Comparative transcriptomics identifies the transcription factors BRANCHED1 and TCP4, as well as the microRNA miR166 as candidate genes involved in the evolutionary transition from dehiscent to indehiscent fruits in Lepidium

Lydia Gramzow
Friedrich Schiller University Jena
Katharina Klupsch
Friedrich Schiller University Jena
Noé Fernández Pozo
University of Marburg
Martin Hölzer
Friedrich Schiller University Jena
Manja Marz
Friedrich Schiller University Jena
Stefan A. Rensing
University of Marburg
Günter Theißen (guenter.theissen@uni-jena.de)
Friedrich Schiller University Jena

Research Article

Keywords: transcriptome, fruit development, dehiscence, differentially expressed genes, Lepidium appelianum, Lepidium campestre

Posted Date: November 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1031457/v1

Comparative transcriptomics identifies the transcription factors BRANCHED1 and TCP4, as well as the microRNA miR166 as candidate genes involved in the evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* (Brassicaceae)

Lydia Gramzow[1], Katharina Klupsch[1], Noé Fernández Pozo[2], Martin Hölzer[3†], Manja Marz[3], Stefan A. Rensing[2,4] and Günter Theißen[1*]

[1]Matthias Schleiden Institute / Genetics, Friedrich Schiller University Jena, 07743 Jena, Germany

[2]Plant Cell Biology, Department of Biology, University of Marburg, 35043 Marburg, Germany

[3]RNA Bioinformatics and High-Throughput Analysis, Friedrich Schiller University Jena, 07743 Jena, Germany

[4]Centre for Biological Signaling Studies (BIOSS), University of Freiburg, 79108 Freiburg, Germany

[†]Current address: Methodology and Research Infrastructure, Bioinformatics, Robert Koch Institute, 13353 Berlin, Germany

*Corresponding author: Günter Theißen, guenter.theissen@uni-jena.de

Keywords: transcriptome, fruit development, dehiscence, differentially expressed genes, *Lepidium appelianum*, *Lepidium campestre*

Abstract

Background: Fruits are the seed-bearing structures of flowering plants and are highly diverse in terms of morphology, texture and maturation. Dehiscent fruits split open upon maturation to discharge their seeds while indehiscent fruits are dispersed as a whole. Indehiscent fruits evolved from dehiscent fruits several times independently in the crucifer family (Brassicaceae). The fruits of *Lepidium appelianum*, for example, are indehiscent while the fruits of the closely related *L. campestre* are dehiscent. Here, we investigate the molecular and genetic mechanisms underlying the evolutionary transition from dehiscent to indehiscent fruits using these two Lepidium species as model system.

Results: We have sequenced the transcriptomes and small RNAs of floral buds, flowers and fruits of *L. appelianum* and *L. campestre* and analyzed differentially expressed genes (DEGs) and differently differentially expressed genes (DDEGs). DEGs are genes that show significantly different transcript levels in the same structures (buds, flowers and fruits) in different species, or in different structures in the same species. DDEGs are genes for which the change in expression level between two structures is significantly different in one species than in the other. Comparing the two species, the highest number of DEGs was found in flowers, followed by fruits and floral buds while the highest number of DDEGs was found in fruits versus flowers followed by flowers versus floral buds. Several gene ontology terms related to cell wall synthesis and degradation were overrepresented in different sets of DEGs highlighting the importance of these processes for fruit opening. Furthermore, the fruit valve identity genes *FRUITFUL* and *YABBY3* were among the DEGs identified. Finally, the microRNA miR166 as well as the TCP transcription factors *BRANCHED1* (*BRC1*) and *TCP FAMILY TRANSCRIPTION FACTOR 4* (*TCP4*) were found to be DDEGs.

Conclusions: Our study reveals differences in gene expression between dehiscent and indehiscent fruits and uncovers miR166, *BRC1* and *TCP4* as possible causes for the evolutionary transition from dehiscent to indehiscent fruits in *Lepidium*.

Background

Flowering plants (angiosperms) form fruits to protect and disperse their seeds. Fruits come in many different types with different morphologies and different properties such as dry or fleshy, and dehiscent or indehiscent (Lorts and Briggeman 2008). There is a tremendous variation in fruit types both across and within different plant lineages (Dardick and Callahan 2014). However, the evolutionary mechanisms that enabled such dramatic shifts to occur, often in a relatively short period of time, remain largely unknown.

The crucifer family (Brassicaceae) includes a number of economically important plants such as cabbage, broccoli, mustard, radish, and turnips. The model plant *Arabidopsis thaliana* is also a member of this family (Initiative 2000). Typical fruits of Brassicaceae species are dehiscent, i.e. that the fruits open upon maturation to release the seeds. Dehiscent fruits also likely represents the ancestral fruit type of Brassicaceae (Hall et al. 2002). However, indehiscent fruits, i.e. fruits that only release the seed upon decomposition of the fruit, are found in many tribes distributed across the Brassicaceae phylogeny (Appel et al. 2003). The scattered distribution of indehiscent fruits indicates that this property evolved independently several times. This situation is mirrored in the genus *Lepidium* belonging to Brassicaceae: Species of this genus typically produce two-seeded dehiscent fruits, but the genus also includes species with indehiscent fruits (Al-Shehbaz and Mummenhoff 2011).

Brassicaceae fruits are composed of two fruit valves that are connected to the replum and enclose the developing seeds. Dehiscent fruits, such as those of *A. thaliana* and *Lepidium campestre* (also known as field pepperwort or field cress), form a well-defined dehiscence zone (DZ) at the valve margin (Mühlhausen et al. 2013). The DZ consists of the lignified layer, a stripe of lignified cells, and a separation layer, a region of small thin-walled cells (Spence et al. 1996; Rajani and Sundaresan 2001). During fruit ripening, the whole fruit dries and shrinks. Only the lignified structures stay rigid. Thereby a spring-like tension is created within the fruit. At the same time, the middle lamellae of the separation layer cells degenerate to form a pre-determined breaking zone at which the pressure tears the valves apart from the replum. Consequently, the fruit bursts open to release the seeds (Meakin and Roberts 1990, 1991;

Spence et al. 1996). In contrast, the indehiscent fruits of the closely related *Lepidium appelianum* do not form a DZ. Instead, a continuous ring of lignified cells surrounds the seeds such that the fruit cannot open (Mühlhausen et al. 2013).

Much of the gene regulatory network underlying the proper formation of the fruit valves, replum and DZ has been elucidated in *A. thaliana* (reviewed in (Ballester and Ferrándiz 2017)). Establishment of the DZ requires expression of the two redundant MADS box genes, *SHATTERPROOF1* (*SHP1*) and *SHATTERPROOF2* (*SHP2*). The SHP1 and SHP2 proteins act as transcription factors and activate the basic helix-loop-helix protein-encoding genes *INDEHISCENT* (*IND*), *ALCATRAZ* (*ALC*) and *SPATULA* (*SPT*), and also autonomously contribute to DZ development (Liljegren et al. 2000; Liljegren et al. 2004; Rajani and Sundaresan 2001; Groszmann et al. 2011).

For correct fruit patterning, it is crucial that the expression of the *SHP* genes is restricted to the DZ. Three transcription factors contribute to this process: The MADS box gene *FRUITFULL* (*FUL*) which is expressed in the fruit valves (Gu et al. 1998; Ferrandiz et al. 2000), the BEL1-like homeobox gene *REPLUMLESS* (*RPL*) (Roeder et al. 2003), also known as *PENNYWISE* (Smith and Hake 2003), *BELLRINGER* (Byrne et al. 2003), *VAAMANA* (Bhatt et al. 2004), and *BLH9* (Cole et al. 2006) which is expressed in the replum, and the floral homeotic gene *APETALA2* (*AP2*) which also negatively regulates *RPL* (Ripoll et al. 2011).

Transcription factors controlling the expression of these negative regulators have also been determined. High levels of the C2H2 zinc finger proteins JAGGED (JAG) and the two closely related YABBY1 group proteins FILAMENTOUS FLOWER (FIL) and YABBY3 (YAB3) activate the expression of *FUL* (Dinneny et al. 2005). In contrast, lower levels of JAG/FIL/YAB3 expression promote expression of *SHP* genes. The expression of RPL is activated by the knotted1-like homeobox protein BREVIPEDICELLUS (BP) (Alonso-Cantabrana et al. 2007) whose gene is in turn activated by the C2H2 zinc finger protein NO TRANSMITTING TRACT (NTT) (Marsch-Martínez et al. 2014). *AP2* is negatively regulated be the microRNA miR172 (Ripoll et al. 2015).

Additionally, other factors which influence the size and the position of the DZ have been identified. The *WUSCHEL-RELATED HOMEOBOX gene 13* (*WOX13*) controls replum width and negatively regulates *JAG*/*FIL*/*YAB3* (Romera-Branchat et al. 2013). The auxin-response factors ARF6 and ARF8, which are regulated by miR167 (Zheng et al. 2019), activate miR172 together with FUL (Ripoll et al. 2015). The MYB protein ASYMMETRIC LEAVES 1 (AS1), likely in collaboration with the leucine zipper protein ASYMMETRIC LEAVES 2 (AS2), negatively regulates BP (Alonso-Cantabrana et al. 2007).

In general, proteins encoded by genes expressed in the replum often negatively regulate genes expressed in the valves and *vice versa*. Apart from the already mentioned interactions, this includes negative regulation of the replum gene *BP* by the valve proteins encoded by *JAG*/*FIL*/*YAB3*, and negative regulation of *JAG*/*FIL*/*YAB3* by the replum protein RPL (González-Reig et al. 2012).

In a previous study, we have shown that orthologues of the valve margin genes are expressed in a similar way in *L. campestre* (dehiscent fruits) as in *A. thaliana* fruits but that expression of the respective orthologues is abolished in the corresponding tissues of indehiscent *Lepidium appelianum* fruits (Mühlhausen et al. 2013). As parallel mutations in different genes are unlikely, we concluded that the changes in gene expression patterns are probably caused by changes in upstream regulators such as FUL, RPL or AP2.

To conduct a more unbiased approach to identify the genetic changes that lead from dehiscent to indehiscent fruits than the analysis of candidate genes, we have sequenced the transcriptomes of floral buds, flowers and fruits of both, *L. campestre* and *L. appelianum* in the present study. We have identified differentially expressed genes (DEGs) and differently differentially regulated genes (DDEGs) where the latter refers to genes for which the change in expression level between two structures is significantly different in one species than in the other. More DEGs were identified in flowers than in fruits and floral buds and a higher number of DDEGs was found in fruits versus flowers than in flowers versus floral buds. Cell wall synthesis and degradation are important processes for fruit opening as revealed by gene ontology (GO) analysis. The fruit valve identity genes *FRUITFUL* and *YABBY3* were identified

as DEGs such that the possible cause for the evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* may even be an upstream factor of these genes. Possible candidates are *BRANCHED1* (*BRC1*), an ortholog of which may determine whether dehiscent or indehiscent fruit develop on the dimorphic plant *Aethionema arabicum*, and *TCP FAMILY TRANSCRIPTION FACTOR 4* (*TCP4*) which may regulate *YABBY3*. These two genes were found to be DDEGs. Our study elucidates differences in gene expression patterns between dehiscent and indehiscent fruits and reveals *BRC1* and *TCP4* as possible causes for the evolutionary transition from dehiscent to indehiscent fruits in *Lepidium*.

Results

Overview of the RNA-seq analysis and transcriptome assembly

Sequencing resulted in an average number of reads per library of 56 Mio. for the mRNA and 12 Mio. for the small RNA (Table 1). An initial analysis of the data revealed contamination with sequences from thrips, likely due to infestation of the plants by these animals. Hence, we removed reads matching to the genome of the thrips *Frankliniella occidentalis* (González et al. 2018) as well as uncorrectable and unpaired reads and reads corresponding to organelle sequences. After this filtering step, 42 Mio. were retained for further analyses for the mRNA sample. For the small RNA sample many reads seem to be derived from organelle RNA. Hence, after removing uncorrectable reads and those matching to the *Frankliniella occidentalis* genome and organelle sequences, only 1.5 Mio. reads remained on average for the small RNA sample (Table 1).

Assembly using Trinity (Grabherr et al. 2011a) resulted in a total of 56,413 transcripts for *L. campestre* and 70,380 transcripts for *L. appelianum* after removing putative contaminant sequences but including potential splice variants or fragmentary sequences. The assemblies also contained chimeric sequences composed of two different transcripts which were likely a result of mis-assembly (Yang and Smith 2013b). Separation of chimeric sequences increased the number of transcripts to a total of 57,209 for *L. campestre* and 71,332 for *L. appelianum*. We used the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool (Simão et al.

2015) with the dataset eudicotyledons_odb10 as reference to assess completeness of our transcriptomes. The BUSCO analyses revealed that 94.6% of the expected eudicotyledonous "near-universal single-copy orthologs" are present in our assembly of the *L. campestre* transcriptome while 94.3% of these BUSCOs are present in our *L. appelianum* transcriptome (Figure 1). It is common that some genes are fragmented in *de novo* assemblies. Hence, we analyzed the length distribution of our assemblies. For both species there are two peaks (Figure 2). One peak appears at a length of about 240 nucleotides and probably represents fragments. The other peak was found at a length of about 1,450 nucleotides which indicates that there are also a number of full-length transcripts.

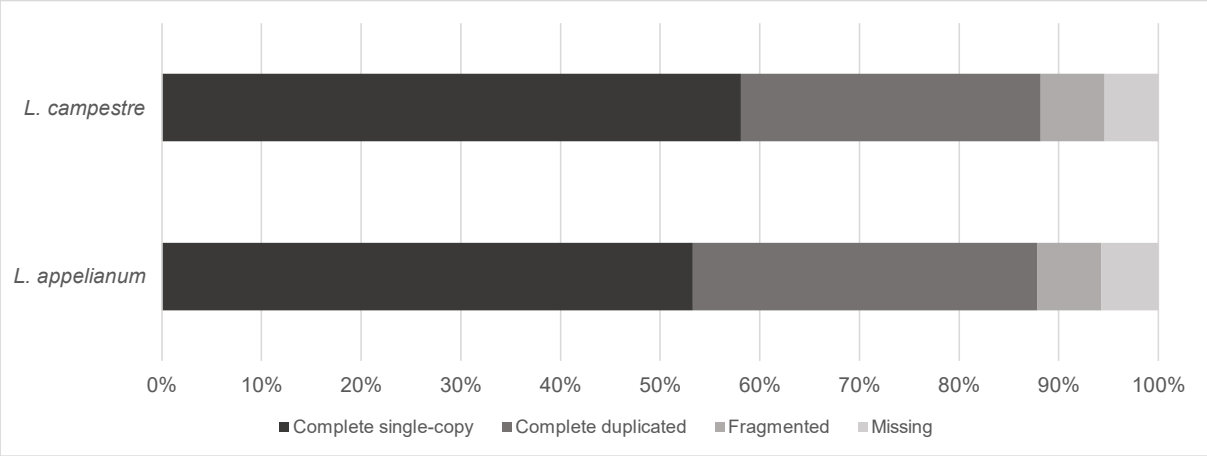


Figure 1: BUSCO completeness analysis. Transcripts from the *L. campestre* and *L. appelianum* assemblies were compared to 2121 Eudicotyledons reference orthologs for completeness assessment.

To detect conserved miRNAs, we mapped the small RNA reads onto the mature miRNAs of *A. thaliana* as provided by miRBase (Kozomara et al. 2019a). We found reads for 64 mature miRNAs belonging to 32 miRNA families in the *L. campestre* small RNA data (Table 2). Using ShortStack (Axtell 2013) and the *L. campestre* genome as available from NCBI, we identified three novel miRNAs. However, no putative target genes could be identified in the transcriptome of *L. campestre* using targetfinder (https://github.com/carringtonlab/TargetFinder). Our *L. appelianum* small RNA data contained reads of 60 mature miRNAs belonging to 30 miRNA families (Table 2). No novel miRNAs could be identified for *L. appelianum* using ShortStack and our transcriptome as reference "genome".

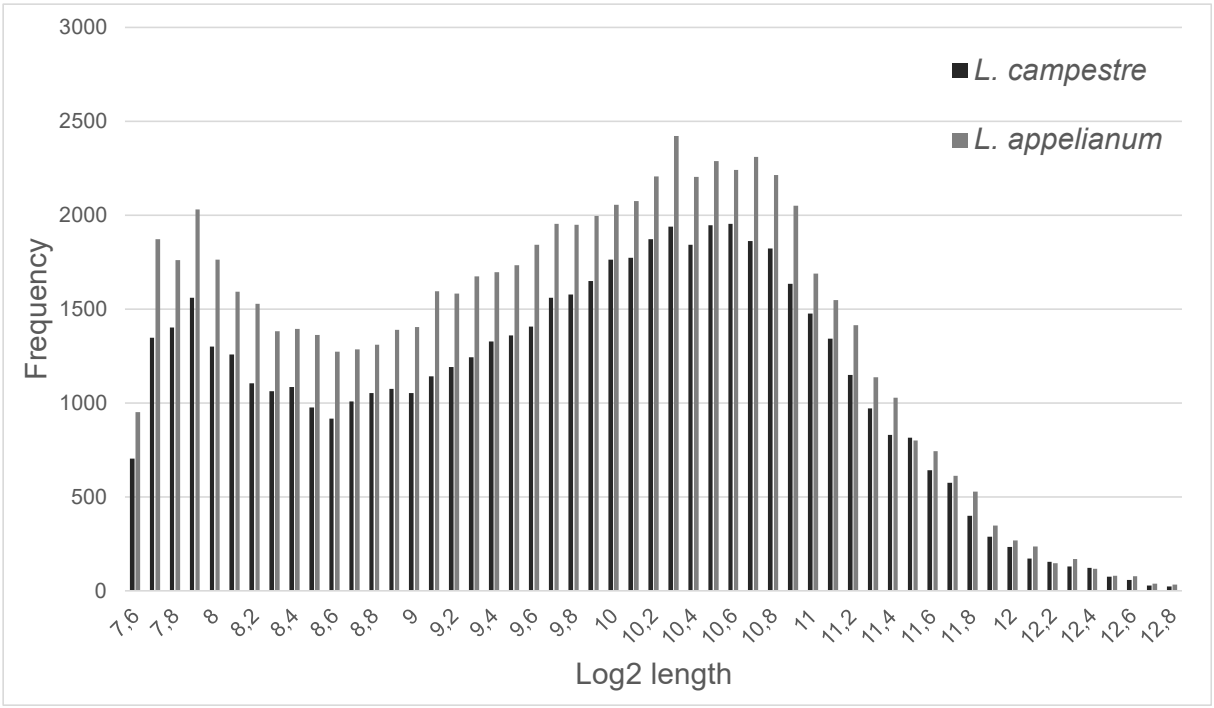


Figure 2: Transcript length distribution of the assembled transcripts of *L. campestre* and *L. appelianum*.

To assess completeness of our small RNA data, we compared our results to the set of conserved and moderately conserved miRNA families as identified by miRNA sample sequencing of vascular plants (Montes et al. 2014). For both species, we identified reads for all 16 miRNA families that were found to have originated before the emergence of eudicots and to be conserved across virtually all corresponding species. Furthermore, we found reads for 6 miRNA families in our *L. campestre* and 7 miRNA families in our *L. appelianum* small RNA data out of 21 miRNA families which were classified as conserved, although missing in a few corresponding species.

Differential gene expression analysis

To conduct differential expression and regulation analysis, we identified putative ortholog pairs between the transcripts of the two *Lepidium* species as described in the methods section. We attained two transcriptome datasets, one for *L. campestre* and one for *L. appelianum*, each containing 17,755 transcripts and where each transcript in one species has exactly one

putative orthologous transcript in the other species. We will refer to these transcriptome datasets as our ortholog-transcriptomes in the following. We reassessed completeness of our ortholog-transcriptomes and found that 89.2% of the BUSCOs remained in our ortholog-assembly for *L. campestre* while this value was slightly lower at 89.1% for our *L. appelianum* ortholog-transcriptome.

Reads were mapped independently to the corresponding ortholog-transcriptome and counted using HTSeq-count (Anders et al. 2015b). A principal component analysis was conducted based on the normalized number of reads mapping to the ortholog-transcriptomes. As expected, the replicates from the same species and structure clustered together (Figure 3). The species are separable based on first component which explains 54% of the variance while the structures are separable based on second component which explains 30% variance (Figure 3).

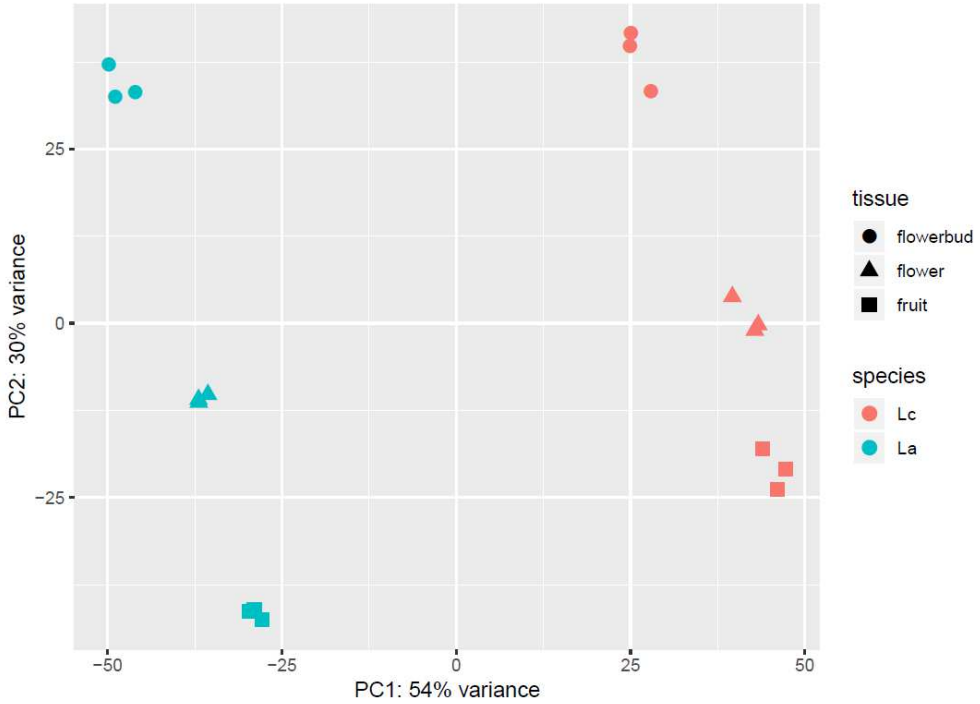


Figure 3: Principal component analysis of gene expression profiles of all samples. Samples from *L. campestre* are shown in red, samples from *L. appelianum* are shown in blue. Samples from floral buds are depicted by circles, samples from flowers by triangles and samples from fruits by squares. PCA shows separation of the two species and the different structures.

To learn more about the differences in fruit development between the *L. campestre* and *L. appelianum*, we analyzed expression in our ortholog-transcriptomes using the programs DESeq2 (Love et al. 2014a) and edgeR (Robinson et al. 2010a). We used a multi-factor design to not only be able to identify differentially expressed genes (DEGs) between the species in the same structure and between structures in the same species, but also to identify genes where the change in expression between the structures is different between the two species. We will refer to the genes identified in the latter analyses as differently differentially expressed genes (DDEGs).

DESeq2 generally identified more DEGs and DDEGs than edgeR, but there is a great overlap of genes identified by both programs (Figure 4). Only this overlap between the two methods will be considered in the following. More DEGs were observed between the same structure of the different species as compared to different structures of the same species. In *L. campestre*, there are similar numbers of DEGs between flower and bud as compared to fruit and flower. In *L. appelianum*, there are more than twice as many DEGs in flowers versus buds as compared to fruits versus flowers (Figure 4). When looking at DEGs in the same structure of the different species, the highest number of DEGs in observed in flowers, followed by fruits and buds.

We also analyzed DDEGs in our dataset, i.e. genes which had a significantly different change in expression in flowers versus buds and in fruits versus flowers, respectively, in *L. appelianum* as compared to *L. campestre*. These genes may have a significantly stronger up- or downregulation in *L. appelianum* as compared to *L. campestre* or these genes may be downregulated in one species and upregulated in the other species. We found 70 DDEGs in flowers versus buds and 158 DDEGs in fruits versus flowers when comparing the two species (Figure 4).

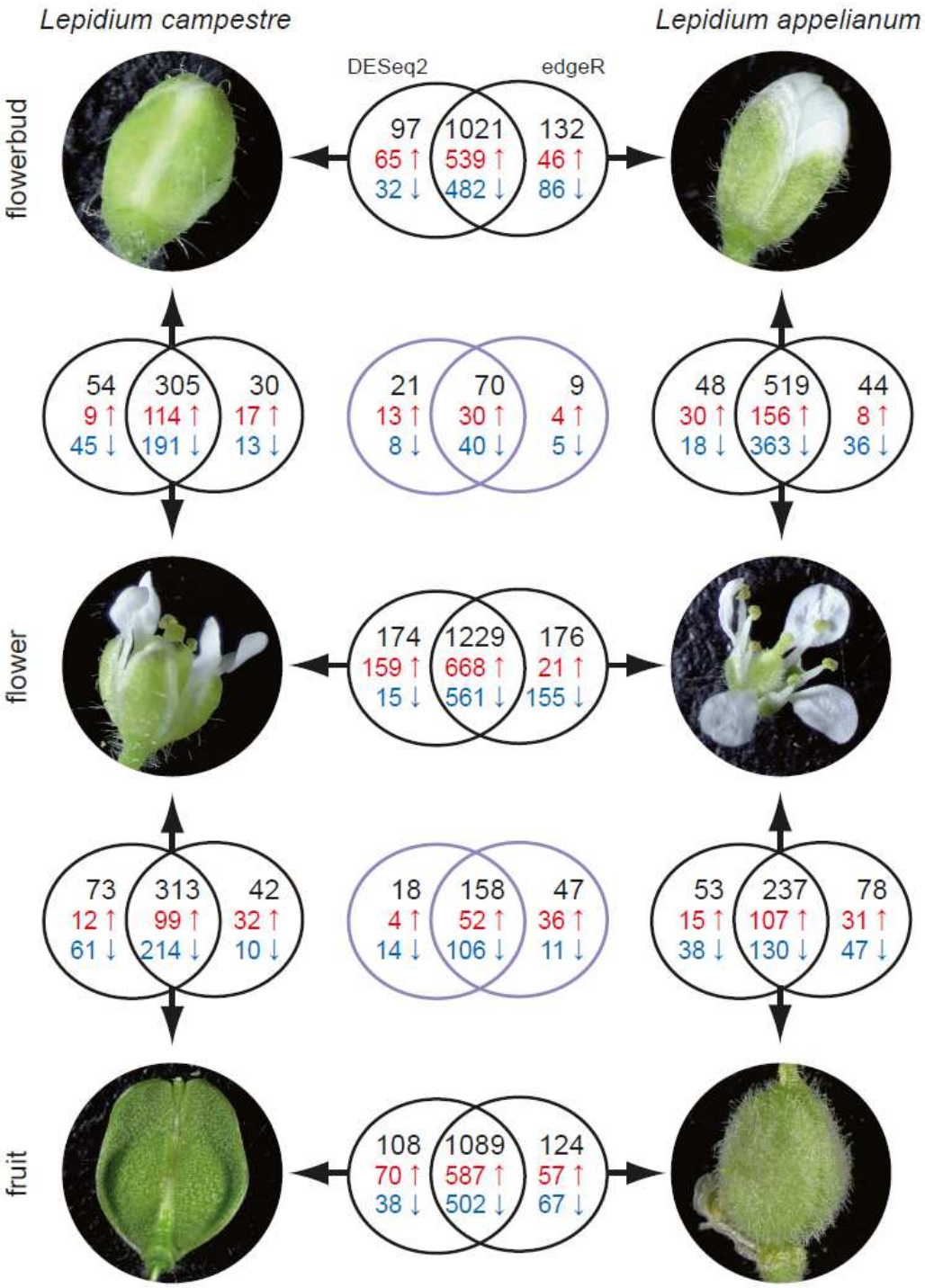


Figure 4: Venn diagrams of the DEGs and DDEGs between different species and different structures. The DEGs and DDEGs were called by the two programs edgeR and DESeq2. Venn diagrams between floral buds, flowers and fruits, respectively, of *L. campestre* and *L. appelianum* represent differentially expressed genes (DEGs) between the two species in the corresponding structure. Venn diagrams between different structures of the same species represent DEGs between those structures in the corresponding species. The two lavender Venn diagramms indicate differently differentially expressed genes (DDEGs) between flower and floral buds and between fruits and flowers,

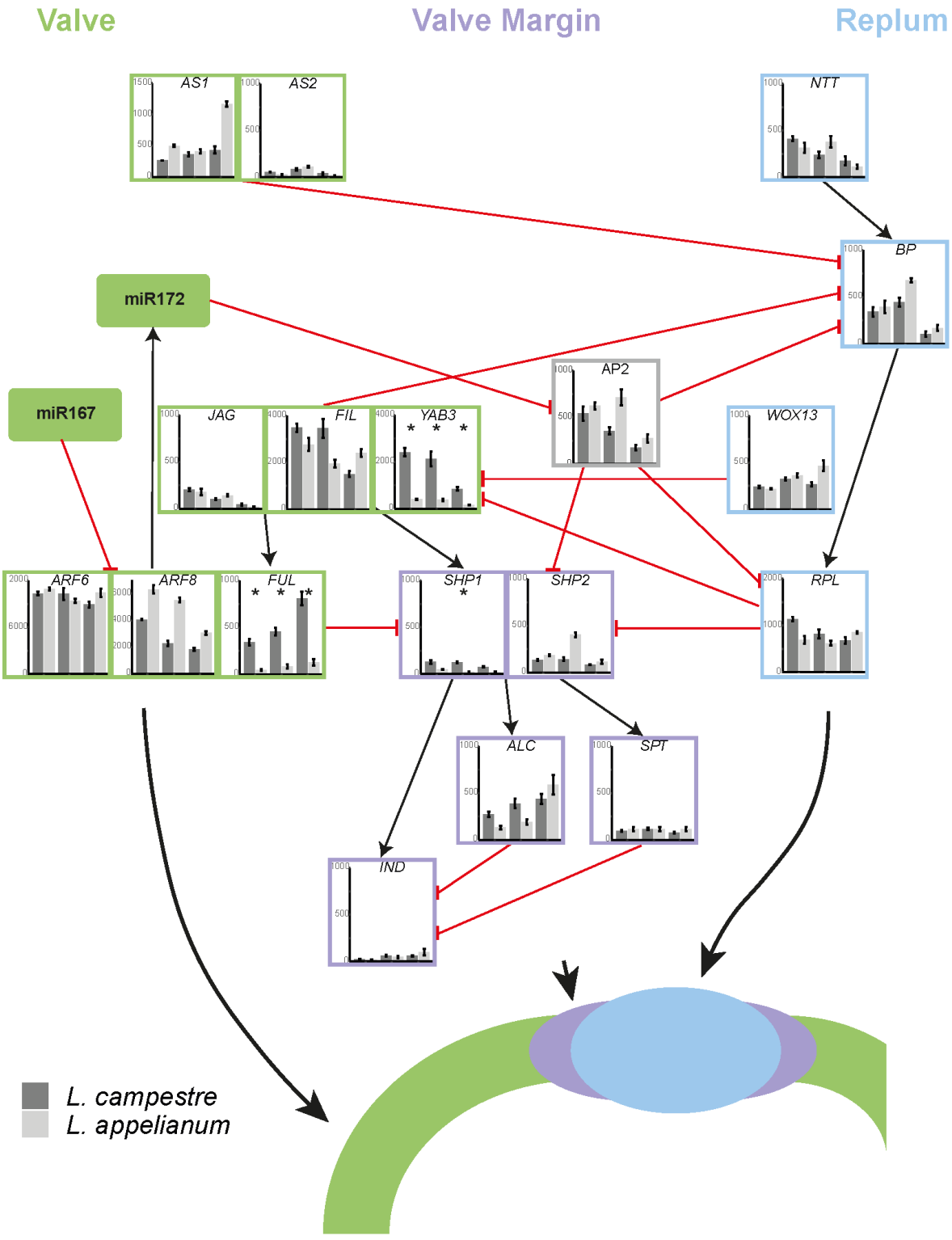
253 respectively, when comparing the two species. Black numbers in the Venn diagram correspond to all
254 DEGs or DDEGs while red numbers represent up- and blue numbers represent downregulated genes.
255
256 We applied the same methods for the identification of DEGs and DDEGs encoding miRNAs.
257 First, we determined orthologs between the miRNAs based on the *A. thaliana* miRNAs they
258 mapped to. For 56 mature miRNAs belonging to 28 miRNA families reads were found in the
259 small RNA data for both species and these mature miRNAs could thus be used for differential
260 expression analyses (Table 2). We will refer to this dataset as our ortholog-miRNAs. All 16
261 miRNA families that are conserved across virtually all species according to (Montes et al. 2014)
262 and 6 out of 21 miRNA families which were classified as moderately conserved belong to our
263 ortholog-miRNAs dataset. Mapping of small RNA reads, counting and differential expression
264 analyses were done as described for the differential expression analysis of the ortholog-
265 transcriptomes.
266 Only one miRNA was found to be encoded by a DEG or DDEG by both programs DESeq2 and
267 edgeR. The miRNA homologous to miR165a-3p, miR165b, miR166a-3p, miR166b-3p,
268 miR166c, miR166d, miR166e-3p, miR166f and miR166g of *Arabidopsis thaliana* (Reinhart et
269 al. 2002) (they all only differ by one nucleotide), which we will refer to as miR165a-3p, was
270 found to be encoded by a DDEG when comparing fruits and flowers. Targets of miR165 and
271 miR166 are HD-Zip transcription factors like PHABULOSA, REVOLUTA and PHAVOLUTA
272 (Rhoades et al. 2002).
273
274 ***Gene Ontology and transcription factor analyses***
275 A number of gene ontology (GO) terms (Ashburner et al. 2000; Gene Ontology Consortium
276 2021) of the category molecular function are significantly over- or underrepresented in the
277 DEGs and DDEGs (Table 3). Among them, the terms protein binding (GO:0005515) and RNA
278 binding (GO:0003723) were underrepresented in two datasets of DEGs. Interestingly, several
279 GO terms related to cell wall synthesis and degradation, i.e. pectinesterase activity
280 (GO:0030599), cellulose synthase (UDP-forming) activity (GO:0016760), polygalacturonase

activity (GO:0004650) and hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553) were overrepresented in different sets of DEGs.

As we were interested in differences in the gene regulatory network involved in fruit dehiscence in the two species, known to be largely composed of transcription factors in *Arabidopsis thaliana* (reviewed in (Ballester and Ferrándiz 2017)), we analyzed genes annotated to have "DNA-binding transcription factor activity" (GO:0003700) in more detail. This set includes transcription factors and transcriptional regulators. For simplicity, we will refer to this dataset as genes encoding transcription factors (TFs).

When comparing flowers and buds, 21 and 28 TFs were DEGs in *L. campestre* and in *L. appelianum*, respectively. Among them, there are 13 TFs that were DEGs comparing these structures in both species, including four genes with known functions in flower development *AGAMOUS-LIKE 104* (*AGL104*) (Adamczyk and Fernandez 2009), *SPOROCYTELESS* (*SPL*, also termed *NOZZLE*) (Balasubramanian and Schneitz 2000), *ORESARA1* (*ORE1*, also termed *ANAC092*, *ATNAC2*, *ATNAC6*) (Gao et al. 2018) and *ZINC FINGER PROTEIN 2* (*ZFP2*) (Cai and Lashbrook 2008) (Table 4). Between fruits and flowers, there are 12 TFs in *L. campestre* and 23 TFs in *L. appelianum* that are DEGs. Five of these genes are DEGs in fruits versus flowers in both species (Table 4). TFs with differential expression between structures in both species are probably those TFs with common functions for flower and fruit development.

When comparing the two species, 43 TFs were DEGs in buds, 68 in flowers and 49 in fruits. Among these TFs, 19 were DEGs in all structures (Table 5). Interestingly, four genes involved in flowering time determination, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 4* (*SPL4*) (Jung et al. 2016), *NUCLEAR FACTOR Y-B2* (*NF-YB2*) (Cao et al. 2014), *NUCLEAR FACTOR Y-B10* (*NF-YB10*) (Wenkel et al. 2006) and *FLOWERING LOCUS C* (*FLC*) (Michaels and Amasino 1999), as well as the fruit development genes *FRUITFUL* (*FUL*) (Gu et al. 1998) and *YABBY3* (*YAB3*) (Dinneny et al. 2005; González-Reig et al. 2012) (Figure 5) were on this list.



309 **Figure 5: Gene regulatory network for the development of valve, valve margin and replum of a**
310 **fruit.** The network is based on what has been determined in *A. thaliana* and is modified after Chavez-
311 Montez *et al.*, 2015. Relative expression levels of genes in *L. campestre* and *L. appelianum* as

determined in this study by transcriptome analysis are shown. Significant differences between *L. appelianum* and *L. campestre* are indicated by asterisks (P ≤ 0.05).

Two TFs were found to be DDEGs when comparing flowers and buds in the two species (Table 6), among them *MASSUGU 2* (*MSG2*, also known as *INDOLE-3-ACETIC ACID INDUCIBLE 19*) (Tatematsu et al. 2004) which has been shown to be involved in stamen filaments development (Tashiro et al. 2009). Comparing fruits and flowers, seven TFs, *PHY-INTERACTING FACTOR 1* (*PIF1*, also known as *PHYTOCHROME INTERACTING FACTOR 3-LIKE 5*) (Huq et al. 2004), *MYB DOMAIN PROTEIN 57* (*MYB57*) (Bender et al. 2013), *TCP FAMILY TRANSCRIPTION FACTOR 4* (*TCP4*, also known as *MATERNAL EFFECT EMBRYO ARREST 35*) (Nag et al. 2009), *BRANCHED 1* (*BRC1*, also known as *TCP FAMILY TRANSCRIPTION FACTOR 18*) (Aguilar-Martínez et al. 2007), *REVEILLE 6* (*RVE6*) (Hsu et al. 2013), *TRIPTYCHON* (*TRY*) (Schnittger et al. 1999) and *OBF BINDING PROTEIN 4* (*OBP4*, also termed *DOF5.4*) (Xu et al. 2016) are DDEGs in *L. appelianum* as compared to *L. campestre*.

Extension of the gene regulatory network for fruit development

We next investigated how the TFs shown to be differentially regulated between fruits and flowers may be involved in the gene regulatory network controlling fruit development (Figure 5). Therefore, we searched for binding sites of the seven TFs identified by chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments in the promotors of the genes known to be involved in fruit development. On ChIP-Hub (Chen et al. 2019), no ChIP-seq data is available for BRC1 and for TRY.

Binding of OBP4 was found in the promotor of all but one of the 18 fruit development genes (Table 7). Binding of RVE6, MYB57, PIF1 and TCP4 was detected in the promotors of 11, 7, 5 and 2 fruit development genes, respectively. PIF1 predominantly binds to the promotors of valve identity genes, with binding to four out of eight valve identity gene promotors and apart from that only binding to one of five valve margin genes. *ARF8* is the only fruit development gene for which none of the differentially regulated genes was found to bind to its promotor. To

the promotors of *YAB3* and *FUL*, which were found to be differentially expressed in all structures between *L. campestre* and *L. appelianum*, binding of TCP4, RVE6 and OBP4 and of MYB57, RVE6 and OBP4, respectively, was found.

Discussion

Transcriptomes and small RNA datasets of L. campestre and L. appelianum are nearly complete

We have sequenced the transcriptomes of floral buds, flowers and fruits of *L. campestre* and *L. appelianum*. Benchmarking of Universal Single-Copy Orthologs (BUSCO) analysis revealed that the transcriptome assemblies of the two species contain more than 94% of the eudicotyledonous “near-universal single-copy orthologs”. This number is similar to or more than that for transcriptome assemblies of other Brassicaceae (Chandler et al. 2020; Yao et al. 2020; Fernandez-Pozo et al. 2021). Furthermore, we found members of all 16 miRNA families that were found to have originated before the emergence of eudicots and conserved in eudicots (Montes et al. 2014). These findings reveal that our transcriptome and small RNA data includes most of the expected transcripts and miRNAs.

Differences in gene expression mainly between floral structures

We identified more DEGs when comparing the same structure between the two species than comparing different structures in the same species (Figure 4). This indicates that gene regulation has diverged between the two species. This is different to what has been observed other flowering plant species, where the correlation of gene expression is higher in the same structure of different species than in different structures of the same species (Chanderbali et al. 2010). However, in this case microarray expression data was analyzed which may select for conserved genes and structures at similar developmental time points were compared.

The highest number of DEGs was observed in flowers, followed by fruits and buds, and the highest number of DDEGs was found in fruits versus flowers as compared to flowers versus floral buds. This indicates that the differences in gene expression between *L. campestre* and *L. appelianum* are most pronounced between flowers and in the transition from flowers to fruits.

This is expected as the developmental program leading to fruit dehiscence or indehiscence needs to be initiated before the fruits are formed. Supportingly, in *Aethionema arabicum*, a plant that develops dehiscent and indehiscent fruits on the very same individual, differences between the fruit types start to occur in flowers after anthesis (Lenser et al. 2018).

A number of GO terms related to cell wall synthesis and degradation, e.g. pectinesterase activity, cellulose synthase activity and polygalacturonase activity were overrepresented in different sets of DEGs. It has been recognized that secondary cell wall formation at the valve margins (Mitsuda and Ohme-Takagi 2008) and degeneration of cell walls in the separation layer are essential processes for fruit dehiscence after the DZ is correctly specified (Ogawa et al. 2009). Hence, the overrepresentation of GO terms related to cell wall synthesis and degradation is not surprising.

Confirmation of previous expression study

In a previous study, we have compared expression of the valve margin genes as well as the valve gene *FUL* and the replum gene *RPL* between *L. campestre* and *L. appelianum* by *in situ* hybridization (Mühlhausen et al. 2013). We showed that their orthologues from *L. campestre* (dehiscent fruits) are similarly expressed as in *A. thaliana* while expression of the respective orthologues is abolished in valve margins of indehiscent *L. appelianum* fruits. Analysis using qRT-PCR revealed that the valve margin genes *IND* and *SHP1* are expressed at a significantly higher level in flowers and early fruits (the fruit stage for which the transcriptome was sequenced here) of *L. campestre* than in *L. appelianum*. Significantly higher expression was confirmed in the present study for *SHP1* in flowers (Figure 5). qRT-PCR analysis revealed significantly higher expression of *SHP2* in flowers and of *ALC* in early fruits in *L. appelianum*. Expression was not significantly different in the present transcriptome analysis, but expression was also found to be higher for *SHP2* in flowers and for *ALC* in fruits. Like qRT-PCR analysis, our transcriptome analysis also found significantly higher expression of *FUL* in fruits of *L. campestre*. Similarly, *RPL* was found to be expressed at a higher level in the flowers of *L. campestre* though the difference was only found to be significant in qRT-PCR but not in transcriptome analysis. *AP2* was found to be expressed at a lower level in flowers and fruits of

L. campestre by both analyses. Again, the difference was significant in qRT-PCR analysis but not in transcriptome analysis. Hence, our transcriptome analysis is in good agreement with the previous qRT-PCR analyses, but differences were less often significant.

Known flower and fruit development genes are differentially expressed

To identify differences in the regulation of flower and fruit development between *L. campestre* and *L. appelianum*, we focused on differentially expressed or differentially differentially expressed transcription factors (TFs). Among 19 TFs which were found to be differentially expressed in all three examined structures, four TFs are involved in flowering time determination. *L. appelianum* and *L. campestre* have different flowering periods according to the Jepson Herbarium (Jepson Flora Project (eds.) 2021, Jepson eFlora, https://ucjeps.berkeley.edu/eflora/, accessed on May 25, 2021), which may be caused by differences in the expression of the identified flowering time genes.

As mentioned above, the fruit development genes *SHP1* and *FUL* were found to be differentially expressed. *FUL* expressed at a significantly lower level in *L. appelianum* in all three structures. In *A. thaliana*, the *ful* knockout mutation causes indehiscence in (Gu et al. 1998; Ferrandiz et al. 2000). Furthermore, *YAB3* (Dinneny et al. 2005; González-Reig et al. 2012) (Figure 5) was also differentially expressed in floral buds, flowers and fruits. FUL is activated by *JAG*, *FIL* and *YAB3* (Dinneny et al. 2005). Concordantly, we found that *YAB3* has a significant higher expression level in *L. campestre* than in *L. appelianum*. Expression of *JAG* and *FIL* is not significantly different between both species. *yab3* single mutants do not have any major defects in dehiscence but *fil yab3* double mutants are largely indehiscent (Dinneny et al. 2005). Hence, decreased expression of *YAB3* in *L. appelianum* as compared to *L. campestre* may have been an important factor for the evolutionary shift from dehiscent to indehiscent fruits in *L. appelianum*. This also shows that there was not only a change in the control of valve margin identity genes but also of the valve identity genes and shifts the causative mutation further upstream in the gene regulatory network of fruit development.

MiR166 is differentially regulated in fruits versus flowers

Our smallRNA sequencing revealed that the miRNA homologous to miR165a-3p, miR165b, miR166a-3p, miR166b-3p, miR166c, miR166d, miR166e-3p, miR166f and miR166g (Reinhart et al. 2002) is encoded by a DDEG when comparing fruits and flowers. Targets of miR165 and miR166 are the mRNAs of HD-Zip transcription factors like PHABULOSA (PHB), REVOLUTA and PHAVOLUTA (Rhoades et al. 2002). Recently, a function of the miR166-PHB module in anther dehiscence has been elucidated (Li et al. 2019). Upregulation of miR166 in the *jba-1D* mutant leads to downregulation of its target gene PHB which results in increased expression of *SPOROCYTELESS/NOZZLE* (*SPL/NZZ*). *jba-1D* mutants do not develop a dehiscence zone in anthers, i.e. overexpression of miR166 leads to indehiscence of anthers. Expression of miR166 in fruits is much higher in *L. appelianum* (indehiscent fruits) than in *L. campestre* (dehiscent fruits), while the opposite is the case in flowers (Figure 6). Hence, miR166 may have a role in the development of indehiscent fruits in *L. appelianum* though the details of the regulation remain to be elucidated.

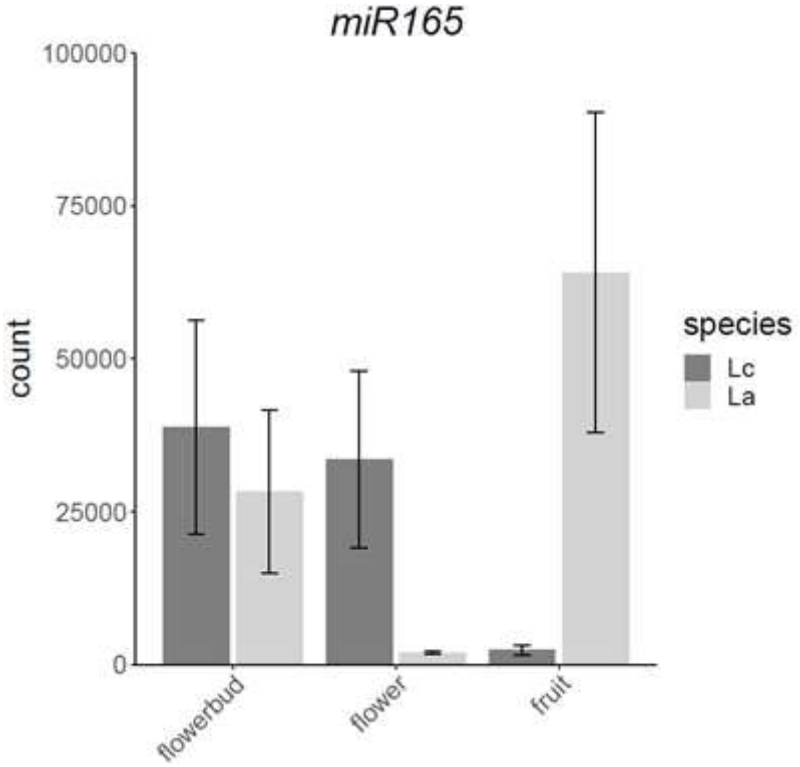


Figure 6: Expression data plot of the miRNA homologous to miR165a-3p of *A. thaliana*. miR165a-3p is identical to miR165b, miR166a-3p, miR166b-3p, miR166c, miR166d, miR166e-3p, miR166f and miR166g such that they cannot be distinguished and hence are summarized here as miR165. Bars

indicate mean normalized count values of reads mapping to miR165 in the corresponding structure and species. Dark and light grey bars represent the mean values for *L. campestre* (Lc) and for *L. appelianum* (La), respectively. The error bars indicate the standard deviation.

BRC1 *and* TCP4 *as candidate genes for the evolutionary shift from dehiscent to indehiscent fruits*

Our transcriptome analysis also identified seven TFs belonging to DDEGs when comparing flowers and fruits (Table 6). PIF1 is a basic helix-loop-helix (bHLH) transcription factor that negatively regulates chlorophyll biosynthesis (Huq et al. 2004); it is involved in a variety of biological processes such as the repression of light-induced seed germination and chlorophyll accumulation in light (Castillon et al. 2007). RVE6 is a MYB gene that controls the pace of the circadian clock together with its close homologs RVE4 and RVE8 (Hsu et al. 2013). The zinc finger gene OBP4 functions in cell cycle progression and cell expansion (Xu et al. 2016) and is involved in root development (Rymen et al. 2017; Ramirez-Parra et al. 2017). So far, involvement of these three factors in flower and fruit development has, to the best of our knowledge, not been reported.

Two other MYB genes, MYB57 and TRY have also been found to be differentially regulated (Table 6). MYB57 functions redundantly with MYB21 and MYB24 to regulate stamen development (Cheng et al. 2009). TRY controls the spacing pattern of trichomes, which are single-celled hairs (Schnittger et al. 1999). Recently it has been found that TRY and other MYB genes of the regulatory network for trichome patterning have been modulated to trigger trichome development in fruits (Arteaga et al. 2021). Hence, these two gene are known to function during flower and fruit development but association with fruit dehiscence is not known so far.

More interestingly, the genes encoding for the two TCP transcription factors BRC1 and TCP4 are DDEGs between fruits and flowers when comparing *L. campestre* and *L. appelianum*. Expression of *BRC1* correlates with bud inhibition (Aguilar-Martínez et al. 2007; Braun et al. 2012) but recently, it has been shown that *BRC1* is neither necessary nor sufficient for bud

inhibition (Seale et al. 2017). Noticeably, it has been hypothesized that *BRC1* may guide fruit morph determination in the dimorphic Brassicaceae plant *Aethionema arabicum* (Lenser et al. 2018). *Ae. arabicum* produces two fruit morphs on the same plant, one of which is dehiscent and the other one is indehiscent. qRT-PCR analyses showed that the expression of *BRC1* in *Ae. arabicum* is high in flowers and decreases strongly in the indehiscent morph but remains at a low level in flowers and fruits of the dehiscent morph. We observe a very similar pattern in our transcriptome analysis for the indehiscent morph *in L. appelianum* and the dehiscent morph in *L. campestre* (Figure 7). It is not known, however, as to how *BRC1* interacts with the genes of the fruit development network (Figure 5).

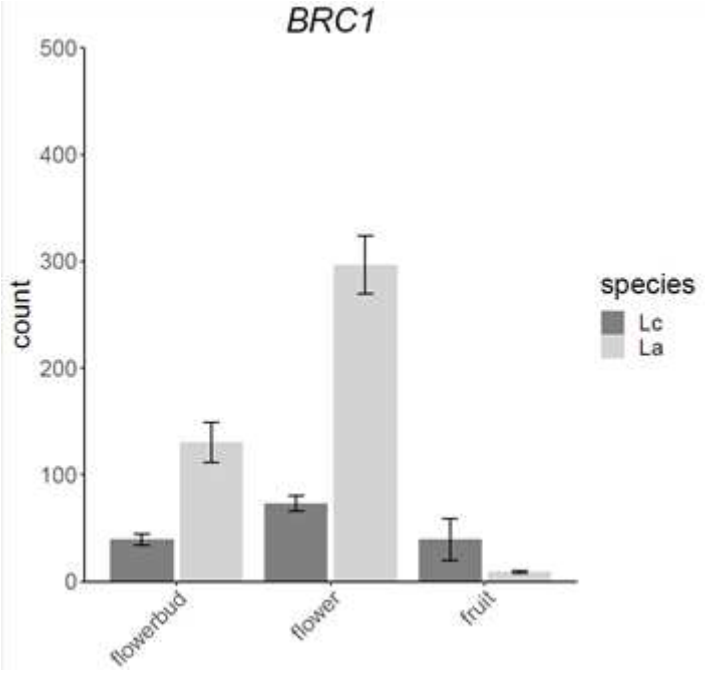


Figure 7: Expression data plot of *BRC1* in *L. campestre* (Lc) and *L. appelianum* (La). Bars indicate mean normalized count values of reads mapping to *BRC1* in the corresponding structure and species. Dark and light grey bars represent the mean values for *L. campestre* (Lc) and for *L. appelianum* (La), respectively. The error bars indicate the standard deviation.

TCP4 has been found to be involved in leaf and flower development as well as in seed oil biosynthesis in *A. thaliana* (Kong et al. 2020; Nag et al. 2009; Palatnik et al. 2003). Furthermore, TCP4 directly activates the expression of *miR167* which targets the TFs ARF6 and ARF8 (Rubio-Somoza and Weigel 2013). This regulation has been hypothesized to be

important for flower maturation, but may also be involved in fruit dehiscence as ARF6 and ARF8 are part of the gene regulatory network of fruit development (Zheng et al. 2019) (Figure 5). Another study found physical interaction of TCP4 and AS2 in yeast-two-hybrid experiments (Li et al. 2012). AS2 has also previously been found to be involved in fruit patterning (Alonso-Cantabrana et al. 2007) (Figure 5). Our analysis of ChIP-seq data on ChIP-Hub (Chen et al. 2019) additionally revealed that TCP4 binds to the promotor of *YAB3* (Table 7), which has been found to be differentially expressed between *L. campestre* and *L. appelianum* in all structures examined. In flowers, TCP4 is expressed at a higher level in *L. appelianum* than in *L. campestre* while the expression pattern is the other way round for *YAB3*. Hence, it is conceivable that TCP4 represses *YAB3* in flowers.

Conclusions

Taken together, our study provides insights into the gene regulatory differences in fruit development between *L. campestre* producing dehiscent fruits and *L. appelianum* forming indehiscent fruits. We confirm differences in the expression of the fruit development genes *SHP2* and *FUL* between the two species and reveal the importance of the valve identity gene *YAB3* for fruit indehiscence in *L. appelianum*. We uncover the microRNA miR166 and the TCP transcription factors BRC1 and TCP4 as new candidates for causing the evolutionary transition from dehiscent to indehiscent fruits in *L. appelianum*.

Methods

Plant material, RNA extraction and sequencing

Seeds of *Lepidium appelianum* (KM 1754) were obtained from J Gaskin, USDA, Fremont County, Wyoming, USA and seeds of *L. campestre* (KM 96) were aquired from the Botanical Garden of the University of Zürich and subsequently mass propagated in the Botanical Garden of Osnabrueck University, Germany. Seeds from these mass propagations were sawn on a mixture of seedling substrate (Kammlott, Kammlott GmbH, Erfurt, Germany)/sand/vermiculite (1–3 mm) (8:1:1) which was supplemented with 1 g L−1 each of Osmocote mini

(www.scotts.com,The Scotts Miracle-Gro Company, Marysville, OH, USA) and Triabon (http://www.compo-expert.com, COMPO Expert GmbH, Münster, Germany). The seeds were placed for 4 days at 4°C for stratification and then put in the greenhouse under a light-dark cycle of 16 hours light, 8 hours dark of artificial light, plus daylight. After 5 weeks in the greenhouse, the plants were vernalized for at least 13 weeks at 4°C with a light-dark cycle of 8 hours light, 16 hours dark. After vernalization, the plants were put back in the greenhouse. Plant material was harvested from two batches of independently grown plants 3 to 5 weeks after the end of vernalization. Plant material was harvested between 12pm and 4pm to minimize the effect of circadian rhythm. Late flower buds, flowers and early fruits were harvested and immediately frozen in liquid nitrogen. Three samples were taken for each species and each structure resulting in 18 samples in total. For each sample, about 100mg plant material was pooled from four individual plants. The material was pulverized in liquid nitrogen using a mortar and pestle. RNA was extracted using Qiazol (Qiagen) according to the manufacturer's instructions. RNA quantity and quality was checked by gel electrophoresis. The samples were sent to the Vienna BioCenter Facility for Next Generation Sequencing where they were quality checked and sequenced on a HiSeqV4. For mRNA sequencing, 125bp paired-end reads were produced and 50bp single-end reads were generated for small RNA sequencing.

Preprocessing of RNA-seq data

Raw reads were corrected using Rcorrector (Song and Florea 2015) with default settings. Uncorrectable reads were excluded using a python script obtained from https://informatics.fas.harvard.edu/best-practices-for-de-novo-transcriptome-assembly-with-trinity.html which was slightly modified for excluding uncorrectable reads from smallRNA libraries. The remaining reads were trimmed with Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using the following settings: `--clip_R1 12, --clip R2 12, --paired, --retain_unpaired, --phred33, --length 36, -q 5, --stringency 5, –e 0.1` for transcriptome reads and the following settings `--phred33, --length 18, --max_length 26, -q 5, --`

545 `stringency 5, -e 0.1, -a adapter` for small RNA reads where `adapter` was replaced
546 by the corresponding adapter sequence identified using FastQC
547 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Thereafter, Poly-A and Poly-T
548 tails were removed from transcriptome reads with PrinSeq (Schmieder and Edwards 2011)
549 using the settings -`trim_tail_left 5` and -`trim_tail_right 5`. Reads that mapped to
550 the genome of *Frankliniella occidentalis* (GenBank: GCF_000697945) or to rRNAs (GenBank:
551 X52320.1), mitochondrial (GenBank: Y08501.2) or chloroplast (GenBank: AP000423.1)
552 sequences from *A. thaliana* as determined using bowtie2 (settings: `--very-sensitive-`
553 `local, --phred33`) (Langmead and Salzberg 2012) were excluded from further analyses
554 from both, the transcriptome and the smallRNA libraries.
555 ***De novo assembly***
556 To simplify *de novo* assembly, also duplicate reads, i.e. reads with the exact same length and
557 sequence, were removed. The remaining reads were assembled using Trinity (Grabherr et al.
558 2011b) with default settings for the two species *L. campestre* and *L. appelianum* separately.
559 To identify remaining contamination in the transcriptome, a BLASTn search was conducted
560 against the nucleotide sequence database of NCBI (nt) using the transcripts in the assembly
561 as query with the option `-max_target_seqs 5`. Transcripts for which the best BLASTn result
562 came from a non-plant species and had an eValue of $E<10^{-10}$ were removed from the
563 transcriptomes. The completeness of the assembled transcriptomes was evaluated using the
564 Benchmarking Universal Single-Copy Orthologs tool BUSCO (Simao et al. 2015) and their
565 accompanying dataset for eudicotyledons with 2121 groups (odb10).
566 ***Separation of chimeras in the assemblies***
567 The initial assemblies contained chimeras composed from two different transcripts. As these
568 chimeras often result from misassembly (Yang and Smith 2013a), we sought to separate
569 chimeras into their separate transcripts. To recognize chimeras, we first conducted a BLASTn
570 search (Altschul et al. 1990) using the transcripts from the *Lepidium* transcriptomes as query
571 and the cDNA sequences of the representative *A. thaliana* gene model as provided by TAIR10
572 (TAIR10_cdna_20110103_representative_gene_model_updated.fasta) as database saving

the best two subjects (i.e. *A. thaliana* cDNAs) for each query (i.e. each transcript from the *Lepidium* transcriptomes) (Figure 8). Using a customized perl-script (Supplementary Data S1) we searched for instances where the two subjects fitted to different regions of the query (i.e. one part of the *Lepidium* transcript has a BLAST hit corresponding to one *A. thaliana* cDNA while another part of the same transcript has a BLAST hit corresponding to another *A. thaliana* cDNA). These instances likely indicate chimeric *Lepidium* transcripts. To identify the best position to split the chimeras, we considered at which position and to what extend the *A. thaliana* cDNAs matched to the *Lepidium* transcripts as shown in Figure 8. Chimeras were split if the overlap was less than 150 nucleotides either in the middle of the overlap or at the positions corresponding to the corrected end and beginning of the involved transcripts (Figure 8).

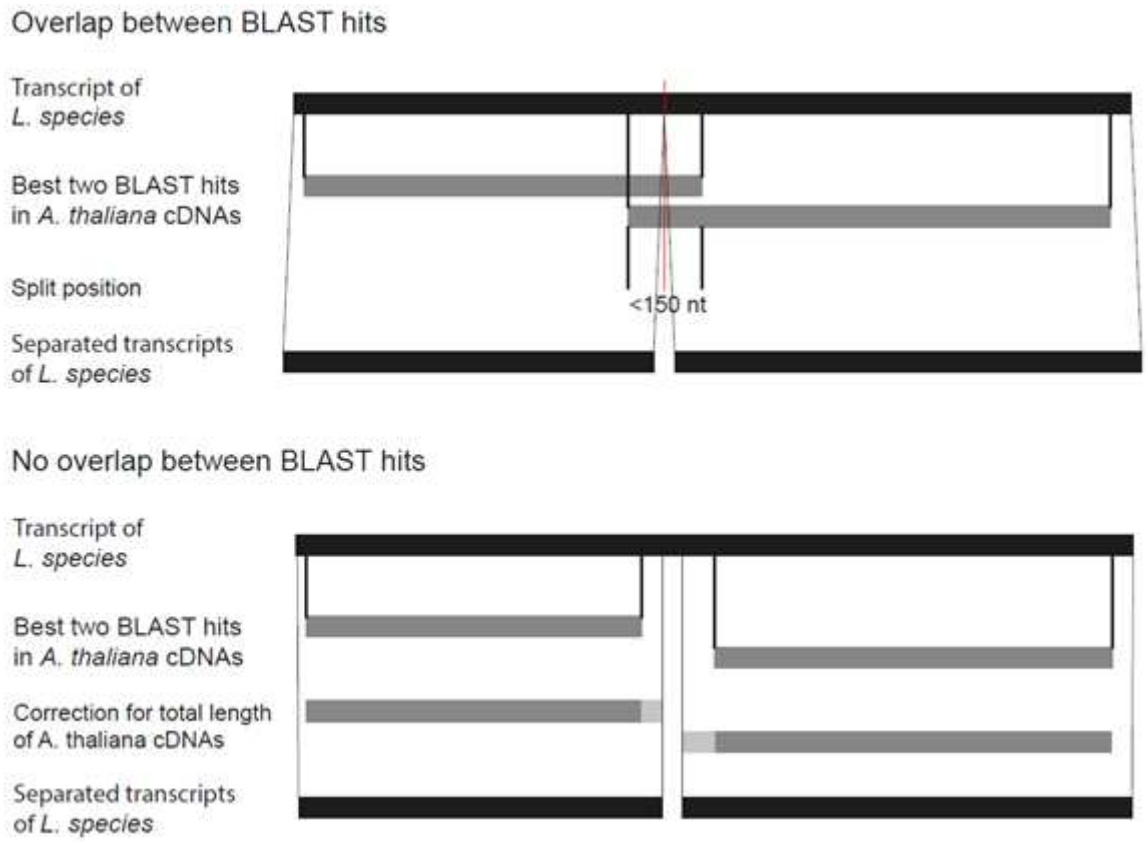


Figure 8: Schematic presentation of the detection and separation of chimeric transcripts in the *Lepidium* transcriptomes. The procedure is slightly different depending on whether the positions of the best two BLAST hits in *A. thaliana* cDNAs overlap on the *Lepidium* transcript or not. If the positions of the best two BLAST hits overlap by less than 150 nucleotides, the *Lepidium* transcript is split in the

middle of the overlap. Otherwise, the beginning and end of the involved transcripts was determined based on the total length of the fitting *A. thaliana* cDNAs.

Identification of miRNAs in smallRNA libraries

SmallRNA reads were mapped to mature miRNAs from *Arabidopsis thaliana* as downloaded from miRBase (Kozomara et al. 2019b) employing bowtie2 with the settings `-N 1, -L 18`. If a read mapped to a specific miRNA from *A. thaliana* this miRNA was considered to be present in the corresponding *Lepidium* species. Mature miRNAs only differing by one nucleotide were combined to avoid multiple mapping during read-counting. To identify novel miRNAs in *Lepidium*, we used ShortStack (Johnson et al. 2016) with the parameters `–foldsize 500, --dicermin 18` and the trinity transcriptome assembly of the corresponding *Lepidium* species as reference "genome". For *L. campestre*, ShortStack was run a second time, this time using the genome sequence of *L. campestre* as available at the National Centre for Biotechnology Information (Sayers et al. 2020) as reference. The stem-loop sequences classified as "N15" or "Y" by ShortStack were used as query sequences for BLAST searches of pre-miRNAs of *A. thaliana* as downloaded from miRBase (Kozomara et al. 2019b) to distinguish known from novel miRNAs. The stem-loop sequences classified as "Y" by ShortStack without similarity to pre-miRNAs of *A. thaliana* were classified as novel *Lepidium* miRNAs.

Determination of orthologs

For transcriptome data, putative ortholog pairs were determined using the reciprocal best hit approach as follows. BLASTn searches were conducted using the transcriptome assembly with chimeras separated of *L. campestre* as query and the transcriptome assembly with chimeras separated of *L. appelianum* as subject and *vice versa*. For each transcript of *L. campestre* the best BLAST hits (having the same eValue, score and alignment length) in *L. appelianum* were recorded and *vice versa*. If a transcript T_c from *L. campestre* had the transcript T_a from *L. appelianum* among its best BLAST hits and transcript T_a from *L. appelianum* had transcript T_c from *L. campestre* in its list of best BLAST hits, these were

617 considered as best reciprocal BLAST hit. Best reciprocal BLAST hits with an alignment length
618 of more than 250 nucleotides and where the length of the shorter sequence was at least 50%
619 of that of the longer sequence were considered as putative ortholog pairs. Additionally, another
620 BLASTn search was conducted using the transcripts in the transcriptomes as query and the
621 *Arabidopsis thaliana* TAIR10 cDNA dataset as database. The set of putative orthologous
622 transcript pairs was pruned such that only one transcript isoform was kept for each species
623 unless different isoforms fitted to different *A. thaliana* genes. The isoform with the longest
624 alignment length between the two species was chosen to be kept. This way, for each transcript
625 in the one transcriptome exactly one transcript in the other transcriptome was kept. We refer
626 to this dataset as the ortholog-transcriptome. The transcripts in the ortholog-transcriptome
627 dataset were named using the TAIR10 identifier of the best BLAST result or numbered if no
628 BLAST result was obtained this way.
629 For the miRNA data, orthologous miRNAs of the two *Lepidium* species were defined as those
630 miRNAs fitting to the same miRNA from *A. thaliana*. Comparison of the novel *Lepidium*
631 miRNAs revealed that none of these was found in both species.
632 ***Read mapping and feature counting***
633 Preprocessed transcriptome and small RNA reads were mapped against ortholog-
634 transcriptome and mature miRNAs, respectively, using bowtie2 (settings: `--very-`
635 `sensitive-local, --phred33` and settings: `--phred33, -N 1, -L 18`, respectively)
636 (Langmead and Salzberg 2012). A custom GFF was generated with one feature for each
637 transcript and miRNA. Mapped reads per feature were then counted using HTSeq-count
638 (Anders et al. 2015a) with the settings `–s no –t transcript –m union`.
639 ***Differential gene expression analysis pipeline***
640 Differentially expressed genes were identified using R (https://www.r-project.org/) and the
641 Bioconductor packages edgeR (Robinson et al. 2010b) and DESeq2 (Love et al. 2014b).
642 Transcript counts were normalized with respect to transcript length. Lowly expressed
643 transcripts with normalized counts and lowly expressed miRNAs with raw counts of less than
644 19 were discarded. Considering the two species *L. campestre* and *L appelianum* and the

structures bud, flower and fruit, the following multi-factor design was used: `species + structure + species:structure`. A Likelihood Ratio Test (LRT) and a quasi-likelihood F-test were conducted in DESeq2 (command: `DESeq(object, test="LRT", reduced=~species + structure)`) and EdgeR (command: `glmQLFit(object, design)`), respectively to identify differentially expressed and differently differentially expressed genes. Only transcripts and miRNAs having a log-fold change to the base of 2 of more than 1 were considered. For DESeq2 the false discovery rate threshold α was set to .001. For principal component analysis, count data was normalized using regularized logarithm with the option `blind=FALSE` in DESeq2 and the principal components were plotted using the plotPCA function in R.

GO enrichment analysis

Gene Ontology (GO) enrichment analysis was conducted on the GO website (http://geneontology.org/) using the PANTHER Overrepresentation Test (Mi et al. 2019). The TAIR10 identifiers of the transcripts in the ortholog-transcriptome were provided as reference list. The TAIR10 identifiers of the transcripts which were identified as significantly differently expressed genes by both programs, DESeq2 and EdgeR, were provided as analyzed list. *Arabidopsis thaliana* was chosen as organism and "GO Molecular function complete" was selected as annotation data set. Enriched GO categories were determined using the Fisher's Exact Test with False Discovery Rate correction.

GO categories and terms were also determined using the AnnotationDbi in R. Transcripts associated with the term "DNA-binding transcription factor activity" were analyzed further.

Promoter analyses

Binding of transcription factors to the promotors of genes involved in fruit opening was analysed using ChIP-Hub (http://www.chip-hub.org/). ChIP-Hub provides access to data on binding sites determined using chromatin immunoprecipitation followed by sequencing (ChIP-seq). On the ChIP-Hub website, *A. thaliana* was chosen as species and binding data was visualized on the WashU EpiGenome Browser. For each fruit development gene, 1,500

nucleotides upstream of the translation start codon were investigated and each occurance of binding of one of the transcription factors found to be differentially regulated was noted.

List of abbreviations

AGL104: AGAMOUS-LIKE 104

ALC: ALCATRAZ

AP2: APETALA2

ARF: auxin-response factor

AS1: ASYMMETRIC LEAVES 1

AS2: ASYMMETRIC LEAVES 2

bHLH: basic helix-loop-helix

BP: BREVIPEDICELLUS

BRC1: BRANCHED1

BUSCO: Benchmarking of Universal Single-Copy Orthologs

ChIP-seq: chromatin immunoprecipitation followed by sequencing

DDEG: differently differentially expressed gene

DEG: differentially expressed gene

DZ: dehiscence zone

FIL: FILAMENTOUS FLOWER

FLC: FLOWERING LOCUS C

FUL: FRUITFULL

GO: gene ontology

IND: INDEHISCENT

JAG: JAGGED

LRT: Likelihood Ratio Test

MSG2: MASSUGU 2

MYB57: MYB DOMAIN PROTEIN 57

NF-YB2: NUCLEAR FACTOR Y-B2

NF-YB10: NUCLEAR FACTOR Y-B10

nt: nucleotide sequence database of NCBI

NTT: NO TRANSMITTING TRACT

NZZ: NOZZLE

OBP4: OBF BINDING PROTEIN 4

PHB: PHABULOSA

PIF1: PHY-INTERACTING FACTOR 1

RPL: REPLUMLESS

RVE6: REVEILLE 6

SHP1: SHATTERPROOF1

SHP2: SHATTERPROOF2

SPL: SPOROCYTELESS

SPL4: SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 4

SPT: SPATULA

TCP4: TCP FAMILY TRANSCRIPTION FACTOR 4

TF: transcription factor

TRY: TRIPTYCHON

WOX13: WUSCHEL-RELATED HOMEOBOX gene 13

YAB3: YABBY3

ORE1: ORESARA1

ZFP2: ZINC FINGER PROTEIN 2

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

728 ***Availability of data and materials***

729 The datasets supporting the conclusions of this article are available at NCBI under the

730 BioProject identifier PRJNA769250 https://www.ncbi.nlm.nih.gov/bioproject/769250.

731 ***Competing interests***

732 The authors declare that they have no competing interests.

733 ***Funding***

734 Part of this work was supported by a grant from the Deutsche Forschungsgemeinschaft to

735 G.T. (TH 417/6-2).

736 ***Authors' contributions***

737 GT conceived the project. LG and GT designed the experiments. KK performed the

738 experiments. LG, NFP and MH analyzed the transcriptome data. LG, GT, MM and SR wrote

739 the manuscript. All authors read and approved the final manuscript.

740 ***Acknowledgements***

741 We thank Klaus Mummenhoff for seeds of *L. appelianum* and *L. campestre*. We are grateful

742 to Heidi Küster for skillful technical assistance as well as Tobias Horst Rogalla and Georgia

743 Daraki for preliminary analyses.

744

745 **References**

746 Adamczyk BJ, Fernandez DE (2009) MIKC* MADS domain heterodimers are required for

747 pollen maturation and tube growth in Arabidopsis. Plant Physiol 149 (4):1713-1723.

748 doi:10.1104/pp.109.135806

749 Aguilar-Martínez JA, Poza-Carrión C, Cubas P (2007) Arabidopsis BRANCHED1 acts as an

750 integrator of branching signals within axillary buds. The Plant Cell 19 (2):458-472

751 Al-Shehbaz I, Mummenhoff K (2011) Stubendorffia and Winklera belong to the expanded

752 Lepidium (Brassicaceae). Edinburgh Journal of Botany 68 (2):165

753 Alonso-Cantabrana H, Ripoll JJ, Ochando I, Vera A, Ferrándiz C, Martínez-Laborda A (2007)

754 Common regulatory networks in leaf and fruit patterning revealed by mutations in the

755 Arabidopsis ASYMMETRIC LEAVES1 gene. Development 134 (14):2663-2671

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215 (3):403-410. doi:10.1016/S0022-2836(05)80360-2

Anders S, Pyl PT, Huber W (2015a) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31 (2):166-169. doi:10.1093/bioinformatics/btu638

Anders S, Pyl PT, Huber W (2015b) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31 (2):166-169

Appel O, Al-Shehbaz I, Kubitzki K, Bayer C (2003) The families and genera of vascular plants. Cruciferae 5:75-174

Arteaga N, Savic M, Méndez-Vigo B, Fuster-Pons A, Torres-Pérez R, Oliveros JC, Picó FX, Alonso-Blanco C (2021) MYB Transcription Factors Drive Evolutionary Innovations in Arabidopsis Fruit Trichome Patterning. The Plant Cell

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25 (1):25-29. doi:10.1038/75556

Axtell MJ (2013) ShortStack: comprehensive annotation and quantification of small RNA genes. Rna 19 (6):740-751

Balasubramanian S, Schneitz K (2000) NOZZLE regulates proximal-distal pattern formation, cell proliferation and early sporogenesis during ovule development in Arabidopsis thaliana. Development 127 (19):4227-4238

Ballester P, Ferrándiz C (2017) Shattering fruits: variations on a dehiscent theme. Current Opinion in Plant Biology 35:68-75

Bender RL, Fekete ML, Klinkenberg PM, Hampton M, Bauer B, Malecha M, Lindgren K, J AM, Perera MA, Nikolau BJ, Carter CJ (2013) PIN6 is required for nectary auxin response and short stamen development. Plant J 74 (6):893-904. doi:10.1111/tpj.12184

784 Bhatt AM, Etchells JP, Canales C, Lagodienko A, Dickinson H (2004) VAAMANA—a BEL1-
785 like homeodomain protein, interacts with KNOX proteins BP and STM and regulates
786 inflorescence stem growth in Arabidopsis. Gene 328:103-111
787 Braun N, de Saint Germain A, Pillot J-P, Boutet-Mercey S, Dalmais M, Antoniadi I, Li X,
788 Maia-Grondard A, Le Signor C, Bouteiller N (2012) The pea TCP transcription factor
789 PsBRC1 acts downstream of strigolactones to control shoot branching. Plant
790 physiology 158 (1):225-238
791 Byrne ME, Groover AT, Fontana JR, Martienssen RA (2003) Phyllotactic pattern and stem
792 cell fate are determined by the Arabidopsis homeobox gene BELLRINGER.
793 Development 130 (17):3941-3950
794 Cai S, Lashbrook CC (2008) Stamen abscission zone transcriptome profiling reveals new
795 candidates for abscission control: enhanced retention of floral organs in transgenic
796 plants overexpressing Arabidopsis ZINC FINGER PROTEIN2. Plant Physiol 146
797 (3):1305-1321. doi:10.1104/pp.107.110908
798 Cao S, Kumimoto RW, Gnesutta N, Calogero AM, Mantovani R, Holt BF, 3rd (2014) A distal
799 CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the
800 FLOWERING LOCUS T promoter and regulates the timing of flowering in
801 Arabidopsis. Plant Cell 26 (3):1009-1017. doi:10.1105/tpc.113.120352
802 Castillon A, Shen H, Huq E (2007) Phytochrome interacting factors: central players in
803 phytochrome-mediated light signaling networks. Trends in plant science 12 (11):514-
804 521
805 Chanderbali AS, Yoo MJ, Zahn LM, Brockington SF, Wall PK, Gitzendanner MA, Albert VA,
806 Leebens-Mack J, Altman NS, Ma H, dePamphilis CW, Soltis DE, Soltis PS (2010)
807 Conservation and canalization of gene expression during angiosperm diversification
808 accompany the origin and evolution of the flower. Proc Natl Acad Sci U S A 107
809 (52):22570-22575. doi:10.1073/pnas.1013395108
810 Chandler JO, Haas FB, Khan S, Bowden L, Ignatz M, Enfissi E, Gawthrop F, Griffiths A,
811 Fraser PD, Rensing SA (2020) Rocket Science: The Effect of Spaceflight on

812 Germination Physiology, Ageing, and Transcriptome of Eruca sativa Seeds. Life 10
813 (4):49
814 Chen D, Fu L-Y, Zhang P, Chen M, Kaufmann K (2019) ChIP-Hub: an Integrative Platform
815 for Exploring Plant Regulome. bioRxiv:768903
816 Cheng H, Song S, Xiao L, Soo HM, Cheng Z, Xie D, Peng J (2009) Gibberellin acts through
817 jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote
818 stamen filament growth in Arabidopsis. PLoS Genet 5 (3):e1000440
819 Cole M, Nolte C, Werr W (2006) Nuclear import of the transcription factor SHOOT
820 MERISTEMLESS depends on heterodimerization with BLH proteins expressed in
821 discrete sub-domains of the shoot apical meristem of Arabidopsis thaliana. Nucleic
822 Acids Research 34 (4):1281-1292
823 Dardick C, Callahan AM (2014) Evolution of the fruit endocarp: molecular mechanisms
824 underlying adaptations in seed protection and dispersal strategies. Frontiers in plant
825 science 5:284
826 Dinneny JR, Weigel D, Yanofsky MF (2005) A genetic framework for fruit patterning in
827 Arabidopsis thaliana. Development 132 (21):4687-4696
828 Fernandez-Pozo N, Metz T, Chandler JO, Gramzow L, Mérai Z, Maumus F, Scheid OM,
829 Theißen G, Schranz ME, Leubner-Metzger G (2021) Aethionema arabicum genome
830 annotation using PacBio full-length transcripts provides a valuable resource for seed
831 dormancy and Brassicaceae evolution research. The Plant Journal:-
832 Ferrandiz C, Liljegren SJ, Yanofsky MF (2000) Negative regulation of the SHATTERPROOF
833 genes by FRUITFULL during Arabidopsis fruit development. Science 289 (5478):436-
834 438
835 Gao Z, Daneva A, Salanenka Y, Van Durme M, Huysmans M, Lin Z, De Winter F, Vanneste
836 S, Karimi M, Van de Velde J, Vandepoele K, Van de Walle D, Dewettinck K,
837 Lambrecht BN, Nowack MK (2018) KIRA1 and ORESARA1 terminate flower
838 receptivity by promoting cell death in the stigma of Arabidopsis. Nat Plants 4 (6):365-
839 375. doi:10.1038/s41477-018-0160-7

840 Gene Ontology Consortium T (2021) The Gene Ontology resource: enriching a GOld mine.
841 Nucleic Acids Res 49 (D1):D325-d334. doi:10.1093/nar/gkaa1113
842 González-Reig S, Ripoll JJ, Vera A, Yanofsky MF, Martínez-Laborda A (2012) Antagonistic
843 gene activities determine the formation of pattern elements along the mediolateral
844 axis of the Arabidopsis fruit. PLoS Genet 8 (11):e1003020
845 González VL, Devine AM, Trizna M, Mulcahy DG, Barker KB, Coddington JA (2018) Open
846 access genomic resources for terrestrial arthropods. Curr Opin Insect Sci 25:91-98.
847 doi:10.1016/j.cois.2017.12.003
848 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
849 Raychowdhury R, Zeng Q (2011a) Trinity: reconstructing a full-length transcriptome
850 without a genome from RNA-Seq data. Nature biotechnology 29 (7):644
851 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
852 Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di
853 Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011b)
854 Full-length transcriptome assembly from RNA-Seq data without a reference genome.
855 Nat Biotechnol 29 (7):644-652. doi:10.1038/nbt.1883
856 Groszmann M, Paicu T, Alvarez JP, Swain SM, Smyth DR (2011) SPATULA and
857 ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for
858 Arabidopsis gynoecium and fruit development. The Plant Journal 68 (5):816-829
859 Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R (1998) The FRUITFULL MADS-box gene
860 mediates cell differentiation during Arabidopsis fruit development. Development 125
861 (8):1509-1517
862 Hall JC, Sytsma KJ, Iltis HH (2002) Phylogeny of Capparaceae and Brassicaceae based on
863 chloroplast sequence data. American Journal of Botany 89 (11):1826-1842
864 Hsu PY, Devisetty UK, Harmer SL (2013) Accurate timekeeping is controlled by a cycling
865 activator in Arabidopsis. Elife 2:e00473

866 Huq E, Al-Sady B, Hudson M, Kim C, Apel K, Quail PH (2004) Phytochrome-interacting
867 factor 1 is a critical bHLH regulator of chlorophyll biosynthesis. Science 305
868 (5692):1937-1941. doi:10.1126/science.1099728
869 Initiative TAG (2000) Analysis of the genome sequence of the flowering plant Arabidopsis
870 thaliana. Nature 408 (6814):796-815. doi:10.1038/35048692
871 Johnson NR, Yeoh JM, Coruh C, Axtell MJ (2016) Improved Placement of Multi-mapping
872 Small RNAs. G3 6 (7):2103-2111. doi:10.1534/g3.116.030452
873 Jung JH, Lee HJ, Ryu JY, Park CM (2016) SPL3/4/5 Integrate Developmental Aging
874 and Photoperiodic Signals into the FT-FD Module in Arabidopsis Flowering. Mol Plant
875 9 (12):1647-1659. doi:10.1016/j.molp.2016.10.014
876 Kong Q, Singh SK, Mantyla JJ, Pattanaik S, Guo L, Yuan L, Benning C, Ma W (2020)
877 Teosinte branched1/cycloidea/proliferating cell factor4 interacts with wrinkled1 to
878 mediate seed oil biosynthesis. Plant Physiology 184 (2):658-665
879 Kozomara A, Birgaoanu M, Griffiths-Jones S (2019a) miRBase: from microRNA sequences
880 to function. Nucleic acids research 47 (D1):D155-D162
881 Kozomara A, Birgaoanu M, Griffiths-Jones S (2019b) miRBase: from microRNA sequences
882 to function. Nucleic Acids Res 47 (D1):D155-D162. doi:10.1093/nar/gky1141
883 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9
884 (4):357-359. doi:10.1038/nmeth.1923
885 Lenser T, Tarkowská D, Novák O, Wilhelmsson PK, Bennett T, Rensing SA, Strnad M,
886 Theißen G (2018) When the BRANCHED network bears fruit: how carpic dominance
887 causes fruit dimorphism in Aethionema. The Plant Journal 94 (2):352-371
888 Li X, Lian H, Zhao Q, He Y (2019) MicroRNA166 monitors SPOROCYTELESS/NOZZLE for
889 building of the anther internal boundary. Plant physiology 181 (1):208-220
890 Li Z, Li B, Shen WH, Huang H, Dong A (2012) TCP transcription factors interact with AS2 in
891 the repression of class-I KNOX genes in Arabidopsis thaliana. The Plant Journal 71
892 (1):99-107

893 Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000)
894 SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. Nature
895 404 (6779):766-770
896 Liljegren SJ, Roeder AH, Kempin SA, Gremski K, Østergaard L, Guimil S, Reyes DK,
897 Yanofsky MF (2004) Control of fruit patterning in Arabidopsis by INDEHISCENT. Cell
898 116 (6):843-853
899 Lorts CM, Briggeman T (2008) Evolution of fruit types and seed dispersal: a phylogenetic
900 and ecological snapshot. Journal of systematics and evolution 46 (3):396-404
901 Love MI, Huber W, Anders S (2014a) Moderated estimation of fold change and dispersion for
902 RNA-seq data with DESeq2. Genome biology 15 (12):1-21
903 Love MI, Huber W, Anders S (2014b) Moderated estimation of fold change and dispersion for
904 RNA-seq data with DESeq2. Genome Biol 15 (12):550. doi:10.1186/s13059-014-
905 0550-8
906 Marsch-Martínez N, Zúñiga-Mayo VM, Herrera-Ubaldo H, Ouwerkerk PB, Pablo-Villa J,
907 Lozano-Sotomayor P, Greco R, Ballester P, Balanzá V, Kuijt SJ, Meijer AH, Pereira
908 A, Ferrándiz C, de Folter S (2014) The NTT transcription factor promotes replum
909 development in Arabidopsis fruits. Plant J 80 (1):69-81. doi:10.1111/tpj.12617
910 Meakin PJ, Roberts JA (1990) Dehiscence of Fruit in Oilseed Rape (Brassica nap us L.) II.
911 THE ROLE OF CELL WALL DEGRADING ENZYMES AND ETHYLENE. Journal of
912 Experimental Botany 41 (8):1003-1011
913 Meakin PJ, Roberts JA (1991) Anatomical and biochemical changes associated with the
914 induction of oilseed rape (Brassica napus) pod dehiscence by Dasineura brassicae
915 (Winn.). Annals of Botany 67 (3):193-197
916 Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD (2019) PANTHER version 14: more
917 genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools.
918 Nucleic acids research 47 (D1):D419-D426
919 Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain
920 protein that acts as a repressor of flowering. The Plant Cell 11 (5):949-956

921 Mitsuda N, Ohme-Takagi M (2008) NAC transcription factors NST1 and NST3 regulate pod
922 shattering in a partially redundant manner by promoting secondary wall formation
923 after the establishment of tissue identity. The Plant Journal 56 (5):768-778
924 Montes RAC, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, Marsch-Martínez N,
925 Meyers BC, Green PJ, de Folter S (2014) Sample sequencing of vascular plants
926 demonstrates widespread conservation and divergence of microRNAs. Nature
927 communications 5 (1):1-15
928 Mühlhausen A, Lenser T, Mummenhoff K, Theißen G (2013) Evidence that an evolutionary
929 transition from dehiscent to indehiscent fruits in L epidium (B rassicaceae) was
930 caused by a change in the control of valve margin identity genes. The Plant Journal
931 73 (5):824-835
932 Nag A, King S, Jack T (2009) miR319a targeting of TCP4 is critical for petal growth and
933 development in Arabidopsis. Proceedings of the National Academy of Sciences 106
934 (52):22534-22539
935 Ogawa M, Kay P, Wilson S, Swain SM (2009) ARABIDOPSIS DEHISCENCE ZONE
936 POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are
937 polygalacturonases required for cell separation during reproductive development in
938 Arabidopsis. The Plant Cell 21 (1):216-233
939 Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control
940 of leaf morphogenesis by microRNAs. Nature 425 (6955):257-263
941 Rajani S, Sundaresan V (2001) The Arabidopsis myc/bHLH gene ALCATRAZ enables cell
942 separation in fruit dehiscence. Current Biology 11 (24):1914-1922
943 Ramirez-Parra E, Perianez-Rodriguez J, Navarro-Neila S, Gude I, Moreno-Risueno MA, Del
944 Pozo JC (2017) The transcription factor OBP 4 controls root growth and promotes
945 callus formation. New Phytologist 213 (4):1787-1801
946 Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants.
947 Genes & development 16 (13):1616-1626

948 Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant
949 microRNA targets. cell 110 (4):513-520
950 Ripoll JJ, Bailey LJ, Mai Q-A, Wu SL, Hon CT, Chapman EJ, Ditta GS, Estelle M, Yanofsky
951 MF (2015) microRNA regulation of fruit growth. Nature plants 1 (4):1-9
952 Ripoll JJ, Roeder AH, Ditta GS, Yanofsky MF (2011) A novel role for the floral homeotic gene
953 APETALA2 during Arabidopsis fruit development. Development 138 (23):5167-5176
954 Robinson MD, McCarthy DJ, Smyth GK (2010a) edgeR: a Bioconductor package for
955 differential expression analysis of digital gene expression data. Bioinformatics 26
956 (1):139-140
957 Robinson MD, McCarthy DJ, Smyth GK (2010b) edgeR: a Bioconductor package for
958 differential expression analysis of digital gene expression data. Bioinformatics 26
959 (1):139-140. doi:10.1093/bioinformatics/btp616
960 Roeder AH, Ferrándiz C, Yanofsky MF (2003) The role of the REPLUMLESS homeodomain
961 protein in patterning the Arabidopsis fruit. Current biology 13 (18):1630-1635
962 Romera-Branchat M, Ripoll JJ, Yanofsky MF, Pelaz S (2013) The WOX 13 homeobox gene
963 promotes replum formation in the Arabidopsis thaliana fruit. The Plant Journal 73
964 (1):37-49
965 Rubio-Somoza I, Weigel D (2013) Coordination of flower maturation by a regulatory circuit of
966 three microRNAs. PLoS Genet 9 (3):e1003374
967 Rymen B, Kawamura A, Schäfer S, Breuer C, Iwase A, Shibata M, Ikeda M, Mitsuda N,
968 Koncz C, Ohme-Takagi M (2017) ABA suppresses root hair growth via the OBP4
969 transcriptional regulator. Plant Physiology 173 (3):1750-1762
970 Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, Funk K, Ketter A, Kim S,
971 Kimchi A, Kitts PA, Kuznetsov A, Lathrop S, Lu Z, McGarvey K, Madden TL, Murphy
972 TD, O'Leary N, Phan L, Schneider VA, Thibaud-Nissen F, Trawick BW, Pruitt KD,
973 Ostell J (2020) Database resources of the National Center for Biotechnology
974 Information. Nucleic Acids Res 48 (D1):D9-D16. doi:10.1093/nar/gkz899

975 Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic
976 datasets. Bioinformatics 27 (6):863-864. doi:10.1093/bioinformatics/btr026
977 Schnittger A, Folkers U, Schwab B, Jürgens G, Hülskamp M (1999) Generation of a spacing
978 pattern: the role of TRIPTYCHON in trichome patterning in Arabidopsis. The Plant
979 Cell 11 (6):1105-1116
980 Seale M, Bennett T, Leyser O (2017) BRC1 expression regulates bud activation potential but
981 is not necessary or sufficient for bud growth inhibition in Arabidopsis. Development
982 144 (9):1661-1673
983 Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO:
984 assessing genome assembly and annotation completeness with single-copy
985 orthologs. Bioinformatics 31 (19):3210-3212. doi:10.1093/bioinformatics/btv351
986 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO:
987 assessing genome assembly and annotation completeness with single-copy
988 orthologs. Bioinformatics 31 (19):3210-3212. doi:10.1093/bioinformatics/btv351
989 Smith HM, Hake S (2003) The interaction of two homeobox genes, BREVIPEDICELLUS and
990 PENNYWISE, regulates internode patterning in the Arabidopsis inflorescence. The
991 Plant Cell 15 (8):1717-1727
992 Song L, Florea L (2015) Rcorrector: efficient and accurate error correction for Illumina RNA-
993 seq reads. GigaScience 4:48. doi:10.1186/s13742-015-0089-y
994 Spence J, Vercher Y, Gates P, Harris N (1996) ‘Pod shatter’in Arabidopsis thalianaBrassica
995 napus and B. juncea. Journal of Microscopy 181 (2):195-203
996 Tashiro S, Tian Ce, Watahiki MK, Yamamoto KT (2009) Changes in growth kinetics of
997 stamen filaments cause inefficient pollination in massugu2, an auxin insensitive,
998 dominant mutant of Arabidopsis thaliana. Physiologia plantarum 137 (2):175-187
999 Tatematsu K, Kumagai S, Muto H, Sato A, Watahiki MK, Harper RM, Liscum E, Yamamoto
1000 KT (2004) MASSUGU2 encodes Aux/IAA19, an auxin-regulated protein that functions
1001 together with the transcriptional activator NPH4/ARF7 to regulate differential growth

responses of hypocotyl and formation of lateral roots in Arabidopsis thaliana. Plant Cell 16 (2):379-393. doi:10.1105/tpc.018630

Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. Plant Cell 18 (11):2971-2984. doi:10.1105/tpc.106.043299

Xu P, Chen H, Ying L, Cai W (2016) AtDOF5. 4/OBP4, a DOF transcription factor gene that negatively regulates cell cycle progression and cell expansion in Arabidopsis thaliana. Scientific reports 6 (1):1-13

Yang Y, Smith SA (2013a) Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. BMC Genomics 14:328. doi:10.1186/1471-2164-14-328

Yang Y, Smith SA (2013b) Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. BMC genomics 14 (1):1-11

Yao S, Liang F, Gill RA, Huang J, Cheng X, Liu Y, Tong C, Liu S (2020) A global survey of the transcriptome of allopolyploid Brassica napus based on single-molecule long-read isoform sequencing and Illumina-based RNA sequencing data. The Plant Journal 103 (2):843-857

Zheng L, Nagpal P, Villarino G, Trinidad B, Bird L, Huang Y, Reed JW (2019) miR167 limits anther growth to potentiate anther dehiscence. Development 146 (14)

Tables

Table 1: Number of reads obtained after sequencing and after correction and pruning steps.

Experiment	Species	Structure	Replicate	Raw reads	Uncorrectable, unpaired reads removed	Thrips and organelle sequences removed
mRNA	L. campestre	Bud	1	56,364,306	47,437,984	42,661,682
			2	52,626,578	44,103,956	42,345,608
			3	46,984,896	38,458,226	35,556,348
		Flower	1	53,973,840	45,071,412	42,388,254
			2	54,176,184	43,661,258	40,639,670
			3	47,473,062	37,866,352	33,766,726

		Fruit	1	66,540,836	59,074,352	48,284,508
			2	67,087,830	57,231,738	45,811,592
			3	57,259,526	48,363,464	41,299,370
	L. appelianum	Bud	1	61,044,624	51,099,508	47,872,858
			2	57,117,360	45,964,748	41,954,354
			3	56,803,334	47,261,624	42,970,772
		Flower	1	54,109,008	45,806,166	41,765,762
			2	60,218,542	49,851,062	46,492,188
			3	59,056,126	49,511,932	46,039,554
		Fruit	1	51,115,752	42,013,148	38,781,852
			2	53,010,164	42,607,090	40,845,816
			3	58,202,346	48,358,214	43,294,808
smallRNA	L. campestre	Bud	1	12,317,448	11,640,550	1,065,134
			2	12,888,040	11,434,250	1,234,296
			3	13,084,802	18,556,186	4,545,624
		Flower	1	13,199,751	12,576,208	1,321,440
			2	14,250,796	21,800,894	4,334,830
			3	11,931,106	10,446,892	991,708
		Fruit	1	12,600,710	10,038,752	908,384
			2	12,171,462	10,194,186	430,032
			3	12,245,803	15,075,818	1,955,400
	L. appelianum	Bud	1	11,107,988	7,005,858	444,744
			2	12,258,200	7,404,808	1,044,044
			3	12,371,624	11,155,240	1,227,638
		Flower	1	11,665,136	7,039,858	359,098
			2	10,806,199	6,161,922	279,758
			3	10,702,759	5,090,392	250,238
		Fruit	1	11,180,533	10,409,022	1,667,408
			2	11,088,927	10,542,716	1,354,590
			3	12,727,467	16,904,122	3,445,396

Table 2: Mature miRNAs identified in short read data of *L. campestre* and *L. appelianum*. bold conserved miRNAs according to Chavez-Montez et al., 2014, bold and italic moderately conserved miRNAs according to Chavez-Montez et al., 2014

miRNA family	*L. campestre* mature miRNA	ShortStack genome	ShortStack transcriptome	*L. appelianum* mature miRNA	ShortStack transcriptome
miR156a-3p/miR156c-3p	x	miR156c		x	miR156a
miR156a-5p/miR156b-5p/ miR156c-5p/miR156d-5p/ miR156e/miR156f-5p/ miR156g/miR156h/ miR156i/miR156j	x	miR156e, miR156b, miR156c		x	miR156a, miR156j
miR156b-3p	x	miR156b		x	
miR156d-3p				x	
miR157a-3p/miR157b-3p	x			x	
miR157a-5p/miR157b-5p/ miR157c-5p/miR157d	x		miR157c	x	
miR157c-3p	x		miR157c	x	

miR158a-3p/miR158b	x			x	miR158a
miR159a/miR159b-3p	x			x	
miR159c	x			x	
miR160a-3p	x	miR160a		x	
miR160a-5p/miR160b/ miR160c-5p	x	miR160a, miR160c	miR160b	x	
miR160c-3p	x	miR160c			
miR161.1	x				
miR161.2	x				
miR162a-3p/miR162b-3p	x	miR162b		x	miR162b
miR162a-5p/miR162b-5p	x	miR162b		x	miR162b
miR164a/miR164b-5p/ miR164c-5p	x	miR164a		x	
miR164b-3p	x			x	
miR164c-3p	x			x	
miR165a-3p/miR165b/ miR166a-3p/miR166b-3p/ miR166c/miR166d/ miR166e-3p/miR166f/miR166g	x	miR165b, miR166d, miR166g		x	miR166a, miR166b, miR166e, miR166f
miR165a-5p	x			x	
miR166a-5p/miR166b-5p	x			x	miR166a, miR166b
miR166e-5p	x			x	miR166e
miR167a-3p	x			x	
miR167a-5p/miR167b/miR167d	x	miR167b		x	
miR167c-5p	x	miR167c		x	
miR168a-3p	x		miR168a	x	miR168a
miR168a-5p/miR168b-5p	x		miR168a	x	miR168a
miR169a-5p/miR169b-5p/ miR169c	x			x	
miR169d/miR169e/ miR169f-5p/miR169g-5p	x			x	
miR169f-3p	x			x	
miR170-5p/miR171a-5p	x	miR171a		x	miR170, miR171a
miR171a-3p	x	miR171a		x	
miR171b-5p/miR171c-5p	x		miR171b	x	miR171b, miR171c
miR171b-3p/miR171c-3p	x		miR171b	x	miR171b, miR171c
miR172a/miR172b-3p/ miR172c/miR172d-3p/ miR172e-3p	x	miR172b (partial)	miR172e	x	miR172b (partial)
miR172b-5p/miR172e-5p	x	miR172b (partial)	miR172e	x	miR172b (partial)
miR172d-5p	x			x	
miR2111b-3p	x				
miR319a/miR319b	x			x	
miR319c	x			x	
miR390a-3p	x			x	
miR390a-5p/miR390b-5p	x			x	
miR390b-3p	x			x	
miR393a-3p/miR393b-3p	x		miR393b	x	miR393b
miR393a-5p/miR393b-5p	x		miR393b	x	miR393b
miR394a/miR394b-5p	x			x	miR394b
miR395a/miR395b/miR395c/ miR395d/miR395e/miR395f	x			x	miR395d, miR395f
miR396a-3p	x	miR396a		x	
miR396a-5p/miR396b-5p	x	miR396a		x	
miR396b-3p	x			x	
miR398a-3p/miR398b-3p/ miR398c-3p				x	miR398b
miR399a/miR399b/miR399c-3p	x	miR399a		x	
miR399f				x	
miR403-3p	x			x	
miR403-5p	x				

miR408-3p	x			x	
miR408-5p	x			x	
miR472			miR472		
miR8174				x	
miR8175	x			x	
miR824-3p	x			x	
miR824-5p	x			x	
miR827	x		miR827	x	
miR845a	x				
miR845b	x				
miR858a/miR858b	x			x	
miR863-5p	x				

Table 3: Gene ontology (GO) terms significantly over- or underrepresented in DEGs and DDEGs. Terms of the category molecular function of GO were analysed.

Dataset	GO term	Fold enrichment	FDR
La vs. Lc bud	protein binding (GO:0005515)	0.74	3.43E-02
	transferase activity, transferring phosphorus-containing groups (GO:0016772)	0.44	2.84E-02
La vs. Lc flower	none		
La vs. Lc fruit	none		
Lc flower vs. bud	pectinesterase activity (GO:0030599)	10.69	1.48E-02
	RNA binding (GO:0003723)	0.15	8.99E-03
La flower vs. bud	sodium:proton antiporter activity (GO:0015385)	14.17	2.35E-04
	cellulose synthase (UDP-forming) activity (GO:0016760)	10.52	3.18E-02
	polygalacturonase activity (GO:0004650)	8.13	8.50E-03
	iron ion binding (GO:0005506)	3.12	3.48E-02
	oxidoreductase activity acting on paired donors with incorporation or reduction of molecular oxygen (GO:0016705)	2.89	4.07E-02
	protein binding (GO:0005515)	0.71	3.99E-02
	RNA binding (GO:0003723)	0.17	1.30E-03
Lc fruit vs. flower	heme binding (GO:0020037)	5.11	9.54E-04
	hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553)	4.28	1.15E-03
La fruit vs. flower	none		
flower vs. bud	acid-amino acid ligase activity (GO:0016881)	40.35	1.53E-02
fruit vs. flower	none		

Table 4: Differentially expressed genes in different structures annotated as "DNA-binding transcription factor activity" (GO:0003700).

	Ortholog ID	Ortholog name	Ortholog description (based on TAIR)	Reg. L.c.	Reg. L.a.
flower vs. bud	AT1G22130.1	AGL104	Pollen development and pollen tube growth	-3,0	-3,6
	AT1G61110.1	anac025, NAC025	Endosperm cell expansion during germination	-3,6	-3,2
	AT1G69490.1	ANAC029, ATNAP, NAP	Leaf senescence (Guo and Gan, 2006), drought stress response (Sakuraba et al., 2015)	3,4	2,8
	AT2G47190.1	ATMYB2, MYB2	Salt tolerance, Phosphate Starvation Response (Baek et al., 2013), Abscisic Acid Signaling (Abe et al., 2003), Plant Senescence (Guo and Gan, 2011)	3,7	2,6
	AT3G04070.1	anac047, NAC047	Flood induced leaf movement (Rauf et al, 2013)	3,3	3,8
	AT3G23050.1	AXR2, IAA7	Auxin response (Timpte et al., 1994), shoot and root gravitopism (Timpte et al., 1992)	2,0	3,1
	AT3G58120.1	ATBZIP61, BZIP61	n.a.	-3,7	-2,5
	AT4G10240.1	bbx23	Temperature-induced hypocotyl elongation together with BBX18 (Ding et al. 2018), photomorphogenesis activated by PIF1 and PIF3 (Zhang et al., 2017)	-4,8	-6,3
	AT4G27330.1	NZZ, SPL	Initiation of micro- and megagametogenesis, patterning of the ovule, differentiation of primary sporogenous cells into microsporocytes, regulation of anther cell differentiation	-7,9	-8,7
	AT4G28500.1	ANAC073, NAC073, SND2	Secondary cell wall development (Hussey et al., 2011), phloem development (Kim et al., 2020)	-2,8	-2,9
	AT5G13180.1	ANAC083, NAC083, VNI2	Xylem vessel formation, leaf senescence (Yang et al., 2011)	2,7	3,5
	AT5G39610.1	ANAC092, ATNAC2, ATNAC6, NAC2, NAC6, ORE1	Leaf senescence (Kim et al., 2018), Termination of flower receptivity (Gao et al., 2018)	4,0	3,5
	AT5G57520.1	ATZFP2, ZFP2	Abscission of floral organs (Cai and Lashbrook, 2008)	2,1	3,4
fruit vs. flower	AT2G01940.3	ATIDD15, SGR5	Auxin biosynthesis and transport, aerial organ morphogenesis and gravitropic responses	-3,3	-4,2
	AT2G20180.2	PIF1, PIL5	Negative regulation of phytochrome-mediated seed germination	-2,5	-5,2
	AT3G23050.1	AXR2, IAA7	Auxin response (Timpte et al., 1994), shoot and root gravitopism (Timpte et al., 1992)	-2,3	-3,2
	AT5G64530.1	ANAC104, XND1	Xylem formation (Tang et al., 2018, Zhao et al., 2017), Regulation of secondary wall synthesis (Zhao et al., 2007)	-3,6	-5,1
	AT5G67300.1	ATMYB44, ATMYBR1, MYB44, MYBR1	Abscisic acid signaling, abiotic stress tolerance	-2,3	-3,9

Table 5: DEGs between the two *Lepidium* species annotated as "DNA-binding transcription factor activity". (GO:0003700)

Ortholog ID	Ortholog name	Ortholog description (based on TAIR)	Reg. bud	Reg. flower	Reg. fruit
AT1G01060.1	LHY	Involved in circadian rhythm	2.5	2.4	2.8
AT1G14687.1	HB32, ZHD14	n.a.	-4.0	-3.1	-2.2
AT1G27370.1	SPL10	Development of lateral organs, lamina shape, lateral root growth (Yu et al., 2015)	-3.9	-3.8	-4.6
AT1G46264.1	HSFB4, SCZ	Asymmetry of stem cell devisions	-5.3	-4.4	-4.1
AT1G53160.2	FTM6, SPL4	Regulation of flowering and vegetative phase change	-5.2	-5.3	-4.5
AT1G79840.2	GL2	Regulation of epidermal cell identity, regulation of seed oil content	3.5	3.6	2.7
AT3G09370.2	MYB3R-3	DNA damage response	-2.4	-2.2	-2.6
AT3G11280.1	n.a.	n.a.	-2.6	-2.9	-2.9
AT3G14020.1	NF-YA6	Involved in male gametogenesis, embryogenesis, and seed development (Mu et al, 2012)	-2.6	-2.2	-2.4
AT3G53340.1	NF-YB10	Flowering time determination (Tao et al., 2017)	3.3	3.4	4.9
AT4G00180.1	YAB3	Specification of abaxial cell fate, involved in fruit patterning along with FIL	-2.6	-2.5	-2.5
AT4G01280.2	RVE5	Clock regulation, growth regulation (Gray et al., 2017)	2.6	2.2	2.2
AT4G31060.1	n.a.	n.a.	-2.9	-3.5	-3.4
AT5G04340.1	C2H2, CZF2, ZAT6	Phosphate homeostasis (Devaiah et al., 2007), Cd accumulation and tolerance (Chen et al. 2016)	4.0	4.6	3.4
AT5G10140.1	AGL25, FLC, FLF, RSB6	Flowering time determination	3.7	4.1	5.6
AT5G39760.1	HB23, ZHD10	Light-induced developmenet (Perrella et al., 2018)	7.2	6.3	4.9
AT5G41920.1	SCL23	Endodermis development (Yoon et al., 2016; Cui et al., 2014)	-5.0	-5.3	-3.9
AT5G47640.1	NF-YB2	Flowering time determination (Hou et al., 2014)	2.2	2.6	2.4
AT5G60910.1	AGL8, FUL	Fruit development (Gu et al, 1998), apical hook development (Führer et al., 2020)	-2.9	-2.4	-2.7

Table 6: DDEGs in different structures annotated as "DNA-binding transcription factor activity" (GO:0003700)

	Ortholog ID	Ortholog name	Ortholog description (based on TAIR)	Reg.
flower vs. bud	AT3G15540.1	IAA19, MSG2	Stamen filaments development	4,8
	AT5G47230.1	AtMACD1, ERF102, ERF5	Stress response, leaf growth	5,3
fruit vs. flower	AT2G20180.2	PIF1, PIL5	Phytochrome-mediated seed germination	2,7
	AT3G01530.1	ATMYB57, MYB57	Stamen and nectary development (Bender et al., 2013)	-2,9
	AT3G15030.1	MEE35, TCP4	Cotyledon, leaf and petal development, seed oil accumulation Overexpression of TCP4 causes increased deposition of lignin and cellulose (Sun et al., 2017), TCP4 interacts with AS2 to regulate BP? (Li et al., 2012)	-4,3
	AT3G18550.1	BRC1, TCP18	Arrests axillary bud development and prevents axillary bud outgrowth. Role in flowering control.	-4,3
	AT5G52660.2	RVE6	Involved in circadian rhythm	-2,6
	AT5G53200.1	TRY	Trichome and root hair patterning, phosphate starvation response	-6,0
	AT5G60850.1	DOF5.4, OBP4	Cell Cycle Progression and Cell Expansion	-2,3

Table 7: Binding of TFs found to be DDEGs to the promotors of known fruit development genes.

		PIF1	MYB57	TCP4	RVE6	OBP4
Valve	AS1	1	1	-	-	3
	AS2	-	1	-	-	1
	JAG	2	-	-	-	1
	FIL	1	-	-	3	2
	YAB3	-	-	1	2	1
	ARF6	1	-	-	2	1
	ARF8	-	-	-	-	-
	FUL	-	1	-	1	1
	AP2	-	1	-	-	2
Replum	NTT	-	1	-	1	3
	BP	-	-	-	1	2
	WOX13	-	-	-	-	4
	RPL	-	-	-	3	1
Valve margin	SHP1	-	-	-	2	5
	SHP2	-	1	-	1	5
	IND	-	-	-	2	3
	ALC	-	1	1	2	2
	SPT	1	-	-	1	2