

Genome-Wide Comparative Analysis of Flowering-Related Genes in Arabidopsis, Wheat, and Barley

Peng et. al, 2015



Karlene Negus, Heather Chamberlain-Irwin, Elizabeth McMurchie, Tyler Foster, and Sarah Minkler

Introduction to Paper

- Although early flowering is important to wheat and barley, the genetics behind it are largely unknown
- Peng et al. (2015) used bioinformatics to predict flowering-related genes in wheat and barley from known *Arabidopsis thaliana* flowering genes
 - Identified 900 putative orthologs in wheat and 275 in barley
 - Conducted domain and phylogenetic analyses of proteins related to flowering at different stages of tissue development in wheat and barley
 - Found several orthologs in PEBP, MADS and BBX gene families showed similar expression patterns in wheat and barley
 - Likely related to early flowering
 - These genes could be used to attempt to improve crops



Barley (left) and wheat (right)

Description of Group Workflow

- Communicated via Slack
- Collectively decided on a study and distributed figure responsibilities
- Coordinated with team mates who needed similar data
- Supported each other through the scripting process
- Used pull requests and Slack to communicate repo changes
- Meet periodically to organize and provide updates

Overview of Documentation

- Karlene Negus replicated Table 1, Figure 2, Supp. File 2
- Sarah Minkler replicated Tables 2 and 3
- Heather Chamberlain-Irwin replicated Figures 1a and 1b
- Tyler Foster replicated Figure 3(a-c)
- Elizabeth McMurchie replicated Figure 4 and Supplemental Figure 3

Table 1

- Data from Supp. file 1.xls
- Summarizes Flowering gene functional group distribution across *Arabidopsis* chromosomes
- Plot in R using kableExtra package

Peng et al. (2015)

TABLE 1: Distributions of 204 flowering genes over five chromosomes and seven known functional groups in Arabidopsis compiled through searches in the literature and TAIR.

Gene type	Functional group	AT1	AT2	AT3	AT4	AT5	Total
Protein coding	Autonomous	0	2	1	3	1	7
	Flower development	19	3	7	3	9	41
	Gibberellin	9	2	3	3	2	19
	Pathway integration	2	1	0	2	2	7
	Photoperiod	9	9	3	1	10	32
	Regulation	8	15	16	10	20	69
	Vernalization	1	1	2	3	8	15
	Subtotal	48	33	32	25	52	190
MicroRNA		2	4	0	2	6	14
Total		50	37	32	27	58	204

Replication

Table 1: Distributions of 204 flowering genes over five chromosomes and seven known functional groups in Arabidopsis compiled through searches in the literature and TAIR.

Gene type	Functional Group	AT1	AT2	AT3	AT4	AT5	Total
Protein coding	Autonomous	0	2	1	3	1	7
	Flower development	19	3	7	3	8	40
	Gibberellin	9	2	3	3	2	19
	Pathway integration	2	1	0	2	2	7
	Photoperiod	9	9	3	1	11	33
	Regulation	8	15	16	10	20	69
	Vernalization	1	1	2	3	8	15
	Subtotal	48	33	32	25	52	190
MicroRNA		2	4	0	2	6	14
Total		50	37	32	27	58	204

Table 2

- Data from Supplemental data file 4
- Replication carried out in Python and R
- Published Table 2 on top
- Minor differences in numbers/percentages as compared to authors.

Ensembl annotation status	New annotation	
	Barley	Wheat
Uncharacterized	93.1% (256)	96.2% (866)
Characterized	6.9% (19)	3.8% (34)
Total	275	900

Ensembl annotation status Barley Wheat
+-----+-----+-----+
Uncharacterized 94.2 % (259) 96.2 % (866)
Characterized 5.8 % (16) 3.8 % (34)
Total 275 900

Table 3

- Data from Supplemental data file 4
- Replication carried out in Python and R
- Published Table 3 on top
- There were some differences in the N numbers calculated for each species.
 - Excluded values not reported
 - Methods on Gene length and protein length not reported

	AT (<i>n</i> = 190)		TA (<i>n</i> = 525)		HV (<i>n</i> = 265)	
	Mean	Range	Mean	Range	Mean	Range
Transcripts per gene	1.4	1–5	1.0	1-1	2.8	1–27
Gene length (bp)	3161	182–16871	3815	240–20952	4328	404–15512
Exons per gene	6.5	1–48	5.7	1–42	4.5	1–20
Exon size (bp)	466	79–4165	565	42–5550	878	87–5211
Intron size (bp)	468	78–2316	924	58–7291	856	44–5912
Protein length (aa)	529	77–3529	444	52–3250	500	50–2056

| Structural Characteristics | AT | (n=204) | TA | (n=833) | HV | (n=271) |

+-----+-----+-----+-----+-----+-----+

| *** | Mean | Range | Mean | Range | Mean | Range |

| Transcripts per gene | 1.4 | 1-5 | 1 | 1-1 | 2.8 | 1-27 |

| Gene Length(bp) | 2297 | 79-11046 | 2675 | 158-12689 | 2998 | 153-9980 |

| Exons per gene | 6.5 | 1-48 | 5.7 | 1-42 | 4.5 | 1-20 |

| Exon size(bp) | 466 | 79-4165 | 565 | 42-5550 | 878 | 87-5211 |

| Intron size(bp) | 468 | 78-2316 | 924 | 58-7291 | 856 | 44-5912 |

| Protein length (aa) | 1828 | 79-10872 | 1751 | 106-10363 | 2132 | 153-7133 |

Supplemental File 2

- TAIR website search of genes in Supp. File 1
- Reciprocal BLAST of *Arabidopsis* gene protein sequences on *Triticum aestivum* and *Hordeum vulgare*
- Ortholog grouping using OrthoMCL

OG5_178217 - Original

AT1G71692

MLOC_53973

Traes_2BL_E0978B1BC

Traes_2DL_71F120931

Traes_7AL_67921A952

Traes_7BL_9BCF391CF

Traes_7DL_CAF83263E

OG1.5_1087 - Duplicate

AT1G71692.1

MLOC_53973.1

MLOC_53973.3

Traes_7AL_67921A952.2

Traes_7BL_9BCF391CF.2

Traes_7DL_CAF83263E.2

OG5_146543 - Original

AT1G65480

AT4G20370

AT5G03840

MLOC_68576

Traes_3B_2A454DB62

Traes_3B_C8DBBCD0E

Traes_3DS_7A9CDCC9A

Traes_5AL_9731E2D53

Traes_7AS_EBD5F1F54

Traes_7BS_581AA844D

Traes_7DS_12C14942B

OG1.5_1123 - Duplicate

AT1G65480.1

AT4G20370.1

MLOC_68576.1

Traes_7AS_EBD5F1F54.1

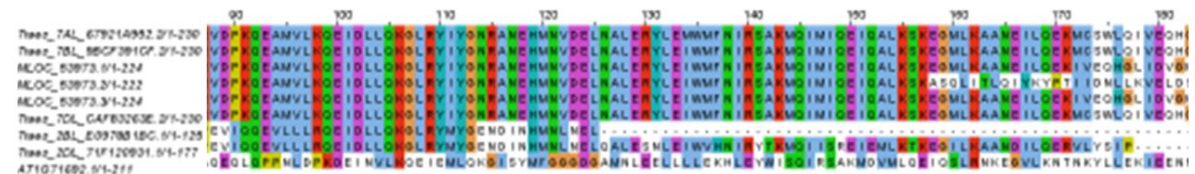
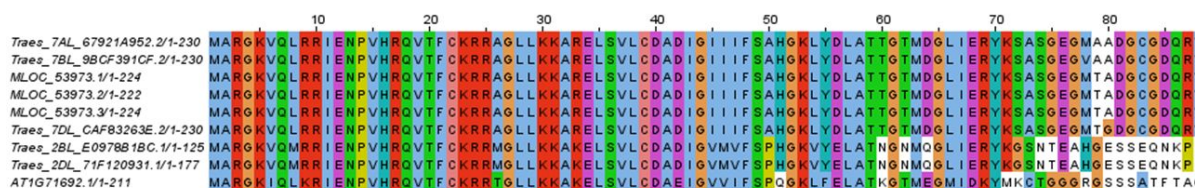
Figure 1a: MADS-box domain and K-box region

OG5_178217 Protein Domain

No parameters given

Differences:

- Included 3 MLOC
- Base instead of “-”
- Jalview newer version, so position is at the top, not the side



Ours



Theirs

Figure 1b: PEBP family proteins

Differences:

- Traes_3B_2A454DB62 unavailable
- Beginning and end of sequences have some genes of different lengths

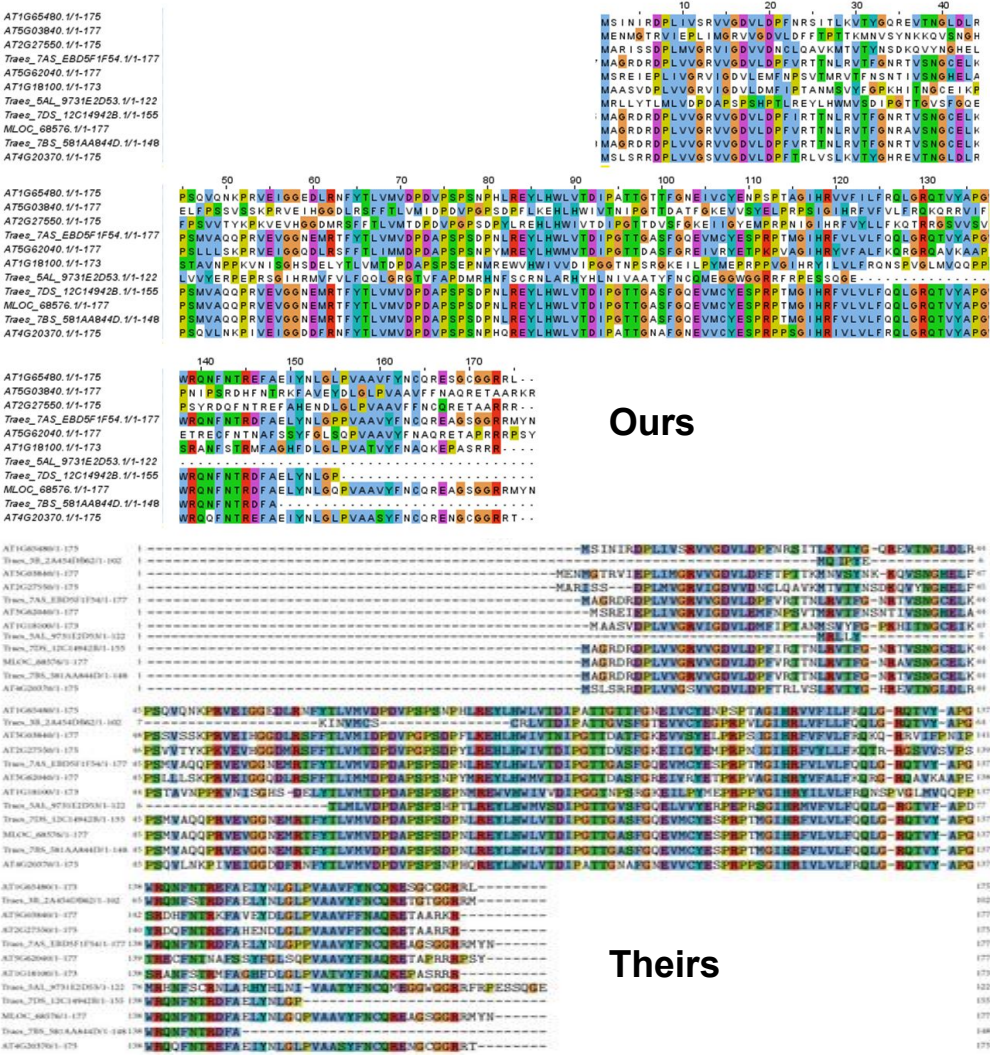


Fig1 Quality Scores

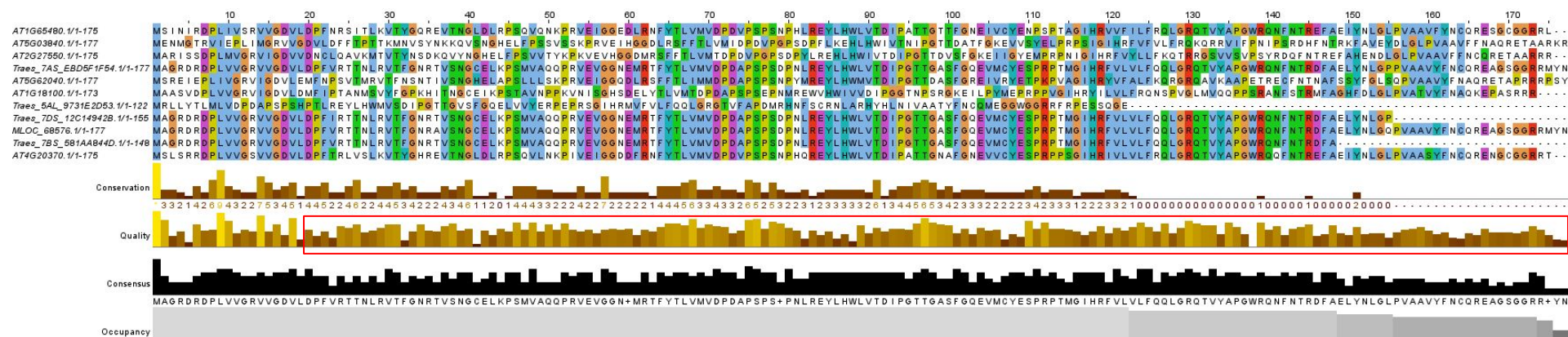
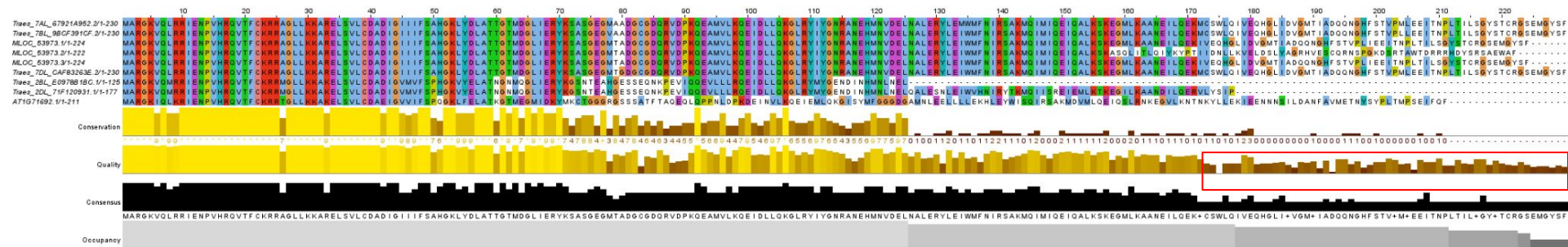


Figure 2

MSA from Figure 1 used for gene phylogenetic trees

- File formatting in BEAUTi
 - GUI interface required
- Tree analysis in BEAST v1.8.2
- Tree annotation in TreeAnnotator
- Tree plotting in R using ggTree package
 - Peng et al. used FigTree which is GUI-based
- Outgroup sequence usage unclear

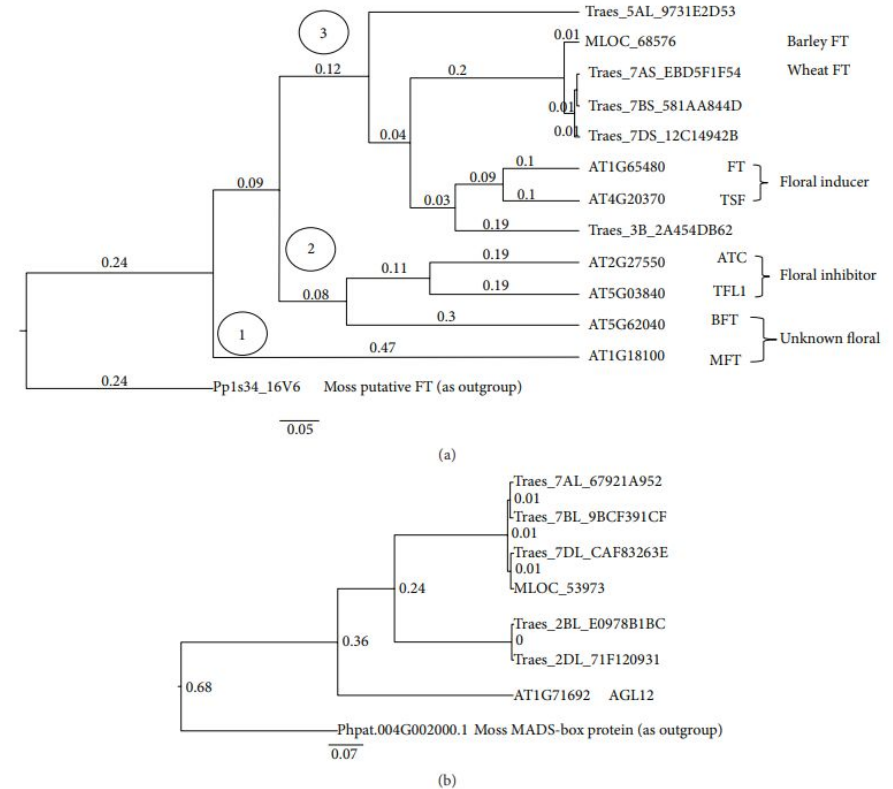
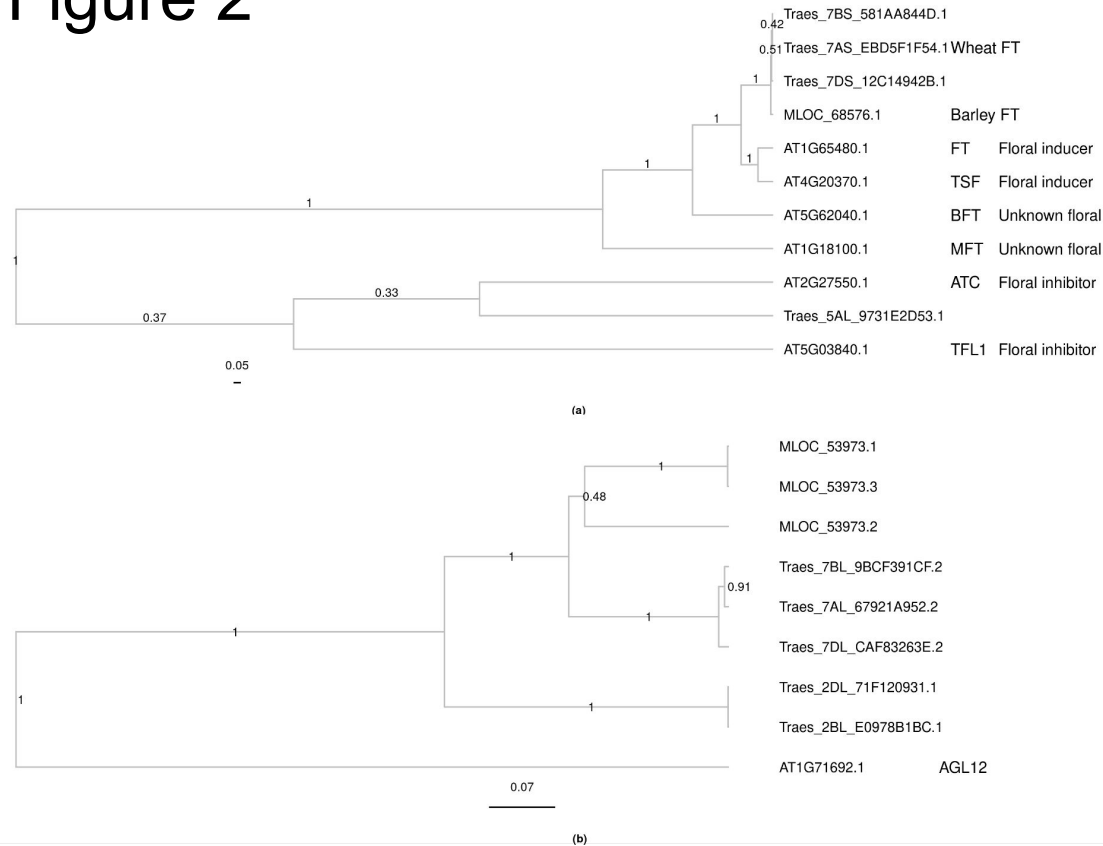


FIGURE 2: Phylogeny of PEBP (a) and MADS (b) family proteins in Arabidopsis, wheat, and barley. The PEBP proteins include 11 sequences in OG5.146543 (see Supplemental file 2), and a PEBP protein Pp1s34_16V6 in moss was used as an outgroup to root the phylogenetic tree. The MADS AGL12 proteins include six sequences in OG5.178217, and a MADS protein Phpat.004G002000.1 in moss was used as an outgroup to root the phylogenetic tree. The support value on each node is the Bayesian posterior probability. The scale bar denotes the number of nucleotide replacements per site.

Replication

Figure 2



Peng et al. (2015)

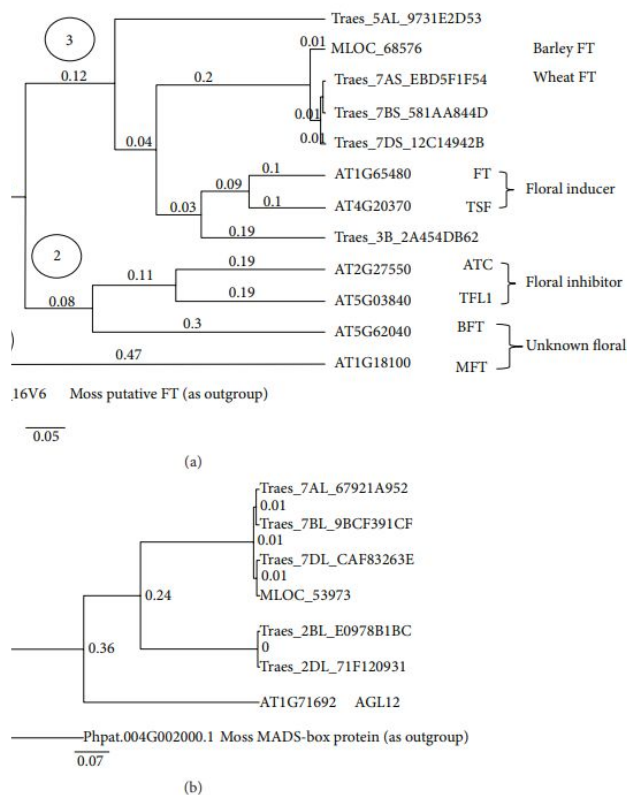


Figure 2: Phylogeny of PEBP (a) and MADS (b) family proteins in Arabidopsis, wheat, and barley. The support value on each node is the posterior probability. The scale bar denotes the number of nucleotide replacements per site

(b) family proteins in Arabidopsis, wheat, and barley. The PEBP proteins include 11 sequences and a PEBP protein Pp1s34.16V6 in moss was used as an outgroup to root the phylogenetic tree. The support values in OG5.178217, and a MADS protein Phpat.004G002000.1 in moss was used as an outgroup. The support value on each node is the Bayesian posterior probability. The scale bar denotes the number of

Figures 3(a-c)

- Expression of flowering genes in different tissues and developmental stages in three species:
 1. Arabidopsis - 63 tissue types
 2. Wheat - 13 tissue types
 3. Barley - 15 tissue types
- Information provided:
 1. Original authors noted they used heatmap.2 function in R, but excluded original code
 2. Comma separated files for each species in supplemental information
 - 'Supplemental file 5.xls'

Figure 3(a) - Arabidopsis

Original

Re-creation

Noticeable Differences

- Key scale
- Key size
- Heat map size
- Heat map text size

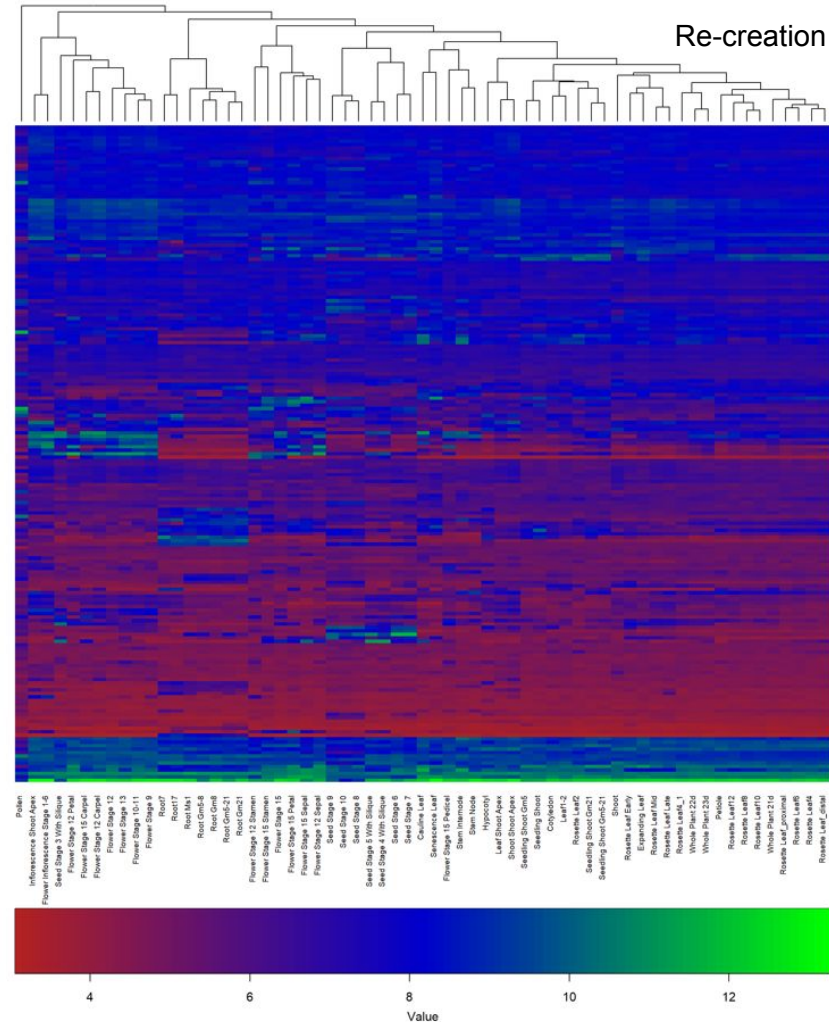
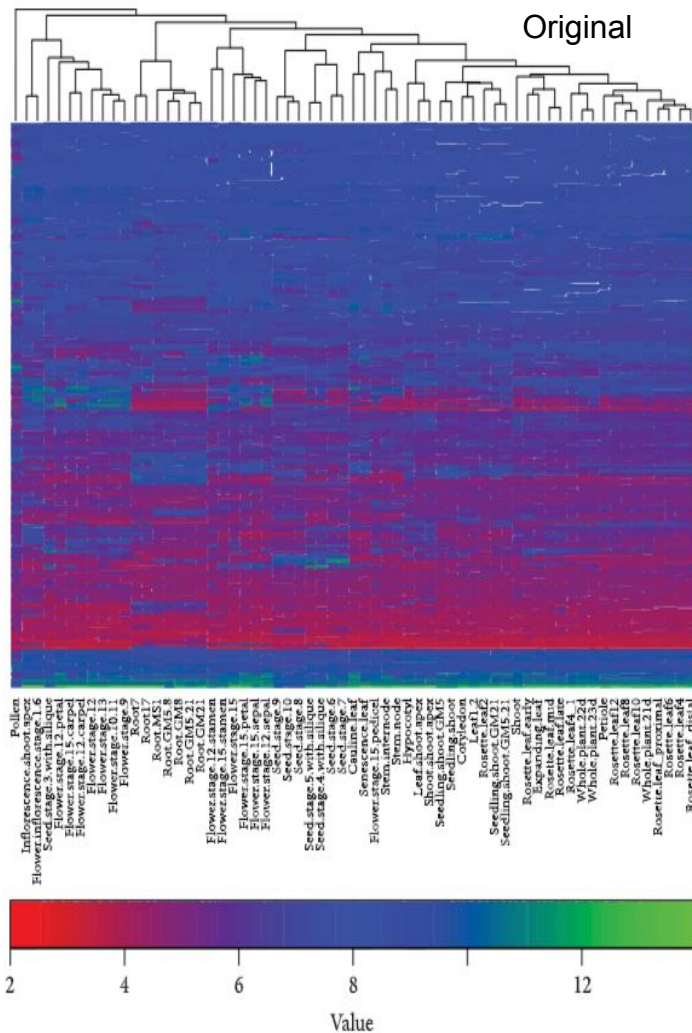


Figure 3(b) -
Wheat

Noticable Differences

- Key scale
- Key size
- Heat map size
- Heat map labeling
- Dendrogram size

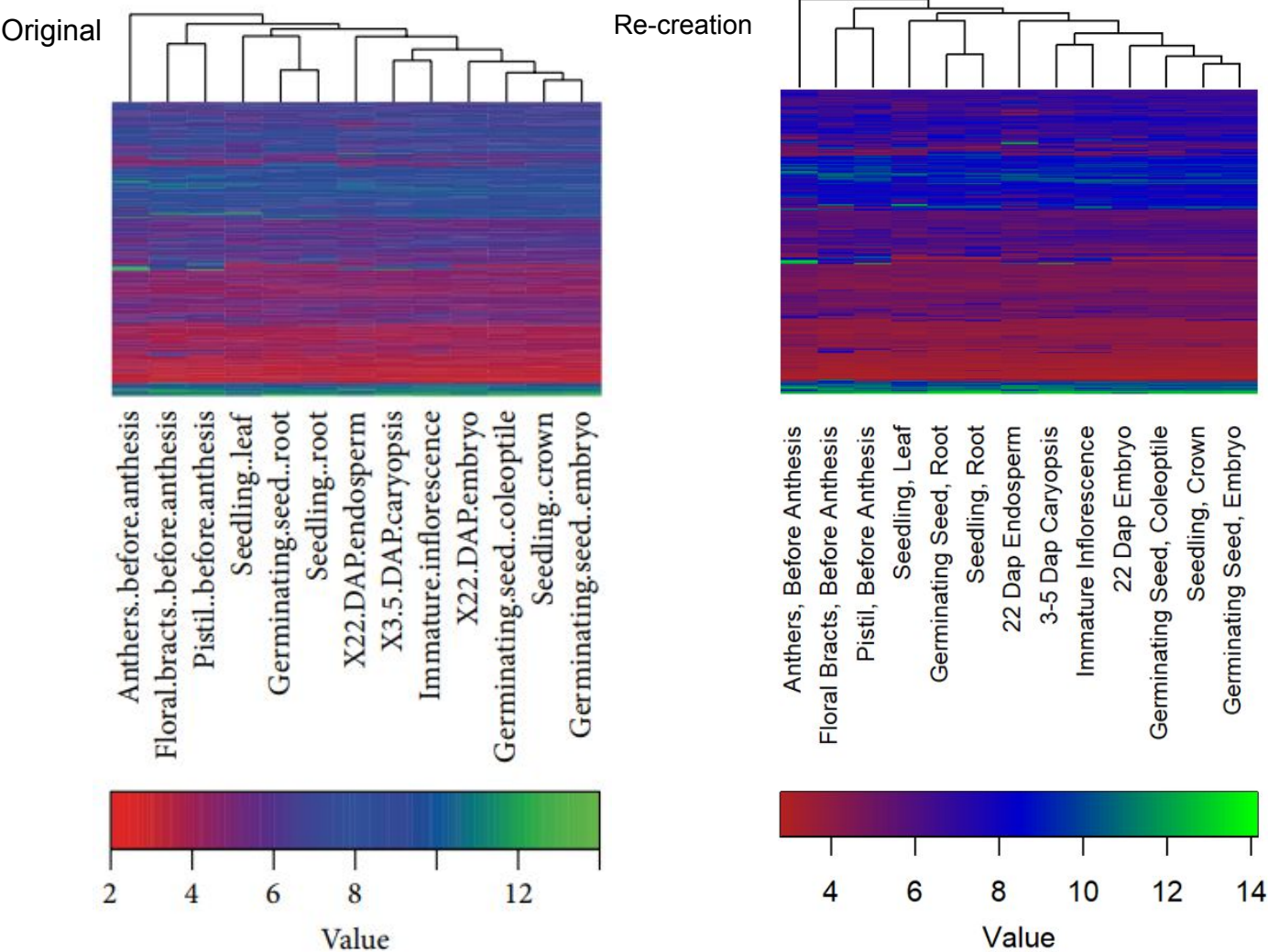
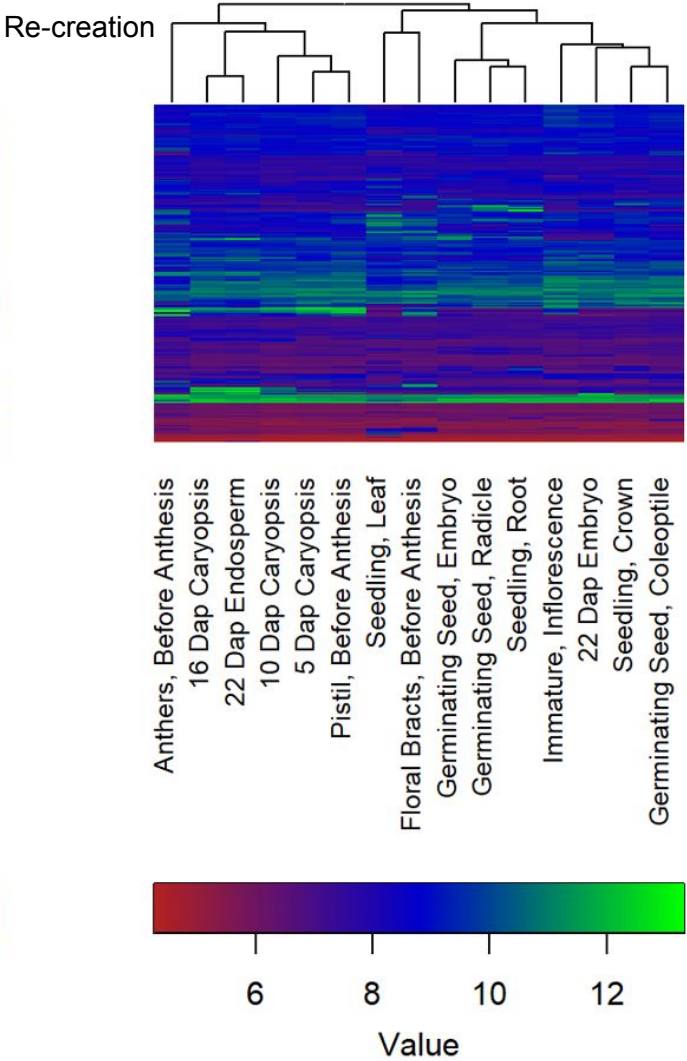
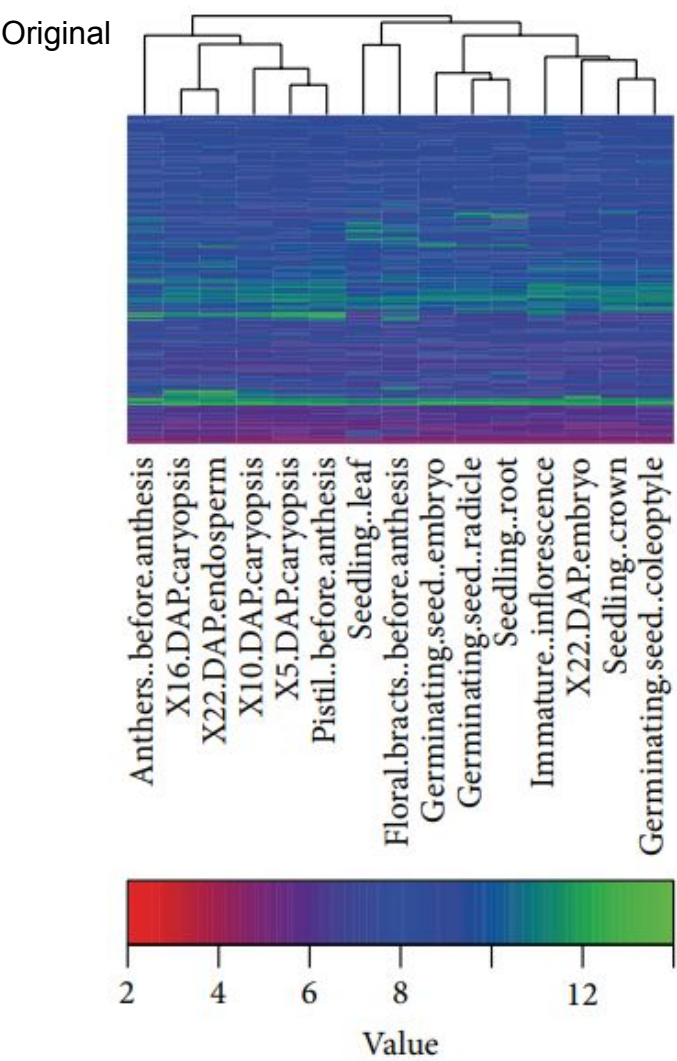


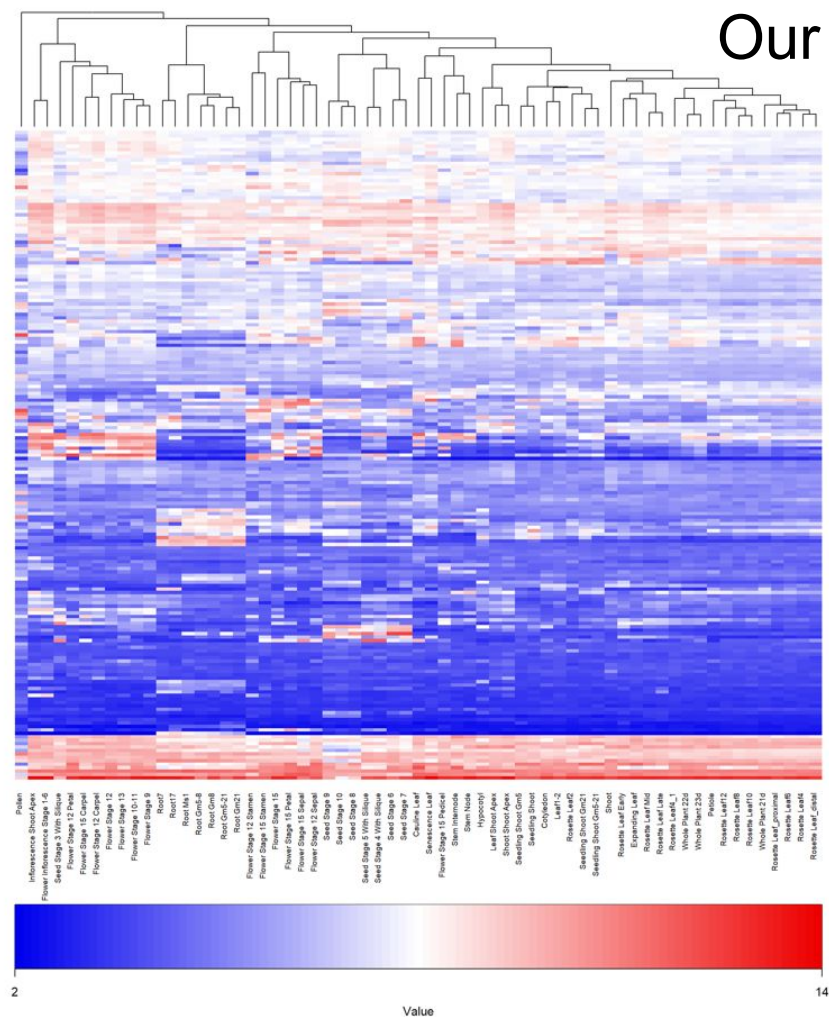
Figure 3(c) -
Barley

Noticable Differences

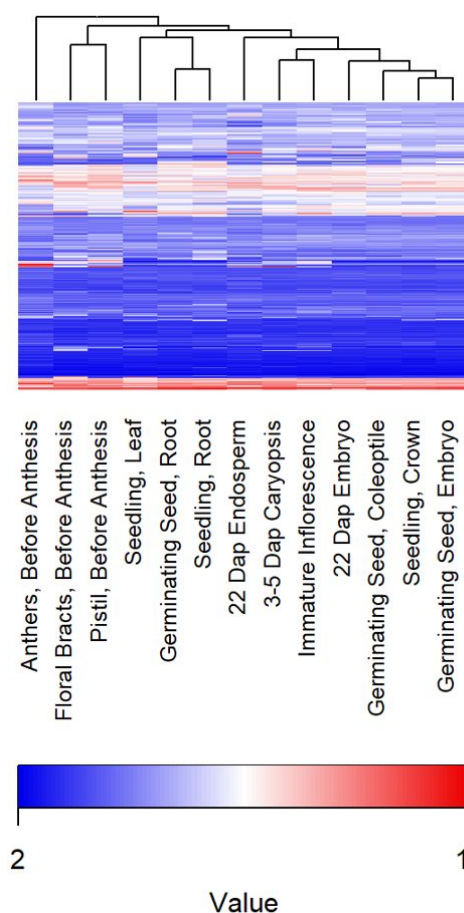
- Key scale
- Key size
- Heat map size
- Heat map labeling
- Dendrogram size



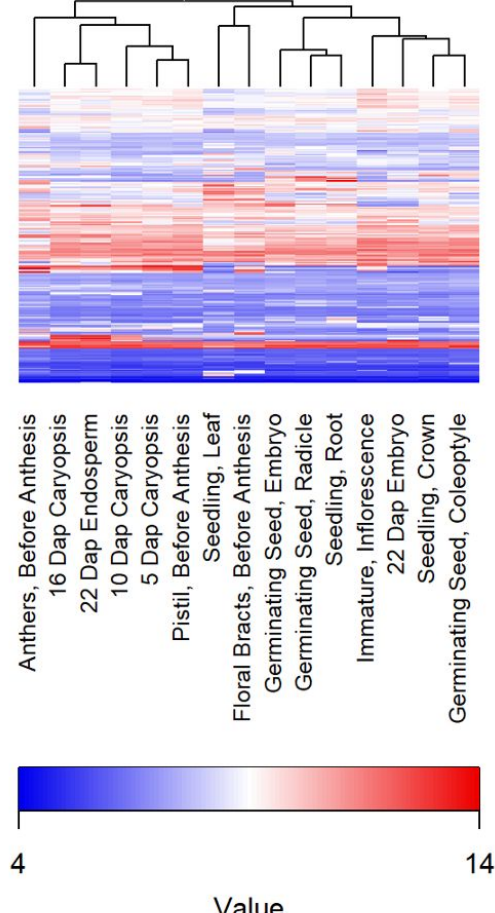
Our “Take” on Figure 3(a-c)



Arabidopsis



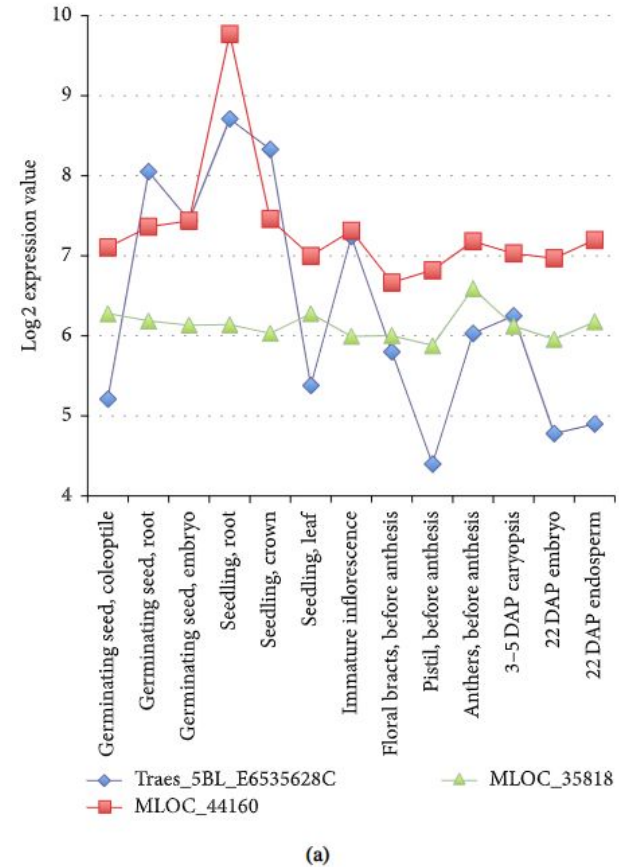
Wheat



Barley

Figure 4

- Figure 4 depicts expression profiles of major ortholog groups in three gene families (PEBP, MADS-box, and B-box) in wheat and barley in different tissues
 - Essentially six figures for (a): PEBP OG5_158796, (b): PEBP OG5_146543, (c): MADS OG5_178217, (d): MADS OG5_144912, (e): BBX OG5_178368, and (f): OG5_170758.



Original Figure 4a. Expression profiles of orthologous genes in the two ortholog groups of PEBP OG5_158796

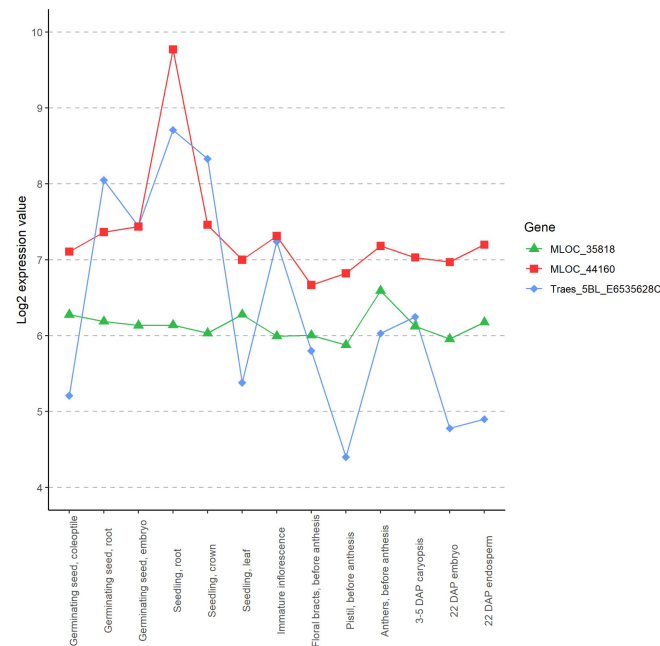
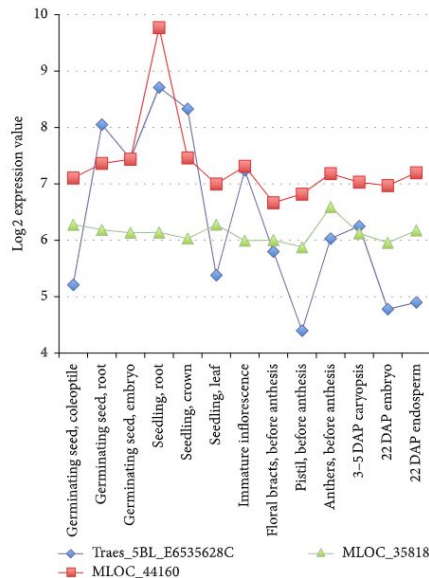
Figure 4

- Apparently originally made in Excel using 'Supplemental file 5.xls'
 - Required considerable data wrangling to use: `tidyverse`, `readxl`, `dplyr`, `data.table`, `janitor`, and `tibble` R libraries required

Side-by-side comparison: Figure 4a from paper (left) and reproduced in R (right)

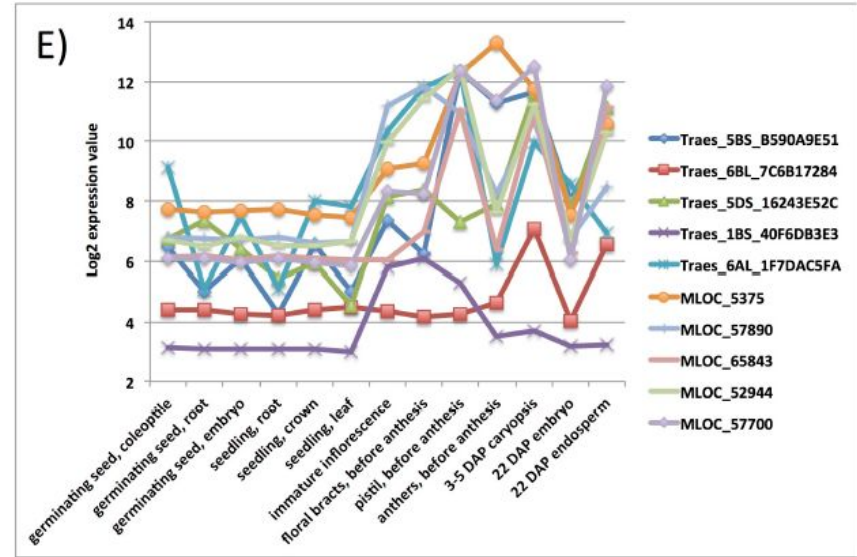
Expression profiles of orthologous genes in the two ortholog groups of PEBP

OG5_158796



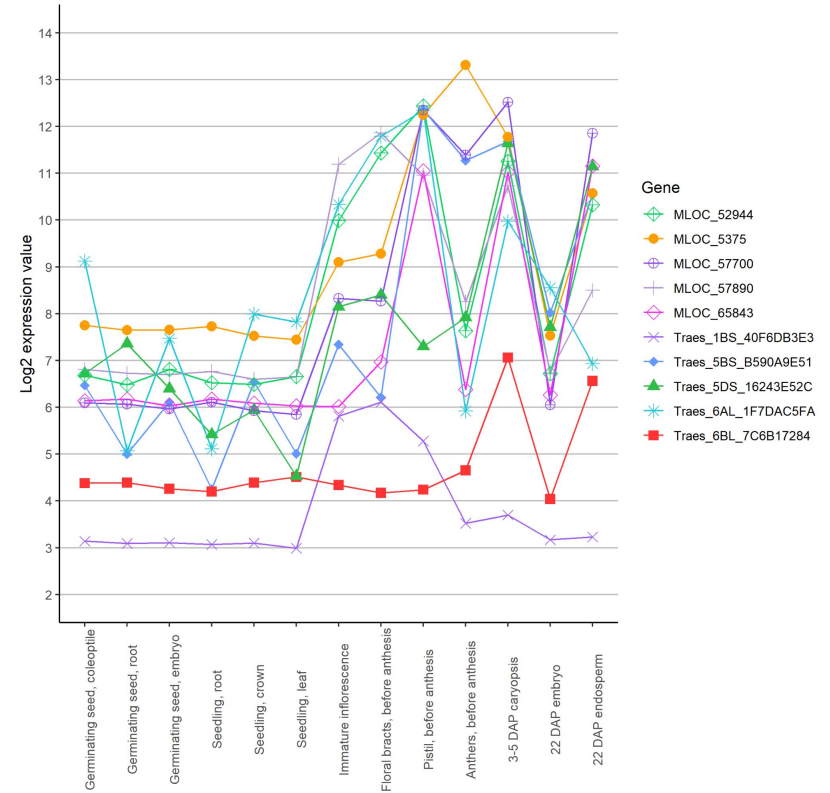
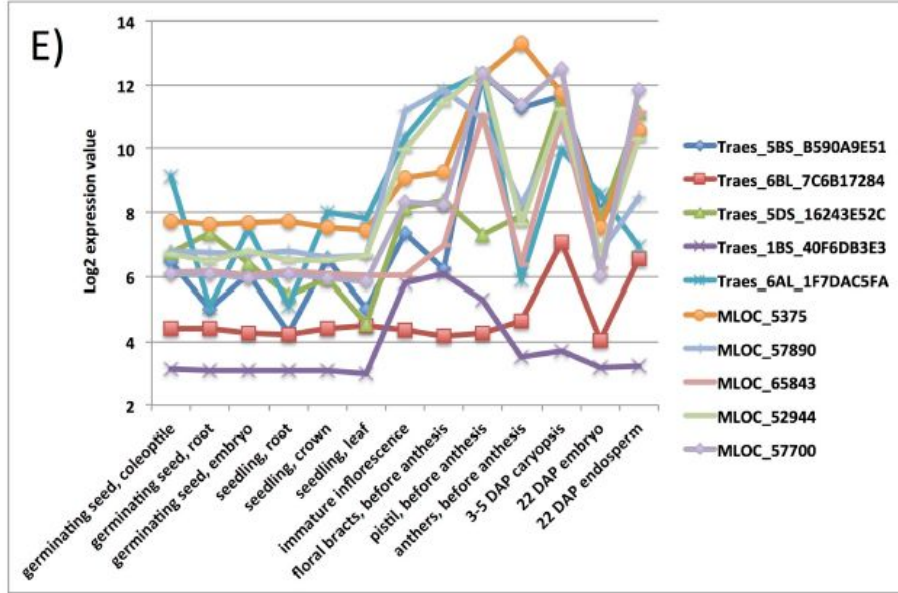
Supplemental Figure S3

- Supplemental Figure S3 was essentially a continuation of Figure 4
 - Depicted were: (A) OG5_127642, and (B) OG5_163093 in the PEBP family; (C) OG5_144912, (D) OG5_177438, (E) OG5_135817, and (F) OG5_190130 in the MADS-box family; (G) OG5_139246 and (H) OG5_156319 in the BBX family.
- Only differences between Figure S3 and Figure 4 were minor formatting changes and high number of genes in some figures



Original Figure S3e. Expression profiles of orthologous genes in the two ortholog groups of OG5_135817

Supplemental Figure S3



Side-by-side comparison: Figure S3e from paper (left) and reproduced in R (right)
 Expression profiles of orthologous genes in the two ortholog groups of PEBP
 OG5_135817

Thank you!