentiation

Whole M^+ and M^- seeds were imbibed in 0.01% (w/v) ruthenium red (Sigma-Aldrich, 11103-72-3) or 0.01% (w/v) methylene blue (VWR, 3470.0025) for two minutes, then visualised under a Leica MZ-125 stereomicroscope. Developing gynoecia from 0 to 7 DAP were harvested and fixed in vacuum-infiltrated FAA fixation solution (2% formaldehyde, 5% glacial acetic acid, 60% EtOH, 0.1% Tween-20) at 4°C for 24 h. After sample dehydration and clearing using HistoClear™, samples were transferred into fresh melted paraffin and embedded into Peel-A-Way® moulds. A Microm HM 355 S rotary microtome (Walldorf, Germany) with Leica 819 low-profile disposable blades was used to prepare 6 µm sections, which were mounted on glass slides using Mayer's egg albumin solution, and deparaffinised using HistoClear™ and EtOH. Slides were stained using 0.05% toluidine blue, and inspected using a Nikon Eclipse (Ni-E) upright motorised microscope (Nikon, Japan). Photographs were acquired using Nikon Imaging Software (NIS) Elements Basic Research (v4.2), and contrast adjusted using Adobe Photoshop Lightroom CC.

RNA extraction for RNA-Seq

Floral buds (0 DAP), flowers at anthesis (1 DAP), and immature fruits at their full length (30 DAP, prior to the onset of yellowing and drying) were harvested from second-order branches of plants that grew undisturbed (IND) or from the main branch of plants where side branches were constantly removed during development (DEH), as previously described by (Lenser *et al.*, 2018). Total RNA was isolated from 50 mg of bud, flower, and fruit tissue using QIAzol Lysis Reagent (Qiagen, Hilden, Germany). Genomic DNA was removed by DNase I (Roche, Manheim, Germany) digestion in solution, followed by RNA purification using RNeasy Mini spin columns (QIAGEN, Hilden, Germany). RNA quantity and purity were determined using a NanoDrop™ spectrophotometer (ND-1000, ThermoScientific™, Delaware, USA) and Agilent 2100 Bioanalyzer with the RNA 6000 Nano Kit (Agilent Technologies, CA, USA) using the 2100 Expert Software to calculate RNA Integrity Number (RIN) values. Four biological replicate RNA samples were used for downstream applications. Sequencing was performed at the Vienna BioCenter Core Facilities (VBCF) Next Generation Sequencing Unit, Vienna, Austria (*www.vbcf.ac.at*). Libraries were sequenced in 50-bp single-end mode on Illumina® HiSeq 2000 Analyzers using the manufacturer's standard cluster generation and sequencing protocols.

RNA-Seq, data trimming, filtering, and analysis

The cDNA sequence libraries were processed, including data trimming, filtering, read mapping and feature counting, as previously described (Wilhelmsson *et al.*, 2019). Raw RNA-Seq reads were quality control checked (FastQC), processed to remove adapters and low-quality bases (Trimmomatic, PrinSeq), and cleaned reads mapped (GSNAP) to the *Aethionema arabicum* genome v2.5 (Haudry *et al.*, 2013). After normalisation, genome-mapped reads were compared at each developmental stage using *R* (R Core Team, 2013) and the Bioconductor package DESeq2 (Love *et al.*, 2014, Huber *et al.*, 2015), to identify differentially expressed genes using an adjusted *P*-value (FDR) cut-off for optimising the independent filtering set to 0.05 (Data S1). Principal Component Analyses (PCAs) and clustered heat-maps were created using a custom script and the pheatmap (Kolde, 2015) package in *R*. Indehiscent samples were compared against dehiscent samples (baseline) in all comparisons.

Gene Ontology (GO) term, promotor motif analyses and gene orthologs

Transcripts of *Aethionema arabicum* genome (v2.5) were annotated with GO terms as previously described (Wilhelmsson *et al.*, 2019). GO term enrichment was analysed with the topGO Bioconductor package, using the classic method and fisher test. Enriched promoter motifs from the *Arabidopsis* DNA affinity purification motif database were identified using Analysis of Motif Enrichment (AME) in MEME Suite 5.0.4 (McLeay and Bailey, 2010), with DEG promoter sequences

used as the primary sequences and all promoter sequences used as the control sequences (average odds score with fisher's exact test). Promoter sequences were defined as –1000 and +100 base pairs from the transcription start site based on *Ae. arabicum* genome (v2.5) mRNA annotation (Nguyen *et al.*, 2019). Orthologs of *A. thaliana* genes were identified in *Ae. arabicum* by searching query sequences with BLASTP (Altschul *et al.*, 1990) against a plant-specific protein database. To detect homologous sequences, results were filtered for adequate query coverage and amino acid similarity (Rost, 1999). Sequence data from *Ae. arabicum* are available in the CoGe database (*https://genomevolution.org/coge/*) under the following genome ID: v2.5, id33968.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., Dewar, K., Stinchcombe, J.R., Schoen, D.J., Wang, X.W., Schmutz, J., Town, C.D., Edger, P.P., Pires, J.C., Schumaker, K.S., Jarvis, D.E., Mandakova, T., Lysak, M.A., van den Bergh, E., Schranz, M.E., Harrison, P.M., Moses, A.M., Bureau, T.E., Wright, S.I. and Blanchette, M. (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet*, **45**, 891-U228.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.D., Irizarry, R.A., Lawrence, M., Love, M.I., MacDonald, J., Obenchain, V., Oles, A.K., Pages, H., Reyes, A., Shannon, P., Smyth, G.K., Tenenbaum, D., Waldron, L. and Morgan, M. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, **12**, 115-121.

Kolde, R. (2015) pheatmap: Pretty Heatmaps. *R packsge version 1.0.12*.

Lenser, T., Tarkowska, D., Novak, O., Wilhelmsson, P.K.I., Bennett, T., Rensing, S.A., Strnad, M. and Theissen, G. (2018) When the BRANCHED network bears fruit: How carpic dominance causes fruit dimorphism in *Aethionema*. *Plant J*, **94**, 352-371.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**.

McLeay, R.C. and Bailey, T.L. (2010) Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *Bmc Bioinformatics*, **11**, 165.

Nguyen, T.P., Muhlich, C., Mohammadin, S., van den Bergh, E., Platts, A.E., Haas, F.B., Rensing, S.A. and Schranz, M.E. (2019) Genome improvement and genetic map construction for *Aethionema arabicum*, the first divergent branch in the Brassicaceae family. *G3-Genes Genom Genet*, **9**, 3521-3530.

Priya, R. and Siva, R. (2014) Phylogenetic analysis and evolutionary studies of plant carotenoid cleavage dioxygenase gene. *Gene*, **548**, 223-233.

Priya, R. and Siva, R. (2015) Analysis of phylogenetic and functional diverge in plant nine-cis epoxycarotenoid dioxygenase gene family. *Journal of Plant Research*, **128**, 519-534.

Rost, B. (1999) Twilight zone of protein sequence alignments. *Protein Eng*, **12**, 85-94.

Wilhelmsson, P.K.I., Chandler, J.O., Fernandez-Pozo, N., Graeber, K., Ullrich, K.K., Arshad, W., Khan, S., Hofberger, J., Buchta, K., Edger, P.P., Pires, C., Schranz, M.E., Leubner-Metzger, G. and Rensing, S.A. (2019) Usability of reference-free transcriptome assemblies for detection of differential expression: a case study on *Aethionema arabicum* dimorphic seeds. *BMC genomics*, **20**, ARTN 95.

Supporting Figures

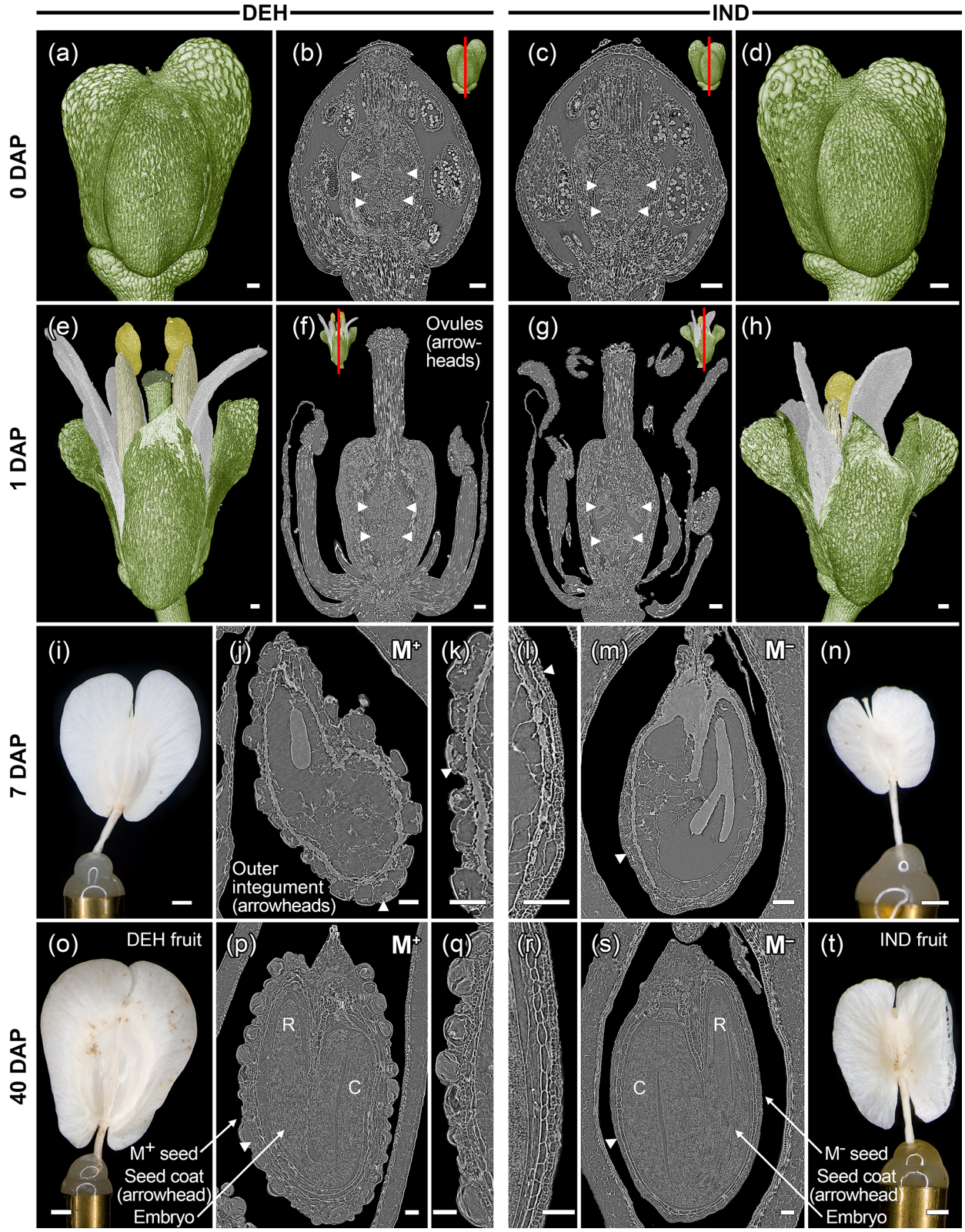


Figure S1. SRXTM imaging of *Aethionema arabicum* fruit morph development.

Comparative Synchrotron Radiation X-ray Tomographic Microscopy (SRXTM) results obtained during diaspore development in the dehiscent (DEH) and indehiscent (IND) fruits of *Ae. arabicum*,

with 3D surface (isosurface) representations and longitudinal slices through tomographic volumes depicting seed and coat development at the respective stage. (a-d) Floral buds, showing morphologically-distinct gynoecium (consisting of two congenitally fused carpels), medial and short stamens. Four anatropous ovules (arrowheads) are present in buds from both morphs. (e-h) Anthesis, the time at which the flower opens and self-pollinates. Ovules (arrowheads) are completely developed and anthers extend above the top of the stigma. Both DEH (f) and IND (g) morphs exhibit differentiated internal structures of an identical nature, with ovules possessing the same nucellus size, number of integument layers, and ovule shape despite subsequently developing into different fruit morphs. (i–n) Immature fruits (i, n) containing immature seeds (j, m) (ca. 7 DAP) from the DEH (i, j, k) and IND (l, m, n) morphs. Details of outer and inner integuments (k, l,) show considerable morph-specific differences (arrowheads), namely in the development and secretion of mucilage between the outer primary wall and the protoplast (in a ring around the area where starch granules are located). (o-t) Mature fruits (o, t) (~40 DAP) containing mature seeds from the the DEH (p, q) and IND (r, s) morphs. The outermost epidermal layer of the seeds forms large mucilage secretory cells in the case of mucilaginous seeds (M^+) from the DEH fruit morph (q), but only a thin layer of mucilage in the single non-mucilaginous (M^-) seed (r) of the IND fruit. Scale bars = 75 μm (a-h, j-m, p-s) or 1 mm (i, n, o, t). Abbreviations: M^+ = mucilaginous; M^- = non-mucilaginous; R = radicle; C = cotyledon.

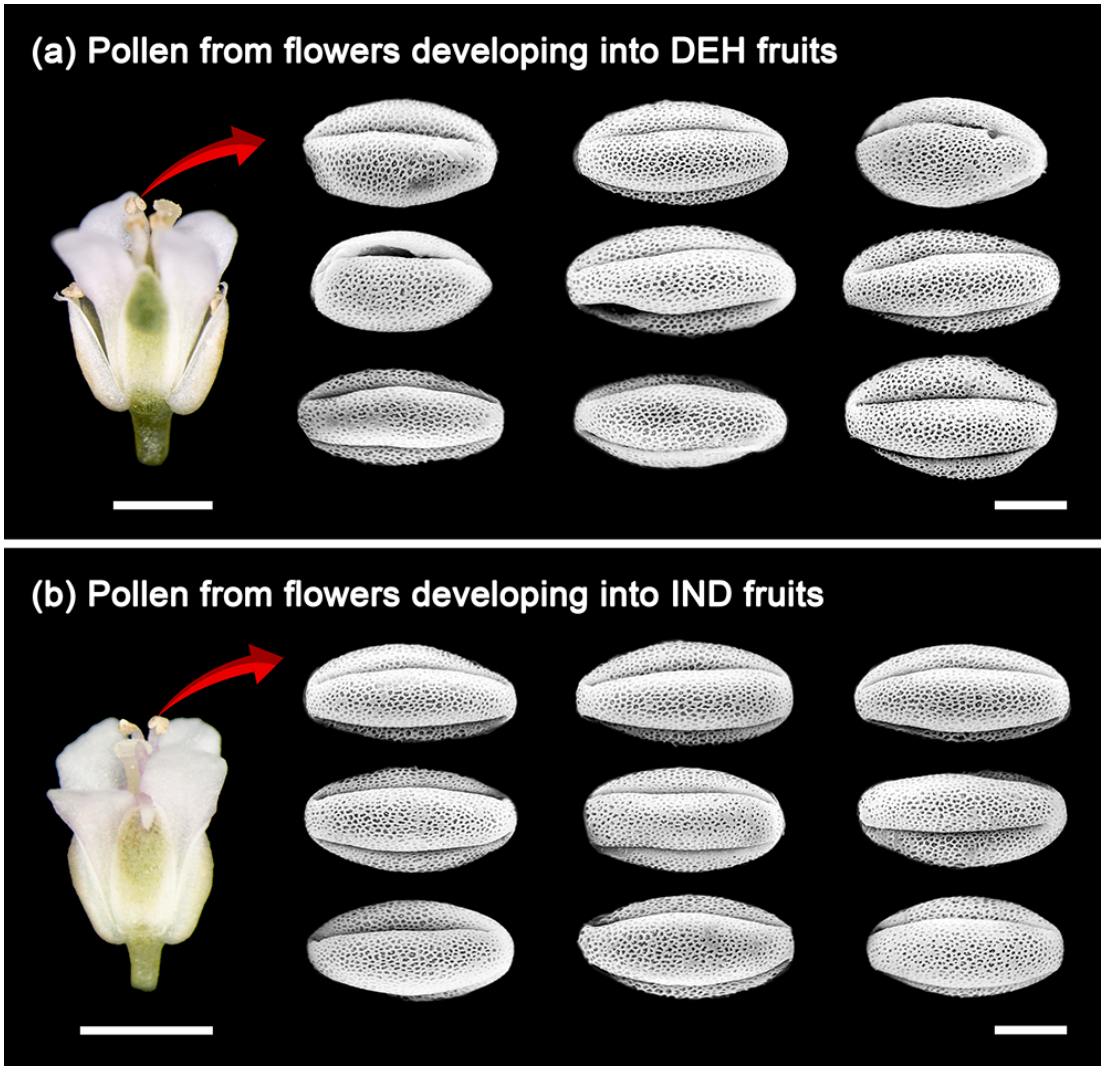


Figure S2. Comparative Scanning Electron Microscopy (SEM) analysis of pollen grains from *Aethionema arabicum* flowers at anthesis. Comparative morphology of pollen from flowers developing into **(a)** dehiscent (DEH) fruits and into **(b)** indehiscent (IND) fruits. Shown is the range of variation of individual pollen grains in equatorial views, all of which exhibit a largely isopolar, prolate-perprolate shape (polar axis : equatorial diameter ratio). All pollen grains are tricolpate, have reticulate exine sculpturing, and do not differ between morphs. Scale bars = 1 mm (flowers) and 10 µm (pollen).

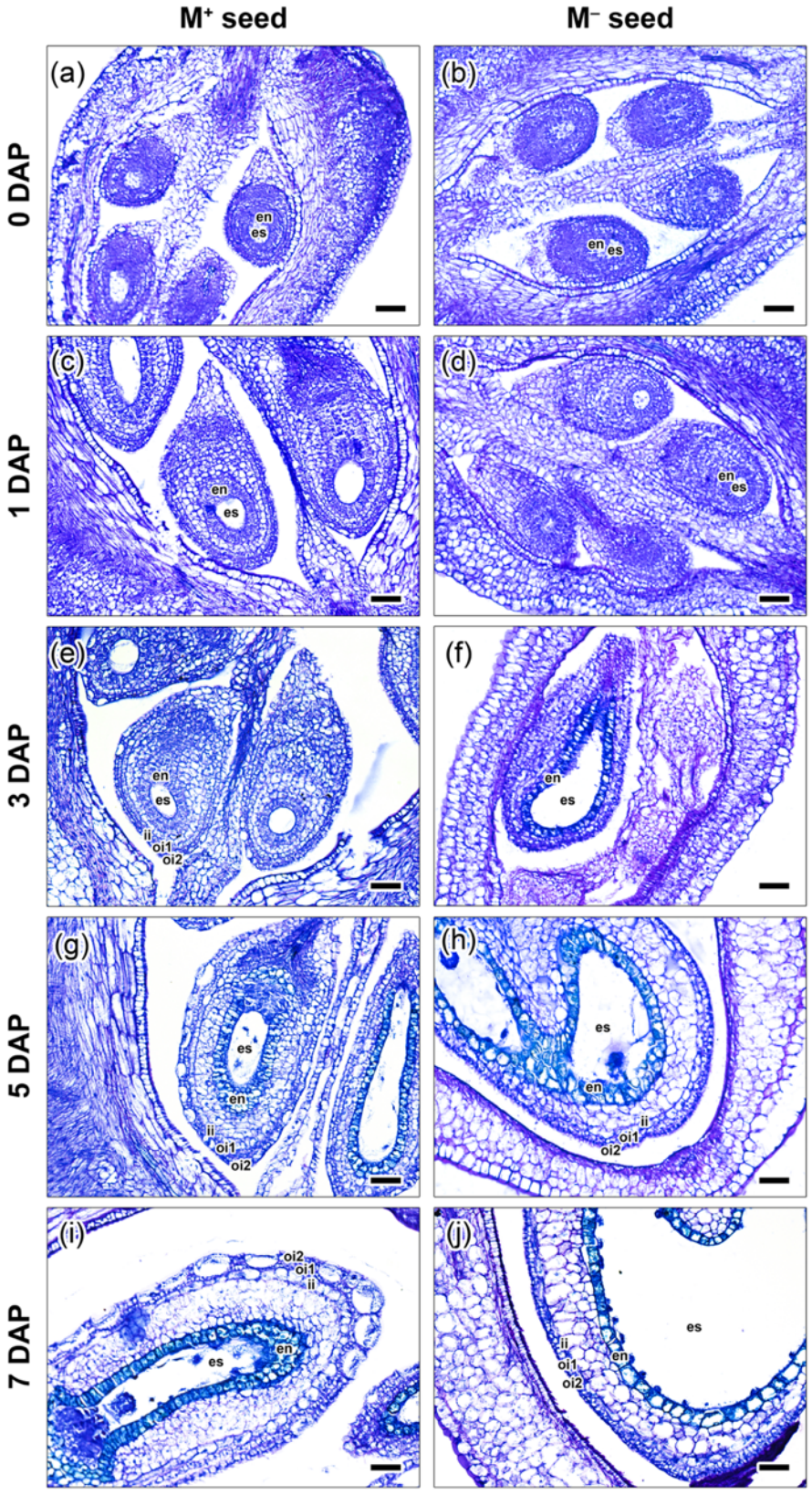


Figure S3. Comparative analysis of morph-specific ovule anatomy during early stages of *Aethionema arabicum* fruit growth after pollination.

Shown are 6 µm thick longitudinal sections through ovules stained with toluidine blue at 0, 1, 3, 5 and 7 days after pollination (DAP). Seeds within dehiscent fruits develop mucilaginous (M^+) seed coats (**a**, **c**, **e**, **g**, **i**), while the single seed within indehiscent fruits (**b**, **d**, **f**, **h**, **j**) develops a non-mucilaginous (M^-) seed coat. Both morphs start with 4 to 6 ovules, and IND fruits undergo an abortive process at 2 to 3 DAP, where programmed cell death of three ovules (**f**) and septum leads to the development of only a single seed. Scale bars = 50 µm. Abbreviations: **en** = endothelium; **es** = embryo sac; **ii** = inner integument; **oi1** = inner epidermis of the outer integument; **oi2** = outer epidermis of the outer integument. Note that unlike *A. thaliana* (the mature ovule of which is amphitropous), the mature ovule in *Ae. arabicum* appeared anatropous; the micropyle is bent towards the funicle to which the body of the ovule is united.

AME
Analysis of Motif Enrichment

For further information on how to interpret these results please access http://alternate.meme-suite.org//doc/ame-output-format.html.
To get a copy of the MEME software please access http://alternate.meme-suite.org.

If you use AME in your research, please cite the following paper:
Robert McLeay and Timothy L. Bailey, "Motif Enrichment Analysis: A unified framework and method evaluation", *BMC Bioinformatics*, **11**:165, 2010, doi:10.1186/1471-2105-11-165. [full text]

ENRICHED MOTIFS | INPUT FILES | PROGRAM INFORMATION | RESULTS IN TSV FORMAT | POSITIVE SEQUENCES FOR EACH MOTIF

ENRICHED MOTIFS

Sequence motif score: avg_odds
Background model source: letter frequencies in (primary) sequences file (FlrINDup.fasta or FlrINDdown.fasta)
Background model frequencies: 0.344464,0.155536,0.155536,0.344464 (FlrINDup)
Background model frequencies: 0.338774,0.161226,0.161226,0.338774 (FlrINDdown)
Total pseudocount added to a motif column: 0.1

Statistical test: Fisher's exact test (optimized over motif scores)
Labeling positives: all 5116 primary sequences are labeled as 'positive'; all 23333 control sequences are labeled as 'negative' (FlrINDup)
Labeling positives: all 5362 primary sequences are labeled as 'positive'; all 23333 control sequences are labeled as 'negative' (FlrINDdown)
Classifying positives: sequences with the motif scores ≥ 'TP Thresh' are classified as positive

E-value threshold for reporting results: 10

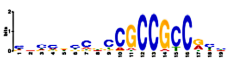
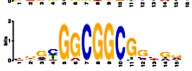
Logo	Database	ID	Alt ID	*p*-value	*E*-value	TP Thresh	TP (%)	FP (%)
bZIP TF binding motifs								
	ArabidopsisDAPv1	bZIP_tnt.ABI5_colamp_v3b_m1	ABI5	1.78e-11	1.55e-8	1.20	843 (16.5%)	2877 (12.3%)
	ArabidopsisDAPv1	bZIP_tnt.ABI5_col_v3h_m1	ABI5	2.19e-11	1.91e-8	1.25	839 (16.4%)	2864 (12.3%)
	ArabidopsisDAPv1	bZIP_tnt.GBF3_colamp_m1	GBF3	1.47e-10	1.28e-7	17.67	721 (14.1%)	2411 (10.3%)
	ArabidopsisDAPv1	bZIP_tnt.GBF3_col_m1	GBF3	1.87e-9	1.63e-6	1.56	813 (15.9%)	2826 (12.1%)
NAC TF binding motifs								
	ArabidopsisDAPv1	NAC_tnt.NAP_col_v3a_m1	NAP	6.98e-10	6.09e-7	1.98	711 (13.9%)	2394 (10.3%)
WRKY TF binding motifs								
	ArabidopsisDAPv1	WRKY_tnt.WRKY28_col_a_m1	WRKY28	3.23e-5	2.81e-2	1.03	1416 (27.7%)	5556 (23.8%)
	ArabidopsisDAPv1	WRKY_tnt.WRKY28_colamp_a_m1	WRKY28	2.71e-3	2.37e0	1.11	805 (15.7%)	3063 (13.1%)
	ArabidopsisDAPv1	WRKY_tnt.WRKY40_colamp_a_m1	WRKY40	9.31e-5	8.12e-2	3.78	742 (14.5%)	2714 (11.6%)
	ArabidopsisDAPv1	WRKY_tnt.WRKY40_col_m1	WRKY40	2.50e-3	2.18e0	1.82	1085 (21.2%)	4240 (18.2%)
Homeobox (HB) TF binding motifs								
	ArabidopsisDAPv1	HB_tnt.ATHB40_col_a_m1	ATHB40	2.40e-3	2.10e0	1.67	538 (10.5%)	1948 (8.3%)
	ArabidopsisDAPv1	HB_tnt.ATHB21_colamp_a_m1	ATHB21	9.94e-3	8.67e0	1.37	711 (13.9%)	2696 (11.6%)
AP2/EREBP TF binding motifs								
	ArabidopsisDAPv1	AP2EREBP_tnt.RAP26_colamp_a_d1	RAP26	2.19e-21	1.91e-18	1.62	2142 (39.9%)	7571 (32.4%)
	ArabidopsisDAPv1	AP2EREBP_tnt.RAP26_col_a_m1	RAP26	1.61e-18	1.40e-15	1.26	1664 (31.0%)	5722 (24.5%)
	ArabidopsisDAPv1	AP2EREBP_tnt.ERF5_colamp_a_m1	ERF5	2.15e-18	1.88e-15	1.53	1451 (27.1%)	4876 (20.9%)
	ArabidopsisDAPv1	AP2EREBP_tnt.ERF5_col_a_m1	ERF5	2.58e-15	2.25e-12	1.28	1800 (33.6%)	6394 (27.4%)
	ArabidopsisDAPv1	AP2EREBP_tnt.ERF11_col_b_m1	ERF11	1.66e-17	1.45e-14	1.12	1771 (33.0%)	6192 (26.5%)
	ArabidopsisDAPv1	AP2EREBP_tnt.ERF11_colamp_a_m1	ERF11	2.77e-16	2.41e-13	2.13	1353 (25.2%)	4558 (19.5%)
	ArabidopsisDAPv1	AP2EREBP_tnt.ERF104_col_a_m1	ERF104	7.51e-17	6.55e-14	1.81	1578 (29.4%)	5435 (23.3%)
	ArabidopsisDAPv1	AP2EREBP_tnt.ERF104_colamp_a_m1	ERF104	3.47e-15	3.03e-12	1.12	1314 (24.5%)	4444 (19.0%)

Figure S4. Analysis of promoter motif enrichment in the *Aethionema arabicum* datasets. The Analysis of Motif Enrichment (AME) tool of the MEME Suite (http://alternate.meme-suite.org//tools/ame) (McLeay et al 2010) was used and Fisher's exact statistical test (P values < 0.01, E value threshold 10) for the comparison against the ArabidopsisDAPv1 database to identify relatively enriched motifs in our *Ae. arabicum* datasets (Data S3). Examples for flower (1 DAP) datasets are presented which delivered 331 enriched motifs (Data S3).

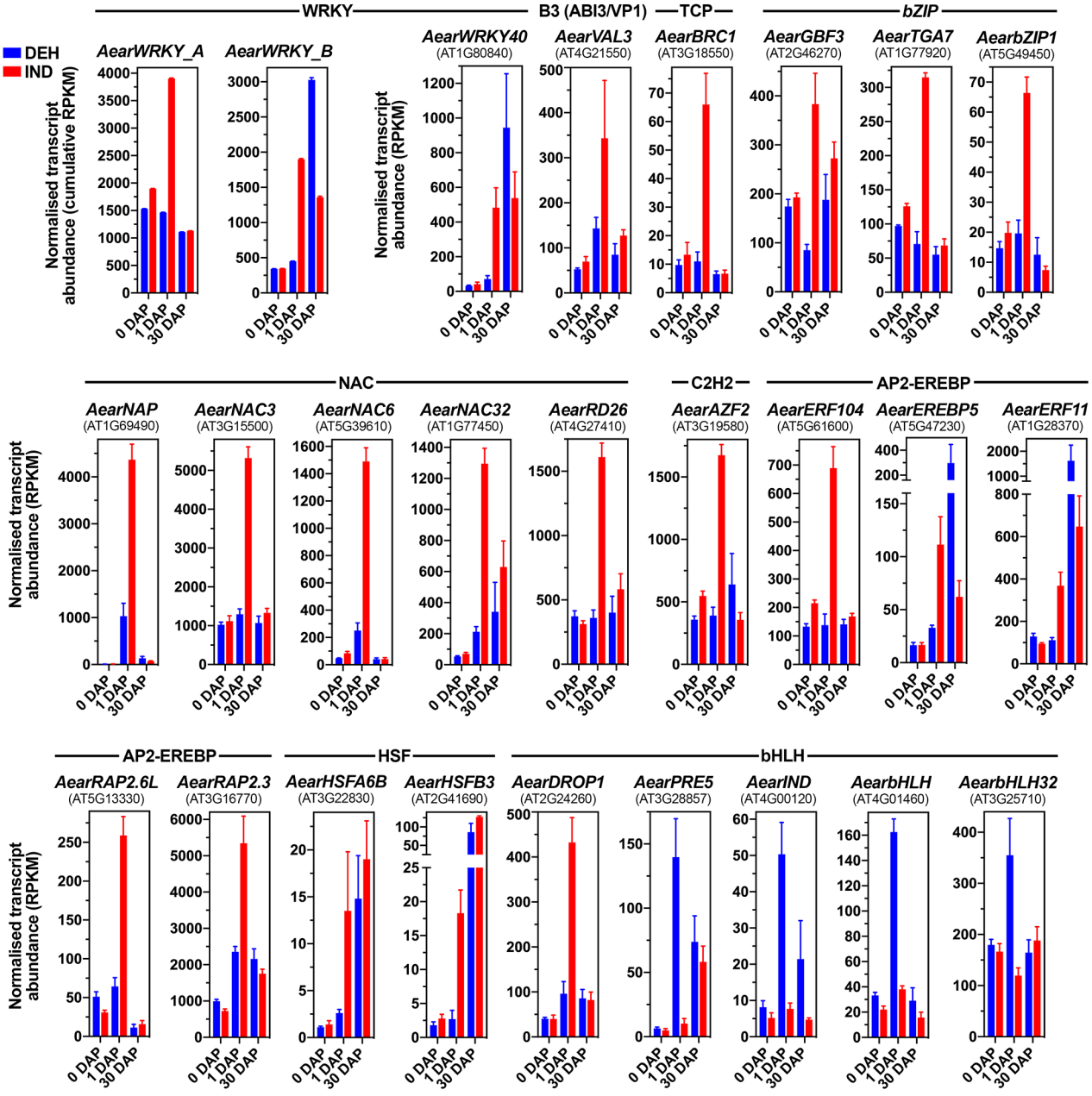
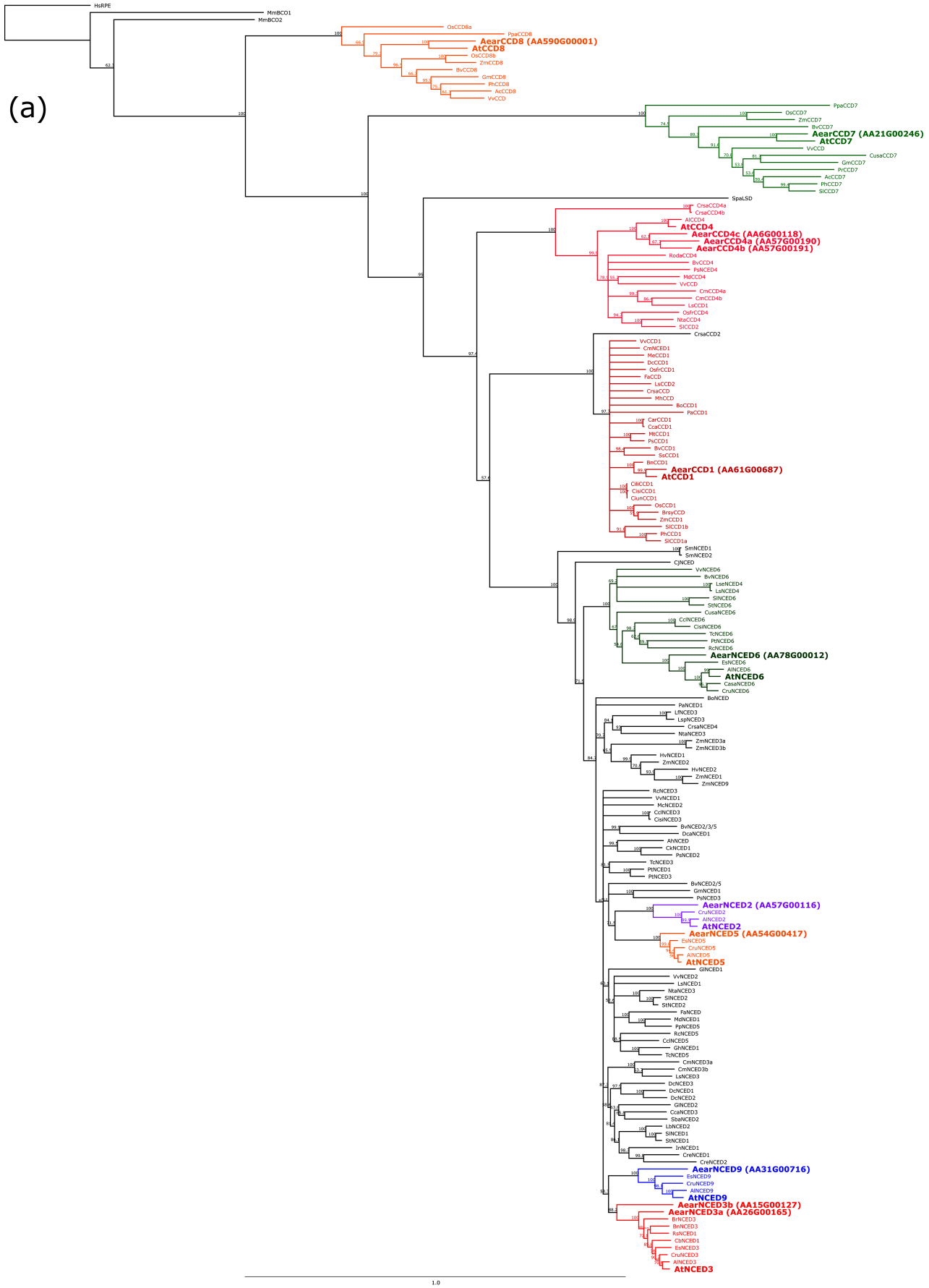


Figure S5. Differentially expressed transcription factor (TF) genes associated with morph-specific anatomical changes in the development of dehiscent (DEH) and indehiscent (IND) *Aethionema arabicum* fruits. Examples for TF DEGs in DEH and IND bud (0 DAP), flower (1 DAP) and fruit (30 DAP) samples. Mean values ± SE from RNA-seq data are presented. For gene names see main text; numbers in brackets refer to the *A. thaliana* orthologous gene.

Arshad et al 2021



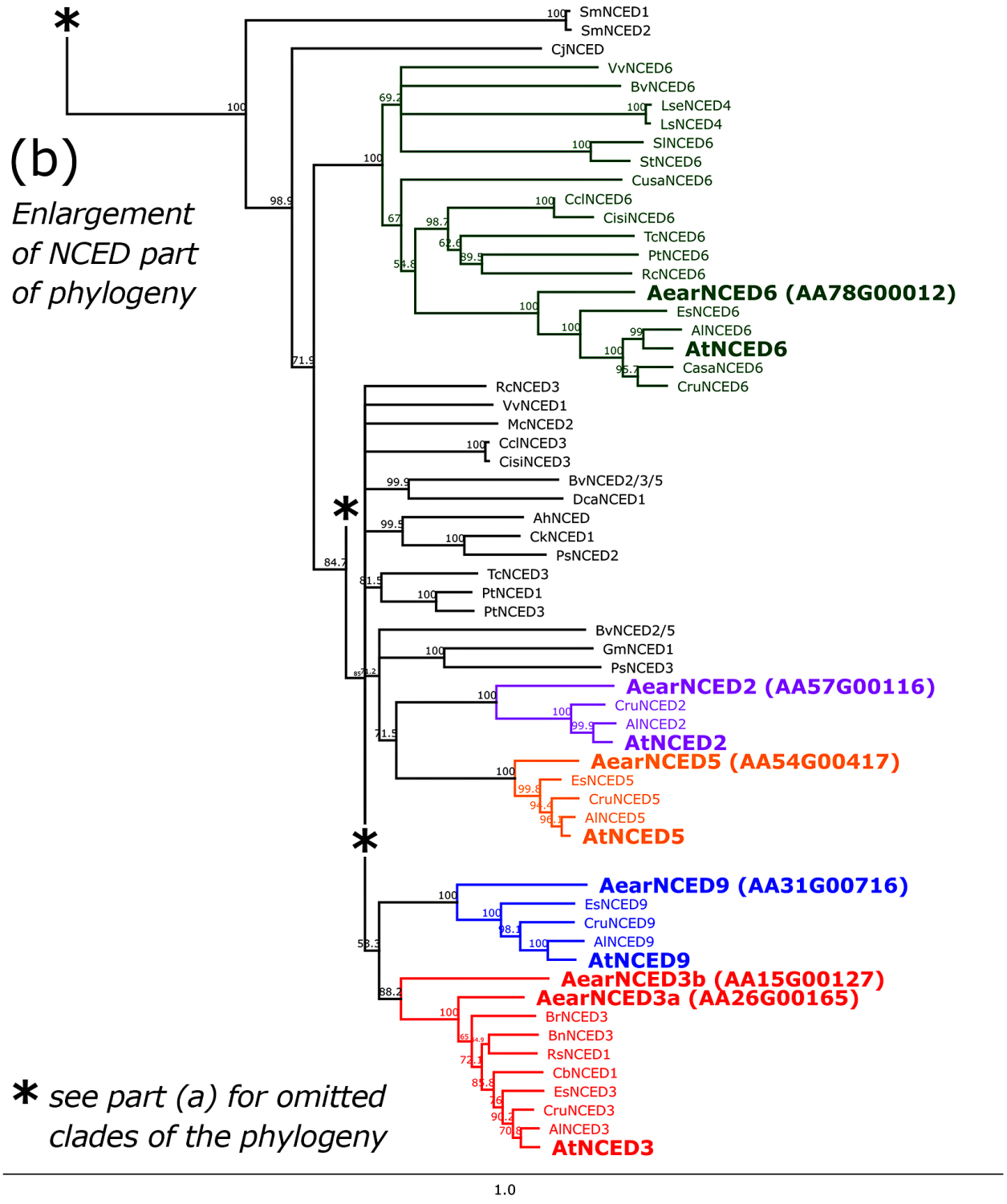


Figure S6. Phylogenetic analysis of *Aethionema arabicum* carotenoid cleavage dioxygenase (CCD) genes. (a) Known and putative CCD amino acid sequences were aligned using ClustalW (BLOSUM cost matrix, Gap open cost 10, Gap extend cost 0.1) and Neighbor-Joining tree was built using Geneious 8.1.9 Tree Builder using Jukes-Cantor distance. Consensus support (mimimum 50%) was determined using bootstrap (1000). *Arabidopsis thaliana* CCD and 9-*cis*-epoxycarotenoid (NCED) sequences were used to identify *Ae. arabicum* orthologs (highlighted in bold). Sequences used are available from NCBI or Uniprot; additional sequences used were from Priya and Siva (2014, 2015). (b) Enlarged view of phylogenetic clades containing *Ae. arabicum NCEDs.*

The Plant Journal - Supporting Information

Supporting Table S1. List of primers used for quantitative RT-PCR analysis.

***Aethionema arabicum* (genome v2.5) ID**	***Gene name***	***Primer direction***	**Primer Sequence (5' – 3')**
AA19G00315	*AA19G00315*	F	TGGTGCACGTGAGCTCTTAG
		R	ATCTTTGGTGGGAGTGCTGG
AA44G00404	*AFP*	F	AGAACGGTGGCTGTAACTGG
		R	GATTCCTTTTCCCGGCATGC
AA10G00283	*CMSP*	F	TTGGTCCGGCTTTGTCTTTG
		R	CACCAAAGTTATCATGGTTTCCCC
AA30G00232	*AP2*	F	AAACGGTGAAAGCGGTTGTG
		R	GAGCTCCGTGATCTTGGACC
AA89G00009	*GL2*	F	ATCCGGACGAGAAGCAAAGG
		R	CTTTCAGCAGCGAGTTCTCG
AA123G00058	*MEEA14*	F	GTGGTTCGGTGTTCAAGCAC
		R	AGTCATCACCTTCCAGCGTC
AA42G00001	*MYB61*	F	GCAGTCCTCGGAAACAGATG
		R	TTGCAGAAGAAGTTGAAGCAGG
AA13G00129	*TTG1*	F	CTTCGTCGCTCATCTACCGG
		R	CAACGGTGCACAGAACTCAC
AA20G00013	*CESA2*	F	TGTGAGATTTGGTTTGCTGTTTC
		R	TGCTAATCCTGATGGCTTCCC
AA4G00210	*MUM3/CESA5*	F	CGAATGGAGGAATGGAAGCG
		R	TCGTGATAACGGTTGCCTCC
AA15G00120	*MUM4/RHM2*	F	CAGAGGGGTCAGGGATTGG
		R	ATGAAGTTTCGCGGGTTCTC
AA21G00495	*PMEI6*	F	TTCGCCATTCTCCTCTCTCG
		R	CGCCGCGTTTTGAGTAAGAG
AA30G00244	*MEEA59*	F	GGTGAGGAGAGTGTAACCGC
		R	CATTTGTCGCCGGATTTCCC
AA118G00053	*ABF3*	F	TGGAAAATCATGCTCCACTTGC
		R	CTTCCGGGAGATACTGCAGC
AA8G00054	*ENY/IDD1*	F	TCCATCTCCGGCGATTTCTG
		R	CTGCGTAGACACTCTCATTGC
AA30G00184	*HB40*	F	TGCACGGTGGAAGAACAAAAG
		R	CCTTCAGCTTCTACTCTCTCTGC
AA8G00057	*HB53*	F	GTGCACGGTGGAAGAACAAG
		R	TCGGTGCATTATTGGCTTCAAC
AA53G01391	*KNAT7*	F	TCACCATCATGACCGTCACG
		R	GCCATAACAGCTTCAACGGC
AA31G00611	*MYB54*	F	TGGCTGGCAAGATTGGTTTC
		R	TGATCCGAATCTTGAGGAGCAC
AA39G00089	*MYB52*	F	GGTCCTCATAATTGGAACGCC
		R	TCGACCGGGGAAAAATCTAGC

AA15G00071	*NARS2*	F	ACCCAGTCATCGATGTGATTTTG
		R	CCTTATCTGTGTGTAATCGAGCTG
AA15G00127	*NCED3b*	F	TCGATGTCCCGGATTGCTTC
		R	TCCCCGCTTCGAGATTAACC
AA26G00165	*NCED3a*	F	AAACCTCATCTTCTCCCGCC
		R	CTCACGAGTAATCCCTCCGC
AA61G00296	*SHP1*	F	AAAGGCATAAGTCGCGTTCG
		R	GCAACTCCATTTCTCGTTTTTGC
AA21G00070	*SHP2*	F	GCTCTCTGTCTTGTGTGATGC
		R	TGGAAGGAGGATTTACAGCGTC
AA3G00145	*STK*	F	GTGGCCGTCTCTATGAATATGC
		R	TGGATCTGTTGCCTAAGTTTTGC