5-2023

Camelina sativa promoters of seed development genes resemble *Arabidopsis thaliana* orthologs

Benjamin Clark

Jiayin Liu

Haiying Liang

Camelina sativa promoters of seed development genes resemble *Arabidopsis thaliana* orthologs

Benjamin Clark*, Jiayin Liu, and Haiying Liang

Abstract

Camelina sativa, a relative of the model plant *Arabidopsis thaliana*, has gained commercial interest in recent years for its seed oil. Camelina oil is a desirable ingredient in animal feed and for cooking due to its rich concentration of omega-3 fatty acids. It is also emerging as an eco-friendly jet biofuel. However, since *C. sativa* has only recently entered the public eye, regulation of its seed oil synthesis remains largely unexplored. *A. thaliana*, on the other hand, has been extensively studied. Promoters directly upstream of genes harbor hundreds of cis-acting regulatory DNA elements. These motifs are often tissue-specific and associated with specific biological processes. Many motifs are also induced by hormones, which regulate development. Comparing promoter motifs of four prominent seed development genes, *FUS3*, *LEC1*, *LEC2*, and *ABI3*, in *A. thaliana* to their orthologs in *C. sativa* revealed significant similarity between the two. This suggests that regulation of seed development in *C. sativa* may be similar to that of *A. thaliana*. Notably, seed-specific motifs associated with storage proteins and carbon metabolism are most highly represented in both species. However, *C. sativa* seed-specific promoters were found to harbor a greater composition of motifs induced by abscisic acid, which has been implicated in seed maturation and dormancy. Moving forward, expression analysis of these genes in various *C. sativa* tissue types will further validate their function. Understanding specific regulatory factors modulating seed development will reveal molecular targets to improve camelina oil yield for industrial applications.

Introduction

Promoters are regions directly upstream of genes that contain many cis-acting regulatory DNA elements. These elements can serve as binding sites for a wide variety of trans-acting regulatory molecules, leading to sequence-specific interactions that modulate gene expression. These elements are often induced in specific tissue types (Li et al. 2014), and at specific phases of development (Wang et al. 2013). They are also often associated with particular biological processes such as light response (Allison 1995) and can be induced by a wide array of hormones (Che 2002).

Seed development was first described in *A. thaliana* as being controlled by four key genes: *FUS3*, *LEC1*, *LEC2*, and *ABI3*. These genes encode B3 domain-containing transcription factors that promote seed development programs. These loss-of-function mutants in *A. thaliana* result in a premature exit of embryogenesis and early entrance into seedling development programs. Notably, the abundance of seed storage components was lacking in mutants as compared to wildtype seeds. These genes have also demonstrated responsiveness to a number of abiotic factors, including dehydration, thus improving desiccation tolerance (Wang et al. 2013).

While *A. thaliana* harbors only one copy of each of these genes, a whole-genome triplication of the *C. sativa* ancestral genome alongside a series of translocations and hybridization events has resulted in 3 copies of each gene in *C. sativa* (Kagale et al. 2014). In fact, *C. sativa LEC1* harbors 2 variants of each copy, resulting in a total of 6 *LEC1* coding sequences as compared to a single ortholog in *A. thaliana*.

Phytohormones are key drivers of differentiation, seed development, and other physiological processes. Proper seed development requires the differentiation of multipotent localized cells into a seed coat, embryo, and endosperm. Proper hormone secretions, gradients, and receptor binding promote normal differentiation and development. Auxins in particular have been identified for their role in seed development, especially pattern formation, cell division, and expansion

(Locascio et al. 2014). In addition, abscisic acid promotes seed growth, development, and dormancy. It also decreases water uptake in plants, which is associated with seed growth (Schopfer et al. 1979).

In recent years, camelina oil has gained interest for its use in animal feed and for human consumption due to its high concentration of omega-3 fatty acids, as well as for its utility as an eco-friendly jet biofuel. Camelina oil, extracted from the *C. sativa* seed, has sparked new research into genomic factors determining camelina seed oil production. Since *C. sativa* has only recently entered the public eye, regulation of its seed oil synthesis remains largely unexplored. *A. thaliana*, on the other hand, has been extensively studied. The aim of this study is to compare promoter elements between *A. thaliana* and *C. sativa* to gain insight into factors governing seed development in *C. sativa*. Understanding specific regulatory factors modulating seed development will reveal molecular targets to improve camelina oil yield for industrial applications. Given their evolutionary proximity to one another, it is hypothesized that regulation of seed development is largely conserved between these two plant species.

Materials and Methods

Collecting promoter sequences

In order to collect promoter sequences for *FUS3*, *LEC1*, *LEC2*, and *ABI3* genes in *A. thaliana* and *C. sativa*, their 5' UTR and coding sequences were first collected from *The Arabidopsis Information Resource* (TAIR) database and *Ensembl Plants*, respectively. For cases in which *C. sativa* sequences were missing from the *Ensembl* database, a nucleotide BLAST of *A. thaliana* orthologs was used to identify missing *C. sativa* copies of the target gene. After navigating to the specific genomic location of the target gene using the full chromosome sequence in the NCBI database, the 3 kb sequence directly upstream of the transcription start site was collected for each gene copy. These were confirmed using the multiple sequence alignment tool *MultAlin* to align the 5' UTR sequence to the 3 kb promoter sequence.

Promoter analysis

The collected 3 kb promoter sequences were then mined for cis-acting regulatory DNA elements using the *New PLACE* database. These motifs were recorded for each gene copy in both species and annotated for associated tissue types, biological processes, and hormones. Tissue types included in the analysis were seed, leaf, root, pollen, multiple with seed, multiple without seed, and unknown. Biological processes included in the analysis were storage proteins/carbon metabolism, embryogenesis, wounding response, pathogen/disease/elicitor induced, water response, light response, temperature response, anaerobically induced, and hormone induced. Hormones included were gibberellic acid (GA), abscisic acid (ABA), salicylic acid (SA), cytokinin (CK), auxin (IAA), and ethylene (ET). The number of elements associated with each subcategory were totaled using Excel, and graphs were constructed to compare *A. thaliana* and *C. sativa* in terms of cis-acting regulatory DNA elements controlling orthologous seed development genes.

RNA Isolation

Tissue samples were first collected from wildtype *C. sativa* plants. These samples included immature seed, mature seed, young leaf, old leaf, stem, and flower. RNA was extracted from these samples using the TRIzol RNA isolation protocol. However, for recalcitrant samples, including immature and mature seed, CTAB isolation buffer will need to be used for isolation.

Results

Comparison of tissue-type associated motif counts

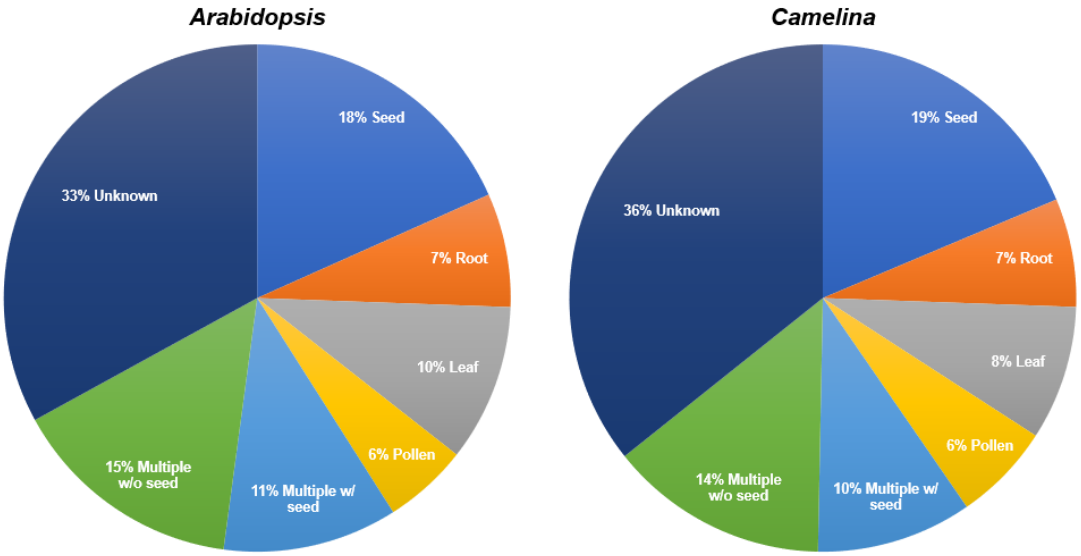

Figure 1. Comparing the promoter composition of seed development genes for *A. thaliana* (left) and *C. sativa* (right) reveals significant similarity in terms of tissue-specific interactions. In both, promoter elements induced in seeds are the most highly represented, followed by leaf, root, and pollen. At least 29% of cis-acting elements in target gene promoters in both species are known to be affected in seeds.

Comparison of biological process associated motif counts

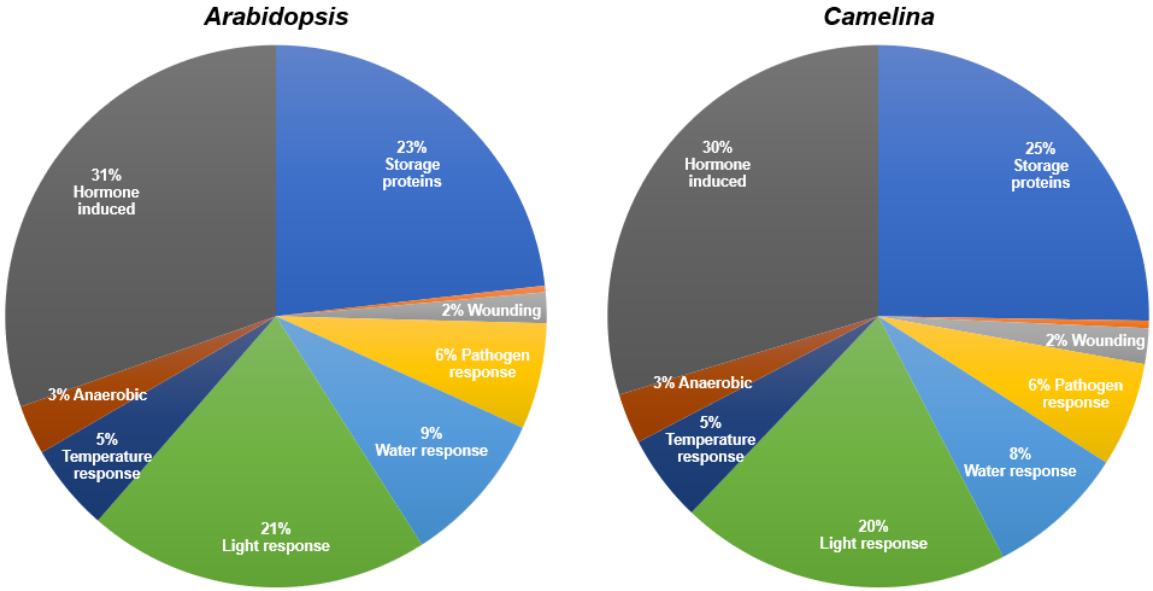

Figure 2. Comparing the promoter composition of DNA elements associated with biological processes for seed development genes in *A. thaliana* (left) and *C. sativa* (right) reveals significant similarity. In both, promoter elements that are activated by hormones, associated with storage proteins or carbon metabolism, and responsive to light are the most highly represented.

Comparison of hormone associated motif counts

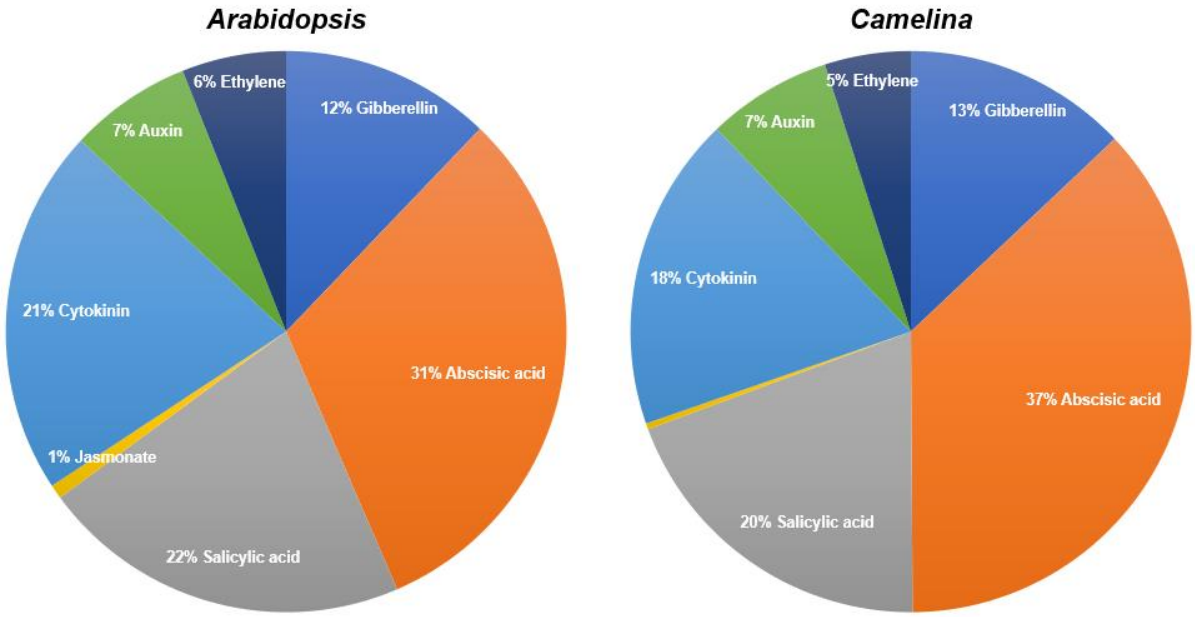

Figure 3. Comparing the promoter composition of DNA elements associated with biological processes for seed development genes in *A. thaliana* (left) and *C. sativa* (right) reveals significant similarity. In both, promoter elements that are modulated by abscisic acid are the most highly represented, followed by salicylic acid and cytokinin. Promoter elements modulated by ethylene are the least represented, followed by auxin and gibberellin. The number of elements modulated by jasmonates and other phytohormones is negligible.

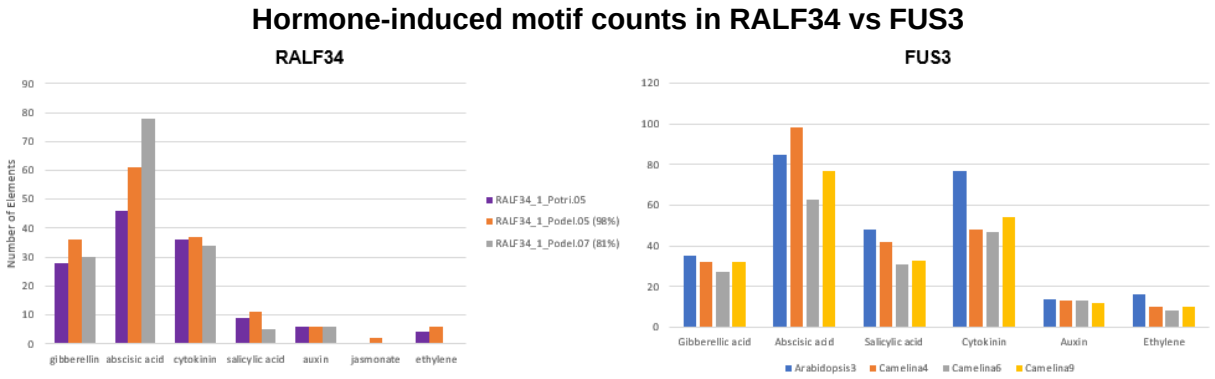

Figure 4. Comparing the promoter composition of DNA elements in a rapid alkalinization factor gene in *Poplar* (*RALF34*; left) and a seed development gene in *Arabidopsis* and *Camelina* (*FUS3*; right). In both, motifs induced by abscisic acid are the most highly represented, followed by cytokinin and gibberellin. Salicylic acid induced motifs are significantly more highly represented in FUS3 than in RALF34.

RNA isolation

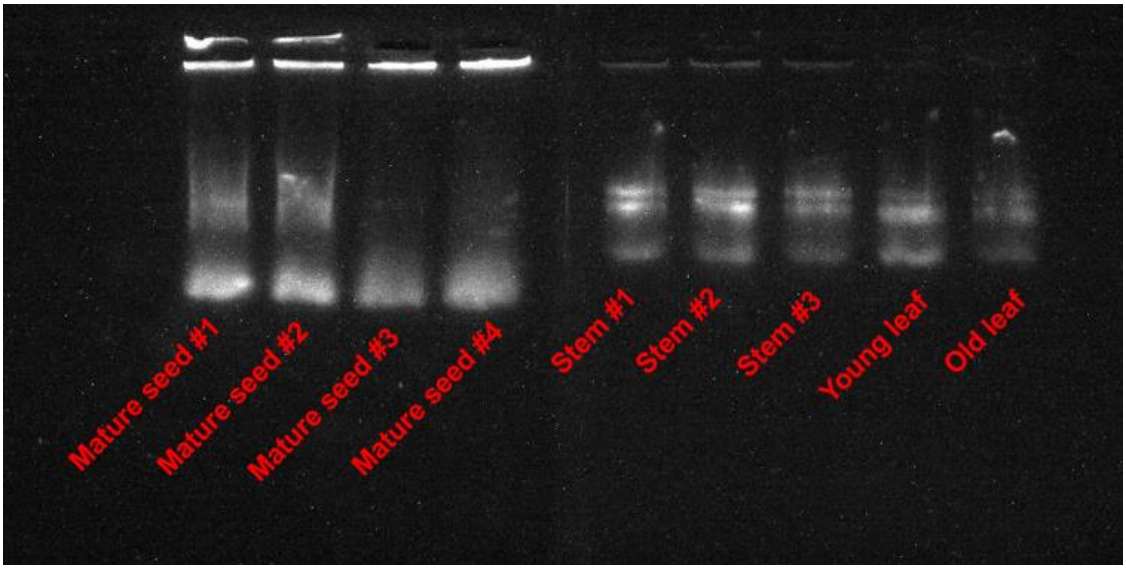

Figure 5. Gel electrophoresis of RNA samples from several *C. sativa* tissue types. Immature seed RNA (not shown) produced no band; mature seed RNA showed heavy, bright bands next to the well and smeared/absent bands in the 28S/18S RNA region; stem and leaf RNA shows distinct 28S/18S RNA bands.

Discussion

The most highly represented biological processes associated with promoter elements of seed development genes include hormone response, seed proteins/carbon metabolism, and light response. The role of each of these in seed development is well-defined. Phytohormones regulate differentiation into seed components and development by establishing a hormone gradient throughout the seed, allowing diversely localized seed cells to respond differently. For analytical purposes, the seed proteins and carbon metabolism categories were merged because of substantial overlap between the two. Induction of these motifs likely contributes to the abundance of protein, carbohydrates, and oil in the seed. Circadian effects, defined by light response, have also been demonstrated to play a critical role in seed development. In fact, response to abscisic acid and gibberellin, two major regulators of seed development, require normal circadian function (Penfield 2009).

Since *C. sativa* research is still a relatively young field, its online databases are often weaker than those of *A. thaliana*, which is the most widely studied plant model organism (Woodward et al. 2018). This has posed two challenges. Firstly, the *C. sativa* database in *Ensembl Plants* often did not contain sequences for all copies of the target gene. This was solved by subjecting the A. thaliana ortholog to a BLAST search in NCBI to identify novel transcripts. Secondly, the 3 kb promoter region directly upstream of *C. sativa* genes occasionally contained a string of unidentified nucleotides which could not be considered for promoter analysis. This fact is prominently reflected in the Chromosome 7 copy of *C. sativa LEC2* (**Figs. 6-8**). However, this loss of cis-regulatory elements data does not seem to skew its overall composition in the analysis in any predictable way.

Upon promoter analysis in New PLACE, the distribution of cis-acting regulatory DNA elements in *FUS3*, *LEC1*, *LEC2*, and *ABI3* in *A. thaliana* highly resembles that of *C. sativa* with respect to

associated tissue types (**Fig. 1**), biological processes (**Fig. 2**), and hormones (**Fig. 3**). However, the proportion of abscisic acid-induced motifs was noticeably higher in C. sativa overall, which may warrant further investigation. Abscisic acid has been implicated in driving seed maturation and dormancy, as well as downstream responses to biotic and abiotic environmental changes (Chen et al. 2019).

A previous study investigated the cis-regulatory promoter elements of *Rapid alkalinization factor* (RALF) genes in *Populus*, which play a key role in molecular response to the fungal pathogen *Sphaerulina musiva* (Randazza et al. 2022). Comparison of cis-regulatory promoter elements in *Arabidopsis*/*Camelina FUS3* to *Populus RALF34* indicates that salicylic acid may play a more important role in modulating seed development than in pathogen response (**Fig. 4**). Notably, abscisic acid plays a substantial role in both processes (Finkelstein 2013).

Descriptions of an entire promoter as being tissue-specific are generally based on the *in-vivo* expression levels of the target gene in various tissue types rather than an *in-silico* analysis of predicted tissue types in which tissue-specific induction of particular cis-regulatory promoter elements may or may not be sufficient for expression of the target gene in that tissue type (**Fig. 1**). Thus, a tissue type-dependent expression analysis by quantitative reverse transcription PCR (RT-qPCR) will determine whether *FUS3*, *LEC1*, *LEC2*, and *ABI3* promoters drive expression in a truly seed-specific manner, or if these genes may play a supplementary role in other tissue processes as well.

In order to perform this analysis, we set out to extract RNA from mature and immature seeds, young and old leaves, stems, and flowers. Roots were omitted from this study due to difficulty growing *C. sativa* roots hydroponically. When RNA isolation was attempted on seed samples using the TRIzol RNA isolation protocol, protein and other non-RNA contaminants were an ongoing threat, due to the recalcitrant nature of seed components. So, a different RNA extraction buffer called cetyltrimethylammonium bromide (CTAB) will be used for these tissue types. CTAB is a surfactant that has demonstrated efficacy in extracting abundant, pure RNA from a polysaccharide matrix (Wang et al. 2009). After cDNA synthesis, primer pairs designed for *FUS3*, *LEC1*, *LEC2*, and *ABI3* will be used for RT-qPCR of RNA samples from each tissue type. Since for each of these genes, *C. sativa* has multiple copies that are nearly identical in sequence, primers were designed such that a single primer pair will target all copies of a particular gene indiscriminately. Consequently, PCR analysis of cDNA samples will measure the combined expression of all *C. sativa* copies against the expression level of its *A. thaliana* ortholog for each gene in each tissue type. This will provide both insight into the function of these genes, as well as an interesting comparison to the *in-silico* analysis of tissue type-related cis-regulatory elements depicted in **Figure 1**.

References

Allison LA, Maliga P. Light-responsive and transcription-enhancing elements regulate the plastid psbD core promoter. EMBO J. 1995 Aug 1;14(15):3721-30. doi: 10.1002/j.1460-2075.1995.tb00042.x. PMID: 7641691; PMCID: PMC394447.

Che P, Gingerich DJ, Lall S, Howell SH. Global and hormone-induced gene expression changes during shoot development in Arabidopsis. Plant Cell. 2002 Nov;14(11):2771-85. doi: 10.1105/tpc.006668. PMID: 12417700; PMCID: PMC152726.

Chen K, Li GJ, Bressan RA, Song CP, Zhu JK, Zhao Y. Abscisic acid dynamics, signaling, and functions in plants. J Integr Plant Biol. 2020 Jan;62(1):25-54. doi: 10.1111/jipb.12899. PMID: 31850654.

Finkelstein R. Abscisic Acid synthesis and response. Arabidopsis Book. 2013 Nov 1;11:e0166. doi: 10.1199/tab.0166. PMID: 24273463; PMCID: PMC3833200.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. "Plant cis-acting regulatory DNA elements (PLACE) database: 1999" Nucleic Acids Research, Volume 27, Issue 1, 1999, Pages 297-300.

Kagale, S., Koh, C., Nixon, J. et al. The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. Nat Commun 5, 3706 (2014). https://doi.org/10.1038/ncomms4706

Li J, Yuan J, Li M. Characterization of putative cis-regulatory elements in genes preferentially expressed in Arabidopsis male meiocytes. Biomed Res Int. 2014;2014:708364. doi: 10.1155/2014/708364. Epub 2014 Aug 27. PMID: 25250331; PMCID: PMC4163388.

Locascio A, Roig-Villanova I, Bernardi J, Varotto S. Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. Front Plant Sci. 2014 Aug 25;5:412. doi: 10.3389/fpls.2014.00412. PMID: 25202316; PMCID: PMC4142864.

Multiple sequence alignment with hierarchical clustering. F CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890.

Penfield S, Hall A. A role for multiple circadian clock genes in the response to signals that break seed dormancy in Arabidopsis. Plant Cell. 2009 Jun;21(6):1722-32. doi: 10.1105/tpc.108.064022. Epub 2009 Jun 19. PMID: 19542296; PMCID: PMC2714941.

Schopfer P, Bajracharya D, Plachy C. Control of Seed Germination by Abscisic Acid: I. Time Course of Action in Sinapis alba L. Plant Physiol. 1979 Nov;64(5):822-7. doi: 10.1104/pp.64.5.822. PMID: 16661062; PMCID: PMC543371.

Wang F, Perry SE. Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol. 2013 Mar;161(3):1251-64. doi: 10.1104/pp.112.212282. Epub 2013 Jan 11. PMID: 23314941; PMCID: PMC3585594.

Wang L, Stegemann JP. Extraction of high quality RNA from polysaccharide matrices using cetyltrimethylammonium bromide. Biomaterials. 2010 Mar;31(7):1612-8. doi: 10.1016/j.biomaterials.2009.11.024. Epub 2009 Dec 3. PMID: 19962190; PMCID: PMC2813910.

Woodward AW, Bartel B. Biology in Bloom: A Primer on the Arabidopsis thaliana Model System. Genetics. 2018 Apr;208(4):1337-1349. doi: 10.1534/genetics.118.300755. PMID: 29618591; PMCID: PMC5887134.

Supplemental figures

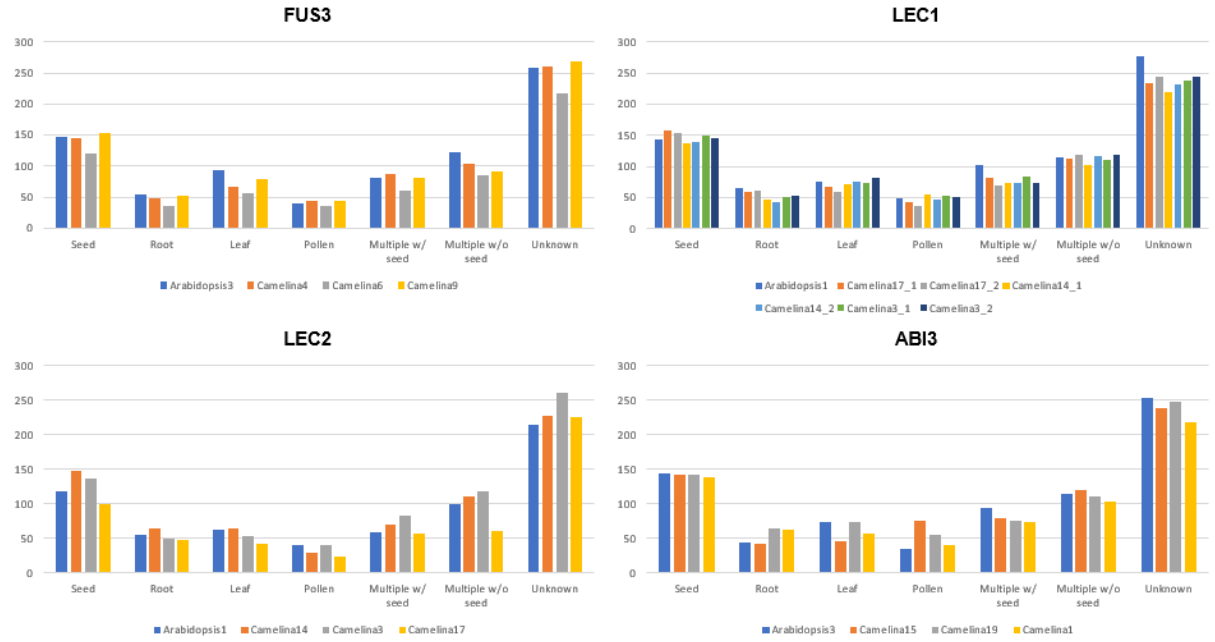

Figure 6. Tissue-specific element counts in the 3 kb promoters of 4 seed development genes in *A. thaliana* and *C. sativa*.

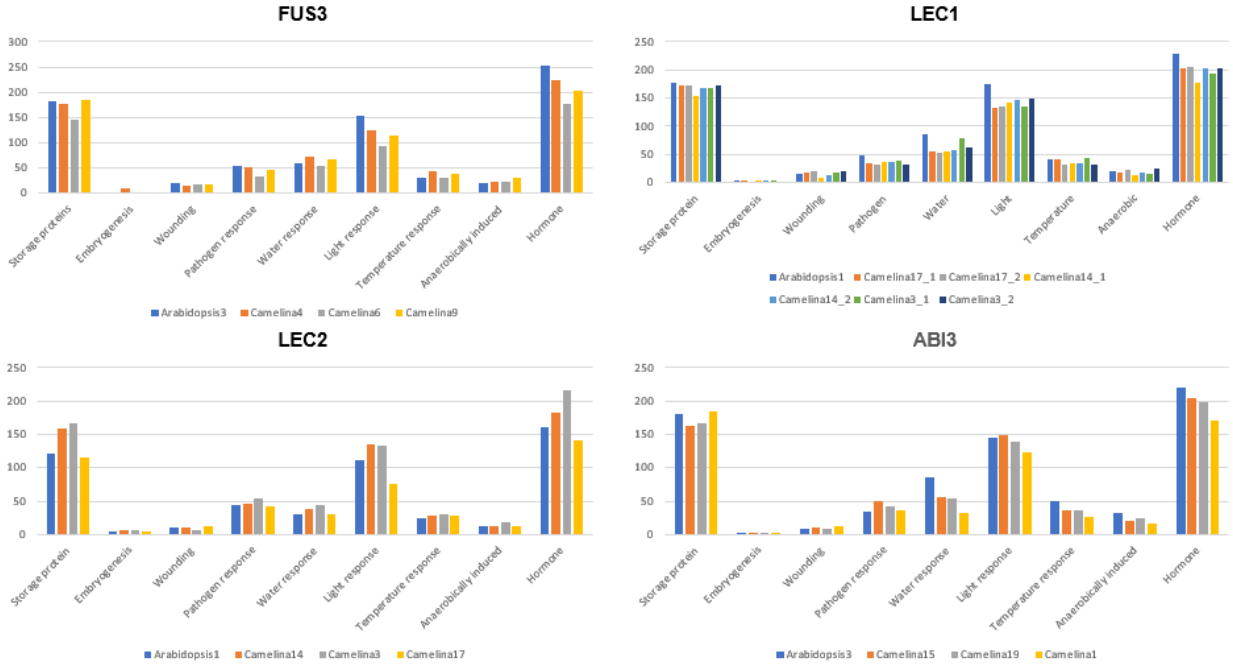

Figure 7. Biological process associated element counts in the 3 kb promoters of 4 seed development genes in *A. thaliana* and *C. sativa*.

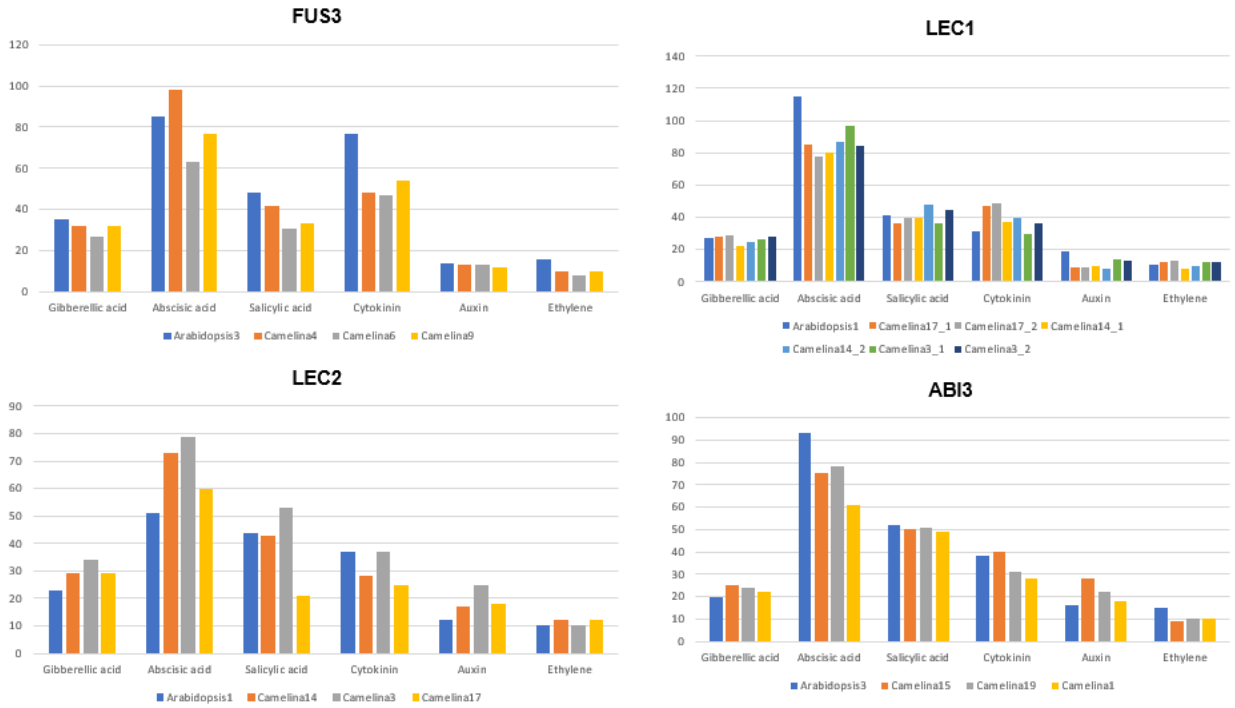

Figure 8. Hormone-induced element counts in the 3 kb promoters of 4 seed development genes in *A. thaliana* and *C. sativa*.