COMPARATIVE ANALYSIS OF SMALL RNAS AND TRANSCRIPTOMIC DATA DURING THE FLORAL TRANSITION IN *BRASSICA RAPA*

AILEEN MAGILIN
STUDENT ID: 100360317

PhD Biological Sciences Research with Rotation Year

University of East Anglia
John Innes Centre

SUPERVISORS:
Rachel Wells & Richard Morris

March 2026

ABSTRACT

Rapid-cycling *Brassica* crops transition from vegetative to reproductive growth without requiring vernalization. Two cultivars of rapid-cycling *B. rapa* differ markedly in flowering time; Sarisha-14, a commercial cultivar, flowers almost two weeks earlier than the lab cultivar R-o-18. These contrasting phenotypes provide a unique opportunity to understand how rapid-cycling cultivars regulate flowering time, particularly for Sarisha-14's precocious flowering.

This thesis tests the hypothesis that the quantitative ratio of two broadly conserved microRNAs, miR156 and miR172, dictate the timing of the vegetative-to-floral transition. To investigate this, I profiled miRNA abundance in the shoot apical meristem across development in both cultivars, and performed bulked segregant analysis (BSA) to identify gene alleles from Sarisha-14 linked to its early flowering phenotype.

My findings reveal that the abundance dynamics of miR156 and miR172 confirm their conserved role in controlling the floral transition. A potential flowering regulatory model emerges, where the miR156/miR172 abundance ratio of log10 value of 1.1 marks the floral transition in both cultivars, and is aligned with Sarisha-14's earlier shift in flowering time. Additional candidate flowering miRNAs also showed cultivar-specific temporal abundance dynamics. The BSA identified several *B. rapa* flowering gene orthologues linked to early flowering, especially a disruptive allele of *TERMINAL FLOWER 1* (*TFL1*), a floral repressor, which may contribute to Sarisha-14's accelerated flowering.

This is the first study to profile microRNA abundance in *B. rapa* from seedling to floral bud emergence. The findings highlight regulatory mechanisms of flowering time and provide potential targets for breeding crops with shorter reproductive cycles and improved environmental resilience.

CONTENTS

1 Introduction **1**
1.1 Background in B. rapa 1
1.1.1 Agronomy 1
1.1.2 Genetics 2
1.2 The importance of regulating flowering time in crops 3
1.2.1 Flowering behaviour of Sarisha-14 and R-o-18 5
1.3 Canonical flowering time pathways 6
1.3.1 Photoperiod pathway 7
1.3.2 Vernalization pathway 8
1.3.3 Gibberellin and autonomous pathways 8
1.3.4 Ageing pathway 8
1.4 Current knowledge on rapid-cycling *B. rapa* flowering 9
1.5 MicroRNAs in developmental regulation 10
1.6 Understanding how microRNAs regulate flowering in rapid-cycling B. rapa 15
1.7 Thesis Hypothesis and Objectives 16
1.8 Thesis overview 17

2 Methodology for microRNA expression profiling **18**
2.1 Introduction 18
2.1.1 18
2.1.2 18
2.2 Material and methods 20
2.2.1 Plant material generation 20
2.2.2 Tissue grinding 21
2.2.3 Total RNA extraction 22
2.2.4 Small RNA sequencing 23
2.2.5 Small RNA-seq data processing and miRNA identification 23
2.3 Optimization of RNA extraction methods 27
2.3.1 Testing RNA extraction kits with *B. rapa* leaf and apex tissue 27
2.3.2 Tapestation analysis of *B. rapa* extracted RNA 28
2.3.3 Testing grinding methods and extraction kits with *B. oleracea* apex tissue 32
2.3.4 RNA yield comparison 33
2.3.5 Bioinformatic analysis of DH1012 samples 33

2.4 Sequencing miRNAs . . . 37
2.4.1 Concern regarding overrepresented sequences . . . 37
2.4.2 Testing miRNA enrichment using Sequence Suppression Probes 38
2.4.3 Bioinformatic analysis of SSP treated R-o-18 apex libraries 40
2.5 B. rapa miRNA timeseries data description . . . 47
2.5.1 Sequencing data quality . . . 47
2.5.2 Overview of identified small RNA . . . 52
2.6 Discussion . . . 57
2.6.1 Challenges in Brassica miRNA identification . . . 57

3 Identification of candidate microRNAs involved in flowering . . . 60
3.1 Introduction . . . 60
3.1.1 Hypotheses . . . 62
3.1.2 Objectives . . . 62
3.2 Methods . . . 63
3.2.1 Plant growth conditions, sampling, and small RNA sequencing . . . 63
3.2.2 MiRNA identification and abundance quantification . . . 64
3.2.3 Fuzzy c-means clustering for grouping miRNAs with similar abundance patterns . . . 64
3.2.4 Curve registration of miRNA abundance profiles between R-o-18 and Sarisha-14 . . . 65
3.2.5 Timeseries differential expression analysis with DESeq2 . 65
3.2.6 MiRNA target gene prediction . . . 66
3.2.7 Target gene mRNA expression quantification and normalisation . . . 67
3.2.8 Plotting results . . . 67
3.3 Results . . . 68
3.3.1 5 distinct abundance pattern clusters are identified in Sarisha-14 and R-o-18 . . . 69
3.3.2 Identifying miRNAs shared between clusters in Sarisha-14 and R-o-18 . . . 72
3.3.3 Case studies of potential miRNAs that repress flowering . 75
3.3.4 Case studies of potential miRNAs that promote flowering 82
3.3.5 Cultivar-specific miRNAs may have a role in Sarisha-14's early flowering phenotype, but their target genes need more investigation . . . 88
3.4 Discussion . . . 96
3.4.1 MiRNAs that share regulatory roles in R-o-18 and Sarisha-14 96
3.4.2 Cultivar specific miRNAs and their potential regulatory role in Sarisha-14's early flowering . . . 98
3.4.3 Future investigation on the functions of candidate miRNAs in regulating flowering time in Sarisha-14 . . . 100
3.5 Conclusion . . . 101

4 Investigating the ageing pathway in rapid-cycling *B. rapa* . . . 103

4.1 Introduction . . . 103
4.1.1 The role of miR156 and miR172 in regulating flowering time 103
4.1.2 miR156 and miR172 abundance in rapid-cycling *Brassica rapa* 106
4.1.3 Hypothesis . . . 107
4.1.4 Objectives . . . 108
4.2 Methods . . . 109
4.2.1 Plant growth conditions, sampling, and small RNA sequencing . . . 109
4.2.2 MiRNA identification and abundance quantification . . . 110
4.2.3 MiRNA target gene prediction . . . 110
4.2.4 Target gene mRNA expression quantification and normalisation . . . 111
4.3 Results . . . 112
4.3.1 Timeseries abundance profiling of Bra-miR156 and Bra-miR172 112
4.3.2 Timeseries expression of predicted targets of miR156 and miR172 118
4.4 Discussion . . . 130
4.4.1 Conservation and divergence of miR156–*SPL* regulation . . . 130
4.4.2 Differential regulation of AP2/TOE targets by miR172 . . . 131
4.4.3 Ratio-based model of flowering time control . . . 131
4.4.4 Future directions . . . 132
4.4.5 Agricultural implications . . . 136
4.5 Conclusion . . . 137

5 Functional annotation of candidate SNPs linked to earlier flowering . . . 138
5.1 Introduction . . . 138
5.2 Methods . . . 139
5.2.1 *B. rapa* crossing and F1 plant generation . . . 139
5.2.2 F1 genotyping . . . 139
5.2.3 F2 growth conditions and phenotyping . . . 140
5.2.4 DNA extraction and sequencing . . . 140
5.2.5 SNP identification . . . 142
5.2.6 Further bioinformatic analysis . . . 142
5.3 Results . . . 143
5.3.1 Phenotyping results from R-o-18 x Sarisha-14 F2 generation 143
5.3.2 SNP comparisons between early flowering and late flowering pool . . . 145
5.3.3 Candidate flowering loci in chromosome A01 . . . 148
5.3.4 Candidate flowering loci in chromosome A02 . . . 153
5.3.5 Candidate flowering loci in chromosome A03 . . . 157
5.3.6 Expression profiles of all *TFL1* copies between Sarisha-14 and R-o-18 . . . 161
5.4 Discussion . . . 163
5.4.1 SNPs in *TFL1.A03* could influence Sarisha-14's earlier flowering phenotype . . . 163
5.4.2 SNPs in other flowering gene orthologues identified in the BSA could contribute to early flowering. . . . 163

5.4.3 Candidate genes with early flowering alleles inherited from R-o-18 in chromosome A02 . . . 165
5.4.4 Potential promoters of flowering in Sarisha-14 . . . 166
5.4.5 Future lines of investigation . . . 167
5.4.6 Promising breeding targets for agriculture . . . 169
5.5 Conclusion . . . 169

6 Discussion . . . 171
6.1 Chapter summaries . . . 171
6.1.1 Methodology for microRNA expression profiling . . . 171
6.1.2 Identification of candidate microRNAs involved in flowering 172
6.1.3 Investigating the ageing pathway in rapid-cycling *B. rapa* . 172
6.1.4 Functional annotation of candidate SNPs linked to earlier flowering . . . 173
6.2 Outlooks and limitations . . . 173
6.3 Concluding remarks . . . 175

A Appendix . . . 177
A.1 Appendix Chapter 2 . . . 177
A.2 Appendix Chapter 3 . . . 188
A.3 Appendix Chapter 4 . . . 212

REFERENCES . . . 215

LIST OF FIGURES

Figure 1.1 The Brassica triangle, proposed by U. Nagaharu, N. Nagaharu, et al. (1935), illustrates the relationships between the diploid A, B, and C genomes and the amphidiploid species derived from interspecific hybridisation between the progenitor species which were cultivated in overlapping geographic locations. 3

Figure 1.2 **Sarisha-14's earlier flowering time is linked to its earlier vegetative-to-floral transition.** Shoot apical meristem morphology of R-o-18 (upper) and Sarisha-14 (lower) sampled at successive timepoints. The vegetative shoot apex is very small, flat and initiates flat leaf primordia. When the shoot apex undergoes the floral transition (green-pink bar) , it looks more domed and is flanked by circular floral primordia (marked with white arrows). In Sarisha-14, this occurs on day 12, compared to R-o-18 which occured on day 19. Note that day 21 was marked as the R-o-18 floral transition timepoint for similar timepoint comparisons with Sarisha-14. By day 26, Sarisha-14 has much more developed floral buds compared to R-o-18, despite being the same age. 6

Figure 1.3 **An overview of flowering pathways in Arabidopsis.** This image shows the main pathways and genes that are though to regulate flowering and meristem development in Arabidopsis. The individual pathways are described in more detail in the main text. Figure taken from (Maple et al. 2024) under CC-BY Creative Commons License. 7

Figure 1.4 **An overview of the key steps in miRNA biogenesis.** This image shows the process and key proteins involved from MIRNA transcription to a mature miRNA. Polymerase II (Pol II) transcribes the MIR encoding loci, producing primary miRNA transcripts (pri-miRNAs) which are processed into a miRNA/miRNA* duplex via DICER-LIKE1 (DCL1). HUA1 ENHANCER1 (HEN1) methylates the 3' ends of both strands of the duplex to prevent degradation. The mature guide miRNA (in red) is then loaded into ARGONAUTE1 (AGO1) to guide target cleavage or inhibit translation. Figure taken from S. Zhang, Y. Liu, and B. Yu (2015) 11

Figure 2.1 **Timeseries** 21

Figure 2.2 **MiRNA identification workflow for identification of *B. rapa* miRNAs.** 24

Figure 2.3 **The RNA extraction kits tested for small RNA extraction from Brassica tissue.** (a) An image of Omega Bio-tek's E.Z.N.A MicroRNA extraction kit, which features two spin-columns that captures small RNA (<200nt) and longer RNA (⩾200nt). (b) An image of Zymo Research's Direct-zol Microprep Total RNA extraction kit, with one spin-column that captures total RNA >17nt in length. 28

Figure 2.4 **The E.Z.N.A. extraction kit successfully extracts small RNA < 200 nt, but miRNA cannot be detected with this Tapestation.** The image above is the electropherogram of RNA samples extracted with different kits. The values shown at the bottom are the respective RNA quality (RNA Integrity, RINe) determined by the Tapestation software. A good quality RNA sample for small RNA sequencing should have a RINe value above 8.0, with no RNA degradation (seen as a smear under the 200 nt line). Lane A1 is the reference RNA ladder. Lane B1 is small RNA extracted by the E.Z.N.A kit (*EN_1L (s)*), C1 is spiked total RNA from the E.Z.N.A. kit (*EN_1L*), while D1 (*DZ_1A*) and E1 (*DZ_1L*) are apex and leaf total RNA extracted with the Direct-zol kit. Lane B1 shows that the E.Z.N.A. kit successfully extracts small RNA <200nt, with a strong band between 200 and 25 nt, however, miRNAs are between 20-22nt which lies outside the reference ladder. Lane B1 which had just small RNA has the lowest RINe value, which is due to the Tapestation identifying RNA under 200nt as degraded RNA. Both EN_1L and DZ_1L have similar RINe value of 7.4, while DZ_1A which has less distinct RNA bands has a slightly lower RINe value of 6.8. This is likely due to low RNA input from the small volume of apex tissue used for Direct-zol RNA extraction. 30

Figure 2.5 **EN_1L (b) and DZ_1L (d) show similar RNA intensity profiles, with a short and broad peak between 25-200nt.** The x-axis of these plots indicate the RNA size in the sample, and the y-axis indicates the intensity of the RNA band based on the gel image in Fig.2.4. Figure (a) corresponds with sample EN_1L (s) at lane B1 in Fig.2.4, and consistent with the gel image, shows a strong peak between 25-200nt. DZ_1A (c) has less distinct peaks between 25-200nt, but a visibly larger area of RNA intensity compared to EN_1L (b) and DZ_1L (d). 31

Figure 2.6 **E.Z.N.A. extracted RNA samples failed Novogene's sample quality checks due to spiked small RNA.** The tables above show QC summaries of samples that went through the first round of quality checks (table A), and samples that were resent due to insufficient RNA initially (table B). These summaries were screenshotted directly from Novogene's QC report. 33

Figure 2.7 **Small RNA libraries show the expected enrichment of 24 nt reads, followed by 21 nt reads.** (a) FASTQC Sequence length distribution of all RNA samples. (b) Percentage of overrepresented sequences of all RNA samples. Overrepresented sequences in small RNA libraries occur due to high abundance of non-coding RNA such as ribosomal RNA. Samples extracted with the E.Z.N.A. have a slightly higher percentage of overrepresented sequences. 34

Figure 2.8 **Non-coding RNA makes up a majority of small RNA sequenced in all samples.** This figure shows the average proportion of RNA types within the samples based on grinding and extraction method. RNA types include *B. oleracea* coding sequences, miRNA, non-coding RNA, unannotated *B. oleracea* reads (unidentified_bol), and a small percentage of reads that didn't align to the reference *B. oleracea* genome (not_bol_genome). "DZ" stands for Direct-zol extraction, while "EN" stands for E.Z.N.A. extraction. On average, the pestle grinding method had a higher proportion of miRNA, but not much difference between the extraction kits. 36

Figure 2.9 **Pestle grinding method extracted the most miRNA compared to steel-ball crushing method.** Each bar represents the extraction kit and grinding method used (4 replicates per bar), and the height of the bar represents the average percentage of the clean library which make up miRNAs. Error bars indicate standard deviation. Two-way ANOVA identified significant effects of kit ($p < 0.01$), but not for grind ($p = 0.95$) and no significant interaction between these two ($p = 0.75$). Post hoc pairwise comparisons (Tukey-adjusted emmeans) indicated that within the Direct-zol kit, the pestle grinding method yielded a significantly higher % of miRNA than the steel-ball crushing method ($p = 0.017$), and it is also significantly higher in the E.Z.N.A. kit ($p = 0.007$). 37

Figure 2.10 **An illustration of RealSeq Biosciences Sequence Suppression Probe technology, which capture overrepresented RNA sequences to enrich miRNAs in small RNA libraries.** Illustration sourced from RealSeq Bioscience's website. 38

Figure 2.11 **All RNA samples sent to RealSeq Biosciences passed their sample quality checks.** The gel image above shows the RNA band sizes of all the RNA samples compared to the reference ladder in lane EL1. The leaf RNA samples had lower RINe values than the apex RNA samples, but this can be attributed to stronger small RNA bands (under 200nt). 39

Figure 2.12 **RNA samples sequenced after SSP treatment had lower overrepresented sequences than RNA samples without SSP treatment.** (a) Total sequence counts of all samples including dz1l and en1l. (b) Per sequence quality scores of all samples. (c) Summary of overrepresented sequences in all samples. Notice how dz1l and en1l had more overrpresented sequences in the library compared to all samples sequenced by RealSeq. (d) Sequence length distribution of all samples. All RealSeq samples showed a peak at 21nt and 24nt read length, which is consistent with plant miRNA (21nt) and siRNA (24nt) read lengths. For all of these plots, RealSeq samples without SSP treatment are highlighted in blue, samples with SSP treatment are in pink, while Novogene leaf samples are highlighted in green. 40

Figure 2.13 **Small RNA libraries had lower proportions of non-coding RNA after SSP treatment.** Each bar represents the average proportion of RNA types of R-o-18 apex and leaf samples before and after SSP treatment, with an additional comparison with Novogene's leaf sequencing data. RNA types include R-o-18 coding sequences (cds), miRNA, non-coding RNA (ncrna), unannotated R-o-18 aligned reads (unidentified_ro18), and a small percentage of reads that didn't align to the reference R-o-18 genome (not_ro18_genome). 43

Figure 2.14 **SSP treatment slightly improved miRNA content, but not significantly.** Each bar represents the average percentage of clean reads which aligned to miRNA sequences, depending on tissue type (apex, leaf) and treatment (no SSP treated (control, blue), SSP treated (pink)), with an additional comparison with two Novogene sequenced leaf samples (en1l, dz1l, in green). 43

Figure 2.15 **SSP treatment does not skew the proportion of the top 10 expressed miRNAs within the sample.** A list of top 10 expressed miRNAs from each sample was combined, and each bar represents the proportion of those top expressed miRNAs per sample. Overall, SSP treated samples had consistent proportions of top 10 miRNA with its corresponding sample. Sample ro4a is slightly skewed due to extremely high expression of Bra-miR319a compared to other apex samples (see Fig.2.16 Bra-miR319a expression in apex tissue), but this is not reflected in ro4a-ssp. 44

Figure 2.16 **SSP treatment improved detection of most highly expressed miRNAs above 1000 CPM (counts per million).** The plot shows expression levels of detected miRNAs in (a) apex and (b) leaf tissue, before and after SSP treatment. Note that the y-axis in plot (a) has a break between 10000-18000 CPM for ease of visualisation. SSP improvement of miRNA detection was more significantly pronounced in leaf miRNAs. In both apex and leaf tissue, Bra-miR168a/b show significantly higher expression in SSP treated samples, alongside Bra-miR159a, Bra-miR165a and Bra-miR166a which also increases but not significantly in the apex. Meanwhile, Bra-miR156a expression is negatively impacted by SSP treatment in apex and leaf tissue, and is significantly reduced in leaf tissue. The high outlier in Bra-miR319a apex is from sample ro4a. Paired t-test was performed between SSP vs Control treatments for each miRNA, and significance is annotated on the plot (*: $p \leqslant 0.05$; **: $p \leqslant 0.005$). 45

Figure 2.18 **Per base sequence quality drops in reads longer than 30 nt after adapter trimming.** The figures above show the mean sequence quality scores across the length of sequencing reads in the raw data (a) and after adapter trimming (b). Each line represents a single sample. Most adapter-trimmed reads under 30 nt long have high sequence quality, which includes miRNA reads (20-22 nt). Meanwhile, longer reads (> 30 nt) have lower quality bases near the end of the reads, but this is not concerning since they are filtered out during miRNA size selection. These figures were generated using MultiQC (Ewels et al. 2016). 49

Figure 2.19 **Adapter trimming step fully removes all adapter sequences.** The figures above show the percentage of sequences in each sample (represented in each line) which have adapters across the sequence lengths in the raw data (a) and after adapter trimming (b). Figure (b) shows that the adapter trimming step was successful in removing adapters from all the sequences. These figures were generated using MultiQC (Ewels et al. 2016). 50

Figure 2.20 **20 nt and 24 nt small RNA reads are abundant in all samples.** The figure above shows the distribution of read lengths in each sample (represented with lines) after adapter trimming, but before size selection. A distinct peak is seen at the 24 nt sequence length, followed by 20 nt sequence length, which shows that these are the two most abundant small RNA read lengths in all the samples. Plant miRNAs are usually between 20-22 nt, while small interfering RNAs (siRNAs) are 24 nt long, which suggests that siRNAs are more abundant than miRNAs in these samples. This figure was generated using MultiQC (Ewels et al. 2016) 51

Figure 2.21 **Summary of small RNAs identified in the timeseries dataset.** Each bar represents the average proportion of RNA types identified in each sample. The miRNA identification pipeline aligns the adapter-trimmed and size-selected small RNA reads (clean reads) against the reference R-o-18 genome, non-coding RNA sequences (ncrna), repeat sequences (repeats), miRNA sequences (mirna), and reference R-o-18 coding sequences (coding_seq). Small RNA reads that did not align to the reference genome are labelled as "non-genome", while other genome-aligned small RNA reads that did not align to any known miRNA/ncRNA/repeat/coding sequences were labelled as "unidentified". On average, non-coding RNA make up the majority of the small RNA libraries (~50-75%), followed by unidentified *B. rapa* reads (~15-30%) and non-genome reads (~10-25%). Around 5-8% of the small RNA libraries aligned to R-o-18 coding sequences, while miRNAs make up around 1-3% of reads. Less than 1% of all the samples aligned to known repeat sequences. 52

Figure 2.22 **Principle Component Analysis using miRNA read counts from all 114 leaf (green) and apex (pink) samples from R-o-18 (circular points) and Sarisha-14 (triangular points).** PCA was performed to assess tissue/stage/cultivar specific clustering in the samples. Each point represents a single replicate coloured by tissue and stage, and genotype (shape). Each developmental stage has a different colour shade depending on the tissue. Initially, library-size normalized counts (counts per million, cpm) were used for PCA (a), which showed some overlap between leaf and apex groups, but no clear genotype or stage-specific clustering was observed. After applying variance stabilizing transformation (VST) and batch correction to miRNA read counts, the PCA was repeated on the corrected samples (b). Here, the samples cluster based on genotype, with R-o-18 samples at the upper half of the plot (positive Dim2 values) while Sarisha-14 samples occupy the lower half of the plot (negative Dim2 values). Samples in the same developmental stage and tissue type also cluster together, and developmental stage progression is observed in the leaf tissue (developmental stage progresses further from the right to the left of the plot). However, the apex samples do not show a similar progression, which could be due to leaf samples contributing more to Dim2 differences. These plots show that VST and batch correction improve clustering results. . . . 55

Figure 2.23 **Principle Component Analysis of apex samples from R-o-18 (circular points) and Sarisha-14 (triangular points) using miRNA read counts.** To assess whether the apex samples showed a similar developmental progression like the leaf timepoints in Fig. 2.22, a separate PCA was performed on the apex samples only (except for samples from the BBCH51 stage, which skewed the data due to being more developmentally advanced than the earlier vegetative/floral timepoints). Each point represents a single replicate, with its timepoint labelled and coloured. Similar to Fig. 2.22, plot (a) is PCA performed with library-size normalized counts, which shows that the samples grouped by genotype, but there was no clear clustering by timepoint. In plot (b), the miRNA read counts are variance stabilized and batch corrected, then PCA was repeated. The apex samples now show clear progression throughout the timeseries, with the youngest samples at the top of the plot (high positive Dim2 value), and the oldest samples at the bottom of the plot (low negative Dim2 value). The progression is more defined in the R-o-18 samples on the right side of the plot, while the Sarisha-14 floral samples (days 12-15) overlapped with each other and no longer separated by timepoints. This confirms that the apex sampling timeseries correctly represents the developmental stages that the plants progress through. 56

Figure 3.1 **miRNA-mediated regulatory networks controlling flowering time.** Overview of microRNA pathways involved in the regulation of the floral transition. While the miR156–SPL and miR172–AP2 modules form the central age-dependent regulatory axis, multiple additional miRNA modules (e.g., miR159, miR169, miR319, miR390, miR393, miR399, miR824, and miR5200) also contribute to flowering time control through interactions with key integrators such as *FT*, *SOC1*, and *FLC*. These pathways integrate developmental, environmental, hormonal, and stress-related signals. Arrows indicate activation; bars indicate repression. This figure was taken from a review by Teotia and Tang (2015). 61

Figure 3.2 **Dmin() elbow analysis for fuzzy c-means cluster optimization.** Minimum centroid distances calculated across c=2-25 (repeats=5) for Sarisha-14 (left) and R-o-18 (right). Optimal resolution indicated at c=5-7 (Sarisha-14, red points) and c=8-10 (R-o-18, red points), but c=5 was selected for direct cross-genotype trajectory comparison while capturing major developmental patterns. 69

Figure 3.3 **Fuzzy c-means clustering of *B. rapa* miRNA reveals similar abundance patterns between R-o-18 and Sarisha-14 clusters.** miRNA abundance dynamics during the floral transition were clustered separately for each genotype using Mfuzz. Each line represents the normalized trajectory of a single miRNA within a cluster, colored by its membership strength (higher membership = red, low membership = green; see colour bar). The number of miRNAs with high membership (value $\geqslant 0.6$) is indicated against the total miRNAs assigned to each cluster (high/total miRs). The solid black line represents the cluster centroid, i.e. the trajectory estimated by fuzzy c-means, which summarizes the overall abundance trend within that cluster. Although cluster numbers are not directly comparable between genotypes, several clusters show similar dynamics (e.g. early upregulation, early/late downregulation, abundance peak). The final developmental stage (BBCH51) was excluded from the clustering analysis, as its distinct abundance profile could overshadow the earlier developmental abundance patterns. 71

Figure 3.4 **Out of 79 clustered miRNAs, 31 miRNAs showed upregulation in Sarisha-14 before the floral transition, while 29 miRNAs were downregulated.** Rows represent individual miRNAs and columns represent developmental timepoints, with its corresponding stage annotated on the top. Abundance values were variance-stabilized and standardized to Z-scores per miRNA in each genotype. miRNAs are grouped according to their cluster in Sarisha-14, allowing direct comparison between R-o-18 and Sarisha-14 abundance dynamics. The abundance patterns of each cluster is annotated on the far left column. 73

Figure 3.5 **Predicted target sites for floral repressor miRNAs** (Bra-miR157a, Bra-miR390a, Bra-miR396a, Bra-miRN347, Bra-miRN367). Each panel shows the miRNA–target duplex discussed in the case studies, with target (5′→3′) above and miRNA (3′→5′) below. Colons (:) indicate matched bases, dots (.) mismatches, and dashes (–) gaps. Lower targetFinder scores indicate stronger predicted interactions. These miRNAs are downregulated at floral transition, correlating with target gene upregulation. 76

Figure 3.6 **Bra-miR157a downregulation aligns with increased abundance of its predicted target gene *A04p003490.1_BraROA*, an orthologue of *SPL9*. (Top)** Overlay of normalized abundance profiles for Bra-miR157a and *A04p003490.1_BraROA*, aligned to the floral transition (day 0). **(Bottom left)** abundance pattern of Bra-miR157a, which shows an earlier shift in downregulation in Sarisha-14. **(Bottom right)** abundance pattern of *A04p003490.1_BraROA*, which increases during the floral transition, consistent with Bra-miR157a downregulation. 78

Figure 3.7 **Bra-miR390a targets *A08p040520.1_BraROA*, a predicted orthologue of ALFIN-LIKE 7 (AL7), a member of PHD domain family proteins which reads H3K27me3 marks. (Top)** Overlay of normalized abundance profiles for Bra-miR390a and *A08p040520.1_BraROA*, aligned to the floral transition (day 0). **(Bottom left)** abundance pattern of Bra-miR390a, which shows downregulation prior to the floral transition (timepoint marked with vertical lines for each genotype). **(Bottom right)** abundance pattern of target gene *A08p040520.1_BraROA*, which rapidly increases as the cultivars approach the floral transition. . 79

Figure 3.8 **Bra-miR396a is predicted to target *A04p008130.1_BraROA*, a potential orthologue of Arabidopsis *Growth Regulating Factor 4 (GRF4)*. (Top)** Overlay of normalized abundance profiles for Bra-miR396a and *A04p008130.1_BraROA*, aligned to the floral transition (day 0). **(Bottom left)** abundance pattern of Bra-miR396a, which shows downregulation prior to the floral transition (timepoint marked with vertical lines). **(Bottom right)** abundance pattern of target gene *A04p008130.1_BraROA*. 80

Figure 3.9 **Novel predicted miRNA Bra-miRN347 targets an isopentenyltransferase orthologue *IPT2*, possibly regulating cytokinin biogenesis. (Top)** Overlay of normalized abundance profiles for Bra-miRN347 and *A04p026570.1_BraROA*, aligned to the floral transition (day 0). Comparison of their relative abundance patterns suggests that Bra-miRN347 abundance represses *A04p026570.1_BraROA*, especially since its variation is also reflected in the target gene expression. **(Bottom left)** abundance pattern of Bra-miRN347, which is downregulated during the floral transition (this stage is marked with vertical lines). **(Bottom right)** abundance pattern of target gene *A04p026570.1_BraROA*, which increases at the same rate in Sarisha-14 and R-o-18, possibly reflecting derepression by Bra-miRN347 downregulation. 81

Figure 3.10 **Bra-miRN367 is predicted to target an orthologue of *Voltage-Dependent Anion Channels 1 (VDAC1)*. (Top)** Overlay of normalized abundance profiles for Bra-miRN367 and *A05p055140.1_BraROA*, aligned to the floral transition (day 0). **(Bottom left)** abundance pattern of Bra-miRN367, which shows rapid downregulation prior to the floral transition (this stage is marked with vertical lines). **(Bottom right)** abundance pattern of target gene *A05p055140.1_BraROA*, which increases almost 2-fold in Sarisha-14 and 1.5-fold in R-o-18 at the floral transition stage. . . 82

Figure 3.11 **Predicted target sites for floral promoter miRNAs** (Bra-miR159c, Bra-miR319c, Bra-miR394a). Each panel shows the miRNA–target duplex discussed in the case studies, with target (5′→3′) above and miRNA (3′→5′) below. Colons (:) indicate matched bases, dots (.) mismatches, and dashes (–) gaps. Lower targetFinder scores indicate stronger predicted interactions. These miRNAs are upregulated leading up to floral transition, correlating with target gene downregulation. 83

Figure 3.12 **Bra-miR159c is predicted to target *A03p019540.1_BraROA*, a predicted orthologue of *TCP Interactor containing EAR motif protein4 (TIE4)* and *A03p040620.1_BraROA*, a predicted Metallo-hydrolase/oxidoreductase orthologue. (Top)** Overlay of normalized expression profiles for Bra-miR159c and its target genes, aligned to the floral transition (day 0). **(Bottom left)** Abundance pattern of Bra-miR159c, which increases in abundance as both cultivars approach the floral transition. This miRNA is expressed higher in R-o-18 than Sarisha-14. **(Bottom right)** Abundance pattern of *A03p019540.1_BraROA* and *A03p040620.1_BraROA*, which decreases in abundance throughout development. *A03p040620.1_BraROA* is expressed slightly higher in Sarisha-14, consistent with Bra-miR159c's lower abundance in Sarisha-14. 85

Figure 3.13 **Bra-miR319c targets several TCP transcription factors, including *TCP4* predicted orthologue *A05p044300.1_BraROA*, suggesting a miR319-TCP4 regulation of floral development. (Top)** Overlay of normalized expression profiles for Bra-miR319c and its target *A05p044300.1_BraROA*, aligned to the floral transition (day 0). **(Bottom left)** Abundance pattern of Bra-miR319c, which increases in abundance as both cultivars approach the floral transition. **(Bottom right)** Expression pattern of *A05p044300.1_BraROA*, which decreases in expression in line with Bra-miR319c upregulation, and then increases again as the cultivars approach the final timepoint (BBCH51). 86

Figure 3.14 **Bra-miR394a targets a potential orthologue of *LEAF CURLING RESPONSIVENESS (LCR)*, repressing it before the floral transition. (Top)** Overlay of normalized expression profiles for Bra-miR394a and *A09p050010.1_BraROA*, aligned to the floral transition (day 0). **(Bottom left)** Abundance pattern of Bra-miR394a, which is upregulated leading up to the floral transition (timepoint marked with vertical lines for each genotype). **(Bottom right)** Expression pattern of target gene *A09p050010.1_BraROA*, which rapidly increases as the cultivars approach the floral transition. 87

Figure 3.15 **Predicted target sites for cultivar-divergent miRNAs** (Bra-miR398a, Bra-miR397). Each panel shows the miRNA–target duplex discussed in the case studies, with target (5′→3′) above and miRNA (3′→5′) below. Colons (:) indicate matched bases, dots (.) mismatches, and dashes (–) gaps. Lower TargetFinder scores indicate stronger predicted interactions. These miRNAs show distinct abundance trajectories between Sarisha-14 and R-o-18. Target predictions for Bra-miR169 and Bra-miR399 failed to identify potential target genes relevant to flowering, nor exhibited a reciprocal expression pattern to the miRNAs, so were not included in this figure. 90

Figure 3.16 **Bra-miR398a is predicted to target an orthologue of *DREB2B*, which is a dehydration and heat inducible transcription factor. However, diverging abundance patterns and levels between Sarisha-14 and R-o-18 do not align with target gene expression patterns. (Top)** Overlay of normalized abundance profiles for Bra-miR398a and *A01p054070.1_BraROA*, aligned to the floral transition (day 0). **(Bottom left)** abundance pattern of Bra-miR398a, which is rapidly upregulated in Sarisha-14 prior to the floral transition (marked with vertical lines for each genotype), while in contrast, R-o-18 shows lower abundance and downregulation before the floral transition. **(Bottom right)** abundance pattern of target gene *A01p054070.1_BraROA*, which increases throughout development, with similar abundance patterns in both genotypes. This shows that the differential expression levels of Bra-miR398a are not reflected in differential regulation of this target gene. 92

Figure 3.17 **Bra-miR169 members display cultivar specific abundance patterns, with early downregulation in Sarisha-14, but late downregulation or peak in abundance before the floral transition in R-o-18.** a) Comparisons between clusters from Sarisha-14 and R-o-18, with Bra-miR169a, Bra-miR169e and Bra-miR169q abundance patterns highlighted in either genotype. These miRNAs are the only members which were strongly associated with its respective cluster using Mfuzz. MiRNAs in S14 cluster 2 display early downregulation pattern, while miRNAs in RO cluster 3 undergo late downregulation, and RO cluster 2 shows peak abundance. b) abundance of Bra-miR169 members across development in R-o-18 and Sarisha-14, highlighting cultivar-specific differences in abundance pattern. Vertical bars mark the floral transition timepoint for each cultivar. 93

Figure 3.18 **Bra-miR399a and Bra-miR399e display cultivar-specific abundance patterns, and is more highly expressed in Sarisha-14 than R-o-18.** a) Bra-miR399a and Bra-miR399e are highlighted in their respective clusters from each genotype, showing early downregulation in S14 cluster 2, and peak in abundance in RO cluster 2. b) abundance of Bra-miR399a and Bra-miR399e throughout development in R-o-18 and Sarisha-14, highlighting differences in abundance patterns and abundance between cultivars. Vertical bars mark the floral transition timepoint for each cultivar. 94

Figure 3.19 **Other miRNAs which are highly expressed in Sarisha-14 but not R-o-18 suggest a role in early flowering, but target gene predictions are not consistent with miRNA abundance patterns.** a) Bra-miR397 abundance and its predicted target gene *A05p018330.1_BraROA* abundance. High Bra-miR397 abundance in Sarisha-14 is consistent with low *A05p018330.1_BraROA* abundance, but this target gene is also downregulated despite low Bra-miR397 abundance in R-o-18. b) Bra-miR408 is predicted to target *A05p017350.1_BraROA*, an orthologue of *LACCASE3 (LAC3)*. Undetectable target gene expression could be due to high Bra-miR408 abundance and cleavage, although this needs to be validated experimentally. 95

Figure 4.1 **An overview of miR156, miR172 and their target genes, and the regulatory network it forms to control flowering.** *LEAFY (LFY), APETALA1 (AP1)* and *FRUITFULL (FUL)* are a group of floral genes that have individual roles in promoting floral organ identity. During the vegetative stage, these genes are repressed by *APETALA2 (AP2)*-like gene family, comprising of *AP2, SCHLAFMÜTZE (SMZ)*, and *TARGET OF EAT (TOE1, TOE2, TOE3)*. At the same time, miR156 maintains the vegetative state by repressing members of the *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* gene family, which promote flowering. As the plant ages, miR156 abundance decreases, leading to the upregulation of its *SPL* targets, where *SPL9* directly activates miR172 abundance, *SPL3, SPL4 and SPL5* promote floral identity genes *LFY, AP1* and *FUL*, and developmental transition genes *SPL9, SPL13, SPL15, SPL2, SPL10* and *SPL11*, which promote expression of floral integrator gene *SUPPRESSOR OF CONSTANS 1 (SOC1)* and are necessary for the floral transition. The activated miR172 targets the *AP2*-like gene family, repressing them and lifting its repression of *SOC1* and floral identity genes *LFY, AP1* and *FUL*. The balance between miR156 and miR172 abundance dynamics is therefore the core of age-dependent flowering pathway. This diagram is adapted from M. Xu, T. Hu, Zhao, et al. (2016). 104

Figure 4.2 *Illustration of temporal changes in miR156 and miR172 abundance and their target gene families during plant development.* MiR156 levels decline with age, permitting the accumulation of SPL transcription factors, whereas miR172 levels increase, leading to repression of AP2-like floral repressors. Together, these antagonistic abundance dynamics form a developmental switch that controls the timing of the floral transition. 106

Figure 4.3 **Expression of pri-miR156 and pri-miR172 precursors in R-o-18 (red) and Sarisha-14 (blue) shoot apex during development.** This figure is adapted from Calderwood, Hepworth, et al. (2021). Calderwood et. al. reported that pri-miR156 expression decreases during development, accompanied by pri-miR172 upregulation. The log ratio of pri-miR156 to pri-miR172 expression are also similar in magnitude at the floral transition, which forms the basis of our hypothesis that miR156/miR172 expression ratio during shoot apex development determines the timing of the floral transition in *B. rapa*. Vertical lines mark the first day of the floral transition in the respective cultivars. 106

Figure 4.4 **A visualisation of the ratio hypothesis.** We hypothesise that miR156 downregulation (orange), accompanied by miR172 upregulation (green), converges to an abundance ratio which determines the timing of the floral transition (vertical dashed red line). The actual ratio may not actually be 1-to-1 as depicted in this diagram, which is merely for illustration purposes. . 108

Figure 4.5 **Multiple sequence alignment of all Bra-miR156 members.** 20 reference *B. rapa* miR156 mature miRNA sequences were downloaded from PmiREN, and multiple sequence alignment was performed using ClustalOmega. These sequences are displayed in 5′-3′ alignment. Bra-miR156 members with similar sequences were aligned together, showing that 13 out of 20 members had the same sequence as Bra-miR156a. Bra-miR156h and Bra-miR156i are the only members with unique sequences, while Bra-miR156d and Bra-miR156g shared the same 21nt sequence. 113

Figure 4.6 **Multiple sequence alignment of all Bra-miR172 members.** 10 reference *B. rapa* miR172 mature miRNA sequences were downloaded from PmiREN, and multiple sequence alignment was performed using ClustalOmega. These sequences are displayed in 5′-3′ alignment. There are three groups of Bra-miR172 members sharing the same mature miRNA sequence. The consensus diagram shows highly conserved nucleotides from position 2-20, with different Bra-miR172 members having different nucleotides in either the first or last position. 113

Figure 4.7 **Timeseries abundance pattern of individual miR156 and miR172 copies during shoot apex development.** Each plot illustrates the average abundance of individual miRNA members across development in R-o-18 (red) and Sarisha-14 (blue). Each point in the plot corresponds to the abundance level in a replicate sample, while the shaded ribbon is the standard deviation of abundance. The first day of the floral transition for either cultivar is denoted with vertical lines (Sarisha-14 on day 12; R-o-18 on day 19). abundance is shown in counts scaled by Trimmed Mean of M factors (TMMC) **A)** All Bra-miR156 members (except Bra-miR156d) are highly expressed in the vegetative stage, followed by downregulation as the age progresses. Bra-miR156a is the most highly expressed miR156 member, and its abundance rapids declines in Sarisha-14 while in R-o-18 the decline is more gradual. **B)** All Bra-miR172 members are lowly expressed during the vegetative stage, and its rise in abundance coincides with the floral transition timepoint. Of all the members, Bra-miR172c is the most highly expressed by the final sampling timepoint (developmental stage BBCH51). 114

Figure 4.8 **Total abundance of Bra-miR156 decline is accompanied by Bra-miR172 upregulation, displaying antagonistic abundance dynamics.** The sum of abundance of all Bra-miR156 members and Bra-miR172 members was calculated for each replicate per timepoint. The centre line in each plot shows the average abundance of the sum of Bra-miR156 and Bra-miR172 members across development in R-o-18 (red) and Sarisha-14 (blue). Each point in the plot corresponds to the total miRNA abundance level in a replicate sample, while the shaded ribbon is the standard deviation of abundance. The first day of the floral transition for either cultivar is denoted with vertical lines (Sarisha-14 on day 12; R-o-18 on day 19). abundance is shown in counts scaled by Trimmed Mean of M factors (TMMC) **A)** Total abundance for Bra-miR156 and Bra-miR172, with the same y-axis scale to demonstrate the differences in abundance level. It is clear that although Bra-miR156 abundance decreases to around half its initial vegetative abundance level at the floral transition timepoint, it is still more highly expressed than Bra-miR172 overall. **B)** A zoomed-in plot for total Bra-miR172 abundance on a smaller scale, which shows that Bra-miR172 rises almost 3-fold, coinciding with the floral transition in both cultivars. 116

Figure 4.9 **Log10 transformed ratio of Bra-miR156/Bra-miR172 expression of 1.1 coincides with the floral transition in both cultivars.** (A) The ratio of miR156 to miR172 expression reduces as Bra-miR156 expression decreases, and Bra-miR172 increases, and when the floral transition starts in either cultivar, the average ratio is a log10 value of 1.1. Each point represents the ratio of total miR156 divided by total miR172 expression from a replicate apex sample. The main line is the average ratio for all apex replicates in a single timepoint, separated by genotype (R-o-18 in red, Sarisha-14 in blue). The standard deviation is represented by the shaded region. The grey dashed line indicates the estimated miR156-to-miR172 ratio which coincides with the floral transition in both R-o-18 and Sarisha-14. (B) Linear models confirm genotype-specific declines. RO: slope=-0.061*** (R^2=0.56), predicted transition at 25d; S14: slope=-0.150*** (R^2=0.79), predicted transition at 14d. Points are individual replicates; lines are fitted slopes. 117

Figure 4.10 **Representative TargetFinder predictions of Bra-miR156a target sites across textitSPL paralogues.** Panels show representative duplexes for highly conserved *SPL9*, *SPL2*, *SPL15.A07* (score=1), the single nucleotide mismatch in *SPL15.A04* (score=1.5), and mismatch in *SPL13* paralogues (score=2). Target (5′→3′) above, miRNA (3′→5′) below. Colons (:) indicate matches, dots (.) mismatches. 119

Figure 4.11 **Sequence comparison of Bra-miR156 mature sequence aligned against *SPL*s′ target binding sites reveals high conservation in target binding sequences, except for a few *SPL*s.** The target binding site sequences for *SPL15.A04*, both copies of *SPL9* and *SPL2* are highly conserved, with one nucleotide difference at position 4 for *SPL15.A04*. In contrast, *SPL5.A03, SPL10.A08* have sequence divergence in the first 6 nucleotides, which could potentially impact miRNA binding efficiency. Target binding site predictions of Bra-miR156 members from TargetFinder from each *SPL* target gene was aligned with the mature Bra-miR156 sequences using ClustalOmege. The miRNA sequences are displayed in reverse 3′-5′ alignment, while the target binding sites are displayed in forward 5′-3′ alignment, displaying the complementary binding between miRNA and target. 120

Figure 4.12 **TargetFinder rediction of Bra-miR172 target sites across *AP2*-like paralogues.** Panels show representative duplexes for highly conserved *TOE2* (score=1), *AP2* (score=2), *SMZ* (score=3), more divergent *TOE3.A07* (score=2.5), and Bra-miR172c:TOE2 (score=2) showing 1-2 nucleotide differences in positions 1–3. Target (5′→3′) above, miRNA (3′→5′) below. Colons (:) indicate matches, dots (.) mismatches. 121

Figure 4.13 **Sequence comparison of Bra-miR172 mature sequence aligned against *AP2*-like genes' target binding sites reveals divergence in some target binding sequences, like *TOE3.A07*.** Most target genes' binding sites have different nucleotides in the first 2-3 nucleotides of the sequence, while the rest of the target site are highly conserved. *TOE3.A07* shows the most sequence divergence from nucleotide position 1-6, potentially impacting miRNA binding to this target gene. Target binding site predictions of Bra-miR172 members from TargetFinder from each Bra-miR172 target gene was aligned with the mature Bra-miR172 sequences using ClustalOmege. The miRNA sequences are displayed in reverse 3'-5' alignment, while the target binding sites are displayed in forward 5'-3' alignment, displaying the complementary binding between miRNA and target. 122

Figure 4.14 **Expression of Bra-miR156 predicted targets in R-o-18 and Sarisha-14 shoot apex.** The expression of target *SPL* orthologues is shown alongside Bra-miR156 vs. individual *SPL* expression profiles to assess aligned reciprocal expression dynamics, or the lack of it. 125

Figure 4.15 **Expression of Bra-miR172 predicted targets in R-o-18 and Sarisha-14 shoot apex.** The expression of target *AP2*-like orthologues is shown alongside Bra-miR172 vs. individual target gene expression profiles to assess aligned reciprocal expression dynamics. 128

Figure 5.1 **Genotyping results for F1 population.** The yellow dotted line indicates the expected band weight for successful Sarisha-14 crosses. 144

Figure 5.3 **Distribution of SNPs associated with flowering time (days to BBCH51)**. (Figure legend continues in the next page) . . 147

Figure 5.5 **A closer look at the first 10Mb of chrA01.** This figure shows the Δ SNP-index plot at the first 10Mb of chromosome A01, with the positions of candidate genes with high impact SNPs annotated with blue lines, and flowering gene *SUS2* in red. The location of *SUS2* within the high Δ SNP-index indicates that it could be a potential contributor to early flowering time. 150

Figure 5.6 **Expression of selected candidate genes influencing flowering time in chrA01.** Time series expression of candidate genes throughout development in R-o-18 (red) and Sarisha-14 (blue). Vertical lines indicate first day of floral transition in R-o-18 (red) and Sarisha-14 (blue). P-values in the top-right corner show whether temporal expression patterns differ between genotypes (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, ns = no significant difference; results from likelihood ratio test for genotype × time spline interaction). 152

Figure 5.7 **A broad negative Δ SNP-index peak near the beginning of chromosome A02 indicates genomic loci inherited from R-o-18 which contribute to earlier flowering.** SNP-index (top) and Δ SNP-index (bottom) plots across chromosome A02. Detailed descriptions of these plots can be found in 5.3.2B and C. . 153

Figure 5.8 **A closer look at the first 10Mb of chrA02.** This figure shows the Δ SNP-index plot at the first 10Mb of chromosome A02, with the positions of candidate genes with high impact SNPs annotated with blue lines, and flowering genes *EMF1* and *PBS3* in red. The location of *EMF1*, *PBS3*, and the orthologue of *AT5G17670* within the high Δ SNP-index suggests that R-o-18 inherited alleles within these genes could influence early flowering time.154

Figure 5.9 **Expression of candidate genes influencing flowering time in chrA02.** Time series expression of candidate genes throughout development in R-o-18 (red) and Sarisha-14 (blue). Vertical lines indicate first day of floral transition in R-o-18 (red) and Sarisha-14 (blue). P-values in the top-right corner show whether temporal expression patterns differ between genotypes (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, ns = no significant difference; results from likelihood ratio test for genotype × time spline interaction). 155

Figure 5.10 **A broader Δ SNP-index peak at the beginning of chromosome A03 indicates another potential loci controlling flowering time closer to the start of this chromosome.** SNP-index (top) and Δ SNP-index (bottom) plots across chromosome A03. Detailed descriptions of these plots can be found in 5.3.2B and C. . 157

Figure 5.11 **A closer look at the first 10Mb of chrA03.** This figure shows the Δ SNP-index plot at the first 10Mb of chromosome A01, with the positions of candidate genes with high impact SNPs annotated with blue lines, and flowering gene *SUS2* in red. The location of *SUS2* within the high Δ SNP-index indicates that it could be a potential contributor to early flowering time. 158

Figure 5.12 **Expression of potential genes influencing flowering time in chrA03.** time series expression of candidate genes throughout development in R-o-18 (red) and Sarisha-14 (blue). Vertical lines indicate first day of floral transition in R-o-18 (red) and Sarisha-14 (blue). P-values in the top-right corner show whether temporal expression patterns differ between genotypes (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, ns = no significant difference; results from likelihood ratio test for genotype × time spline interaction). 159

Figure 5.13 ***TFL1.A03* is the only copy with a high-impact SNP within a region of high SNP divergence.** Comparison of all *TFL1* copies between R-o-18 and Sarisha-14. time series expression of all *TFL1* copies in R-o-18 and Sarisha-14. Data from Calderwood, Hepworth, et al. (2021). 162

Figure A.1 **Expression pattern of Bra-miR156 predicted target genes.** (caption continued in above page) 213
Figure A.2 **Expression pattern of Bra-miR172 predicted target genes.** (caption continued in above page) 214

LIST OF TABLES

Table 1.1 MiRNAs and their roles in plant development. This table was adapted from Dong, B. Hu, and C. Zhang (2022). 14
Table 2.1 **Comparison between the E.Z.N.A. Micro RNA and Direct-zol Microprep Total RNA extraction kits.** Kits were assessed based on RNA size distribution, RNA yield, protocol complexity, and suitability for small RNA sequencing requirements. 29
Table 2.2 **Description of RNA samples sent for RNA quality testing.** Lane ID column refers to the lane ID in the Tapestation gel image result (Fig. 2.4). Sample EN_1L is an approximate 1:18 mixture of small RNA to long RNA. RNA quality testing with the Tapestation 4200 was performed by Dr Saleha Bakht from the JIC Molecular Genetics platform. 29
Table 2.3 **Description of pilot DH1012 (*B. oleracea*) samples sent for small RNA sequencing to assess miRNA extraction quantity.** 33
Table 2.4 **Almost all samples have very low miRNA content (>2% of clean reads.)** The table shows read alignment summary by grinding method. Raw read counts are shown with its percentage of clean reads in smaller font. Clean reads are raw reads that have been adapter trimmed and size filtered(18-30nt). Samples extracted with the Direct-zol kit have sample ID's starting with "dz", while E.Z.N.A. extracted samples start with "en". NcRNA stands for non-coding RNA sequences. * These two dz7a and dz8a samples are read alignment summaries from Novogene's internal bioinformatic analysis from the same samples, which identified less miRNA than our pipeline. 35

Table 2.5 **Summary of samples sent to RealSeq Biosciences for SSP method testing to improve miRNA detection in *Brassica* tissue samples.** Apex and leaf tissue were collected from four independently pooled plants (n = 4 biological replicates per tissue type, pool size = 4 individual plants). * dz1l and en1l were not sent to RealSeq Biosciences, but were included in the later analysis to compare Novogene sequencing with RealSeq sequencing. 39

Table 2.6 **Summary of read alignments from small RNA sequencing of R-o-18 leaf and shoot apex RNA before and after SSP treatment.** Two samples (dz1l and en1l) which were sequenced by Novogene were also included in this analysis. Counts are shown with percentages in smaller font. 42

Table 3.1 **Comparison of abundance patterns between R-o-18 and Sarisha-14 clusters.** Cluster patterns identified by Mfuzz were matched between the two genotypes. Note that R-o-18's cluster 4 upregulation pattern corresponds to Sarisha-14's cluster 5 (up_early) and cluster 3 (up_late). This table is a guide for comparing shared miRNA abundance dynamics. 72

Table 3.2 **List of candidate miRNAs with potential floral promoter/repressor roles based on abundance dynamics during the floral transition.** (caption continued in next page) 74

Table 3.3 **List of miRNAs with cultivar specific abundance patterns.** The padj column is the statistical significance from differential expression analysis of the miRNAs with genotype x time interaction, highlighting miRNAs with genotype-specific abundance patterns ($p<0.05$). The curve registration column is a summary of whether the abundance profiles of these miRNAs are aligned in both cultivars when taking into account time/abundance level (curve registered = TRUE); or not aligned (curve registered = FALSE). The two columns "s14 pattern" and "ro pattern" are the abundance patterns corresponding with the cluster which the miRNAs are assigned to with Mfuzz clustering (see Fig. 3.3 and Table 3.1 for visualisation of the clusters and their abundance pattern description). Some miRNAs were not assigned by Mfuzz to any specific clusters in Sarisha-14 or R-o-18, thus have "NA" in those columns. 89

Table 5.1 **Primer sequences of RA2F04 SSR markers used to genotype F1 R-o-18 x Sarisha-14 lines.** 140

Table 5.2 **Description of each pool sent for sequencing for bulk segregant analysis.** 145

Table 5.3 **Overview of genes with SNPs in regions of interest.** This table summarizes the total number of genes with high-impact SNPs and flowering genes found in chrA01, chrA02 and chrA03. This is within the first 10 Mb of the chromosomes, which contains a region of interest where the Δ SNP-index is higher than the set threshold of $\pm$ 0.45. 148

Table 5.4 **List of candidate genes within chrA01 region of interest with high impact SNPs.** Only genes with detectable expression from either R-o-18 or Sarisha-14 are listed (Calderwood, Hepworth, et al. 2021). 150

Table 5.5 **List of candidate genes within chrA02 region of interest with high impact SNPs.** A few of these genes have no predicted orthologues in Arabidopsis, but their gene IDs and SNP information are still included in this table. Only genes with detectable expression from either R-o-18 or Sarisha-14 are listed (Calderwood, Hepworth, et al. 2021). 156

Table 5.6 **List of genes which contained high-impact SNPs and flowering genes with high or moderate-impact SNPs within chrA03 region of interest.** Only genes with detectable expression from either R-o-18 or Sarisha-14 are listed (Calderwood, Hepworth, et al. 2021). 160

Table A.1 Nanodrop measurements of DH1012 RNA extraction trial. 177

Table A.2 SSP probe sequences used for trial SSP suppression of overrepresented sequences. 178

Table A.3 List of *B. rapa* miRNAs used for miRNA identification. 179

Table A.4 Read alignment statistics of full miRNA timeseries. 184

Table A.5 **Mfuzz soft clustering membership scores for Sarisha-14 miRNAs across 5 expression clusters.** Each miRNA's membership score (0-1) indicates similarity to cluster expression patterns (described in header).Cluster 1 (low confidence/noise), Cluster 2 (early downregulation), Cluster 3 (late upregulation), Cluster 4 (late downregulation), Cluster 5 (early upregulation). The rightmost column indicates the cluster the miRNA is most highly associated with. 188

Table A.6 **Mfuzz soft clustering membership scores for R-o-18**. miRNAs across 5 expression clusters. Each miRNA's membership score (0-1) indicates similarity to cluster expression patterns: Cluster 1 (early downregulation), Cluster 2 (peak abundance), Cluster 3 (late downregulation), Cluster 4 (early upregulation), Cluster 5 (low confidence/noise). The rightmost column indicates the cluster the miRNA is most highly associated with. 192

Table A.7 miRNAs with highest cluster membership (>0.6) in Sarisha-14 (S14) and R-o-18 (RO) Mfuzz clusters. Each column lists the miRNAs most strongly associated with that cluster (membership score >0.6). RO Cluster 5 was not included due to no miRNAs found above this threshold. 197

Table A.8 **TargetFinder predictions for target genes of candidate flowering-associated miRNAs**. The table lists all predicted target genes (brapa_id) for *B. rapa* miRNAs discussed in Chapter 3, with their *A. thaliana* orthologues (ath_orth), targetFinder scores (lower = more complementary bases), target binding site sequence, and base pairing patterns. Colons (:) indicate Watson-Crick matches, dots (.) mismatches, and spaces indicate gaps in the miRNA–target duplex. 198

ACKNOWLEDGEMENTS

The first people I would like to express my deepest gratitude to are my supervisory team: Rachel Wells, Richard Morris, Hugh Woolfenden and Jo Hepworth. I could not ask for a team more supportive than them during my research and thesis writing. Thank you Rachel, for the guidance, probing questions, and constant support from day 1. You have been a steady hand in guiding your students, and have been a deep well of knowledge especially for us Brassica kids in the CSB department. Your passion for Brassicas and support for your mentees has helped so many of us here at JIC. Thank you Richard, for accepting me under your wing and your encouragement to keep pushing the boundaries of modern science. You nurtured an unmatched environment full of creativity, open-mindedness and exploration that encouraged me to go into bioinformatics in the first place. Hugh, thank you so much for your patience and guidance when I would walk up to your desk to ask you yet another question about bash scripting. I have learnt so much about programming languages and proper data management practices, that I hope to take with me on my next steps. And thank you Jo, for your tips and guidance about small RNAs and flowering, and for your unwavering support throughout this project.

I would also like to thank all past and current members of the Morris group, who I spend my day-to-day with, and have made my PhD experience one that I'll always remember fondly. Thank you to the postdocs: Amy, Melissa, and Ruth - I have learnt a lot about machine learning and programming from you all. Thank you especially to the students: Gurpinder, Jared, Georgia, Ella and Sam - your curiosity, discipline and intelligence inspire me every day.

On the lab side, I am eternally grateful to all the current and past members of the Wells group, particularly Emmanuel Solomon and Monica Chhetry. It has been a privilege to work alongside you, and your support was pivotal to this project.

I am also very grateful for the people whose work was essential for this project. I would like to thank the team at JIC Horticultural Services, especially Damian Alger and Catherine Taylor, for taking care of my hundreds of Brassica plants in the CERs and glasshouses, and all the troubleshooting we had to do with the growth rooms. Thank you to the people at JIC Bioinformatics, especially Burkhard Steuernagel, for your guidance when I was unsure about my bioinformatics analysis. Some of the molecular biology work wouldn't have been possible without the support of Saleha Bakht and Roland Wouters at the JIC Molecular Biology platform, for their patience and help with my DNA extractions.

I have also learnt so much from all the past and present PhD students at JIC, all of whom have become dear friends to me. I would like to thank Aaron, Jimmy, Judit, Jiawen, and Sam, who have been a constant source of guidance, encouragement and support during this PhD. I would also like to thank my friends, Thomas, Tom, Julia and Shouchao, for all the coffees, lunches and brunches - your friendship has made this journey so special. A special thank you as well to Ruth and Alf, who have become like family to me, for their support, kindness and patience in our friendship.

My friends and family from Malaysia have been my foundation since the very beginning. Thank you to my lifelong friends, Danisha, Netta and Mei Yao, for your friendship, love and patience. Thank you to my sister, Alicia, for answering my random calls and being there for me even when we are so far apart. To my parents, who have supported my academic journey, and have been my rock during the hardest times, I can only say thank you. You have nurtured my love for plants, nature, and scientific curiosity from a very young age, and have instilled a passion for learning and education which has stayed until now.

To my dearest Alexander, I am so lucky that we could share this PhD experience together. To have a partner who truly understands and can share in the successes and difficulties of a PhD, is a privilege that I never expected. Thank you for being there with me during our late nights of writing - your love, humour and amazing chef skills have kept me going all these years.

DISCLAIMER

ChatGPT-5 (OpenAI) was used during thesis preparation to check grammar and sentence structure. It was used during the later stages of thesis writing from August - September 2025. All research design, data analysis, interpretation, and conclusions are my own.

1

INTRODUCTION

Two cultivars of *Brassica rapa*, Sarisha-14 and R-o-18, flower at drastically different times despite belonging in the same species. Sarisha-14 initiates flowering almost two weeks earlier than R-o-18 under similar growth conditions. Understanding what allows Sarisha-14 to initiate flowering so early would further our fundamental understanding of developmental timing, which could be exploited for crop improvement. This forms the central theme of my thesis : What molecular and genetic mechanisms allow Sarisha-14 to flower substantially earlier than R-o-18?

1.1 BACKGROUND IN B. RAPA

1.1.1 *Agronomy*

The Brassicaceae family includes a wide range of crops used for food, fodder, condiments and ornamentals. They are cultivated globally and represent a major source of both vegetable and oilseed production. One reason for their agricultural importance is their morphological diversity, where different cultivars are grown for their leaves, stems, roots, floral buds, or seeds.

Brassica rapa is a highly diverse crop species within the Brassicaceae family, which includes vegetable and oilseed types which are widely cultivated across Asia. These include varieties such as var. pekinensis (Chinese cabbage), var. chinensis (pak choi), var. parachinensis, var. narinosa, var. japonica and var. rapa (turnip). Brassica rapa also includes oilseed forms, such as var. oleifera (field mustard) and var. sarson, which are commonly grown in India and Bangladesh. In these cultivars, the seeds are used directly in cooking and for oil production.

Short-duration local mustard varieties such as Tori-7 are therefore widely grown in Bangladesh, as they fit into tight intercropping systems (e.g. rice–mustard–rice or rice–mustard–maize) (Sultana, Rahman, and Hasan 2020) . However, these short-duration local varieties are relatively low yielding , and with limited arable land and increasing demand for edible oils (Sultana, Rahman, and Hasan 2020), there is a strong pressure to develop higher-yielding, short-duration mustard cultivars.

To address this need, agricultural research institutes in Bangladesh have developed improved *B. rapa* cultivars, including Sarisha-14 (*B. rapa var. oleifera*) from the Bangladesh Agricultural Research Institute (BARI). Sarisha-14 was developed through crossing local Tori and Sonali cultivars and reaches maturity in approximately 75–80 days under field conditions (Bangladesh Agricultural Research Institute 2006). Its rapid growth and improved yield makes it well suited for existing intercropping systems and economically attractive to farmers. R-o-18 is a double haploid mustard line of Indian ancestry (*B. rapa var. trilocularis*, yellow sarson). It serves as a reference genotype for *B. rapa* research, with a publicly available annotated genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_017639395.1)and extensive genetic resources, including a TILLING population and recombinant inbred line populations (Stephenson et al. 2010; Bagheri et al. 2013). Its self-compatibility, rapid life cycle and well-characterised genome makes it a powerful tool for genetic mapping and functional studies. These features make R-o-18 an ideal reference cultivar for comparison with Sarisha-14 to study the regulation of floweirng time in rapid-cycling cultivars.

1.1.2 *Genetics*

Brassica rapa (2n = 20) has a genome of approximately 350 Mb, more than twice the size of the model species Arabidopsis thaliana (2n = 10; 135 Mb). Although these species share a common ancestor within the Brassicaceae, the Brassiceae lineage underwent a whole genome triplication (WGT) event approximately 24–25 million years ago, after diverging from the lineage leading to Arabidopsis. This polyploidization event was followed by diploidization, during which large-scale gene loss and genome reorganisation reduced redundancy and reshaped transcriptional regulation. As a result, *B. rapa* retains many duplicated gene copies, while still maintaining broad conservation of regulatory pathways with Arabidopsis.

In the context of flowering time, this means that while the core regulatory network is largely conserved, *B. rapa* frequently contains multiple homologues of key flowering regulators. These duplicated genes provide opportunities for regulatory divergence, altered expression dynamics, and fine-tuning of developmental timing. Therefore, insights from Arabidopsis provide a conceptual framework, but flowering regulation in *B. rapa* likely involves additional layers of paralog-specific control.

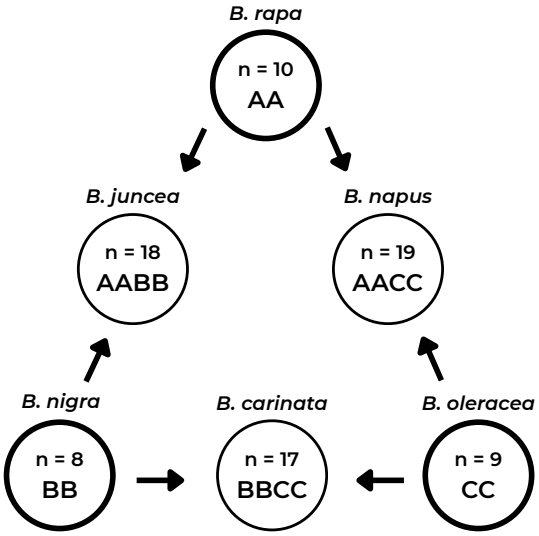


Figure 1.1: The Brassica triangle, proposed by U. Nagaharu, N. Nagaharu, et al. (1935), illustrates the relationships between the diploid A, B, and C genomes and the amphidiploid species derived from interspecific hybridisation between the progenitor species which were cultivated in overlapping geographic locations.

Brassica rapa (AA genome) is also one of the three diploid progenitor species described in the classic "Triangle of U" model (Fig. 1.1, U. Nagaharu, N. Nagaharu, et al. (1935)), alongside *Brassica oleracea* (CC genome) and *Brassica nigra* (BB genome). Hybridisation between these diploids gave rise to the economically important allotetraploid species *Brassica napus* (AACC) and *Brassica juncea* (AABB), as well as Brassica carinata (BBCC). As the A-genome donor, *B. rapa* has contributed substantially to the genetic architecture of several major oilseed crops. Understanding developmental regulation in diploid *B. rapa* therefore provides insights that are directly relevant to its polyploid derivatives.

Within *B. rapa*, further diversification into subspecies occurred through geographic separation, selection and breeding history, all within this shared triplicated genomic background. R-o-18 (*var. trilocularis*) and Sarisha-14 (*var. oleifera*) share the same genome architecture (2n = 20; 350 Mb) but are not near-isogenic lines. Differences in flowering time between them are therefore unlikely to reflect gene presence or absence, but rather variation in allele composition, expression patterns, or regulatory control of paralogous flowering genes. Such divergence may operate at transcriptional, post-transcriptional or epigenetic levels.

1.2 THE IMPORTANCE OF REGULATING FLOWERING TIME IN CROPS

The control of flowering time is critical for plant survival and reproductive success, particularly in seasonal temperate environments. If a plant flowers too early, it risks damage from late frosts or reduced pollinator activity. If it flowers too late, it may fail to set seed before drought or winter, reducing the chances of successful reproduc-

tion. Thus, the developmental switch from vegetative growth to reproductive growth (known as the floral transition) is tightly regulated and integrates both environmental cues and internal developmental signals.

In agriculture, this transition has direct consequences for yield and crop performance. For leafy vegetable crops such as Chinese cabbage and other *Brassica rapa* vegetable types, early bolting limits biomass accumulation because energy is diverted away from leaf and stem production toward flowering and seed development. Similarly, in swollen root crops such as turnips, premature flowering reduces the quality and size of the harvested organ. In contrast, for seed crops such as oilseed rape (Brassica napus) and mustard (*Brassica rapa*), flowering too early can expose developing reproductive tissues to frost damage, while flowering too late can shorten the seed-filling period and reduce final yield. In both cases, the timing of flowering directly shapes economic output.

Flowering behaviour is also closely tied to regional adaptation. In temperate climates, many crops require vernalization (prolonged exposure to cold temperature) to prevent flowering in autumn and ensure reproduction occurs in spring. Oilseed species such as *Brassica napus*, commonly grown in the UK, Canada, and Australia, often have strong vernalization requirements and therefore longer growth cycles. These crops require extended growing seasons and substantial input resources. In contrast, mustard cultivars grown in warmer regions such as Bangladesh typically lack a vernalization requirement and are able to flower rapidly without exposure to cold. This allows them to fit into short intercropping systems and tight agricultural schedules.

However, as global temperatures are rising and winter seasons are becoming less predictable, oilseed yields in temperate regions also become unstable. The warmer or shorter winter seasons mean that local crops fail to fulfil its vernalization requirement, thus flowering later than expected, leading to reduced yield (Brown, Beeby, and Penfield 2019; X. Lu et al. 2022). There is therefore a growing need to develop crop varieties that are both climate resilient and capable of rapid, predictable flowering under variable conditions.

Rapid-cycling oilseed cultivars such as Sarisha-14 and R-o-18 provide an opportunity to study flowering regulation in genotypes that do not require vernalization and complete their life cycle quickly. Sarisha-14 develops visible floral buds at approximately 24 days after sowing, whereas R-o-18 reaches the same stage at around 37 days. Understanding the genetic and molecular basis of this difference may reveal regulatory mechanisms that enable accelerated floral transition without compromising reproductive success. Such knowledge could be exploited to fine-tune flowering behaviour in oilseed crops grown under changing climatic conditions. Furthermore, *B. rapa* forms the A-genome of the polyploid *B. napus* and *B. juncea* oilseed species, so insights into flowering time regulation in *B. rapa* cultivars may inform breeding strategies not only within this species, but also in related Brassica crops grown worldwide.

In summary, regulating flowering time is central to crop adaptation, yield stability, and climate resilience. By investigating how Sarisha-14 and R-o-18 control the timing of the floral transition, this thesis aims to contribute to a broader effort to develop faster-cycling, more adaptable oilseed varieties for modern agriculture.

1.2.1 *Flowering behaviour of Sarisha-14 and R-o-18*

The timing of flowering is linked to the floral transition, where the shoot apical meristem (also referred to as the shoot apex) switches from the vegetative phase to reproductive phase. This irreversible phase transition is determined long before floral organ development. Besides measuring the time it takes for the floral buds to be visible (developmental stage BBCH51), we can identify when the floral transition occurs by visually inspecting the shoot apex under the microscope. During the vegetative stage, the shoot apex is flat and produces leaf primordia, which grow into leaves. During the floral transition, the shoot apex begins to dome, and circular floral primordia flank the shoot apex (Kinoshita et al. 2020). In this thesis, I will refer to the floral transition timepoint as the first day when the shoot apex reflects these signs.

Dissection of the shoot apex reveals that R-o-18 initiated the floral transition at around 19 days after sowing, while Sarisha-14 reached the same stage a week earlier, at day 12. This suggests that Sarisha-14 is able to flower earlier due to transitioning to the reproductive state earlier than R-o-18.

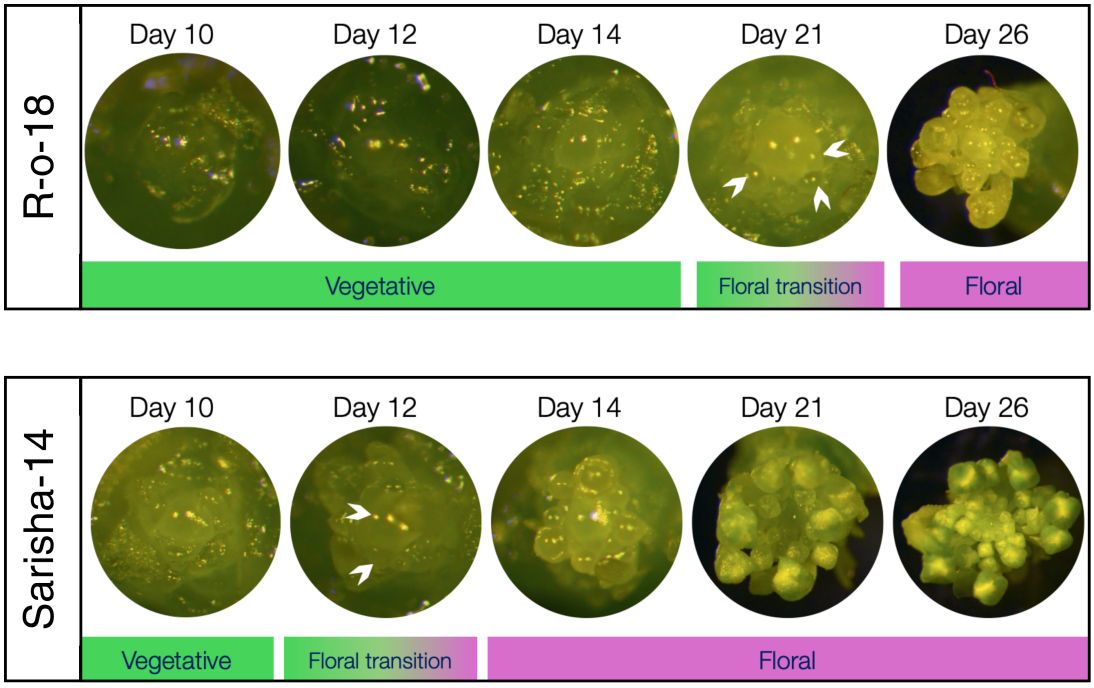


Figure 1.2: **Sarisha-14's earlier flowering time is linked to its earlier vegetative-to-floral transition.** Shoot apical meristem morphology of R-o-18 (upper) and Sarisha-14 (lower) sampled at successive timepoints. The vegetative shoot apex is very small, flat and initiates flat leaf primordia. When the shoot apex undergoes the floral transition (green-pink bar) , it looks more domed and is flanked by circular floral primordia (marked with white arrows). In Sarisha-14, this occurs on day 12, compared to R-o-18 which occured on day 19. Note that day 21 was marked as the R-o-18 floral transition timepoint for similar timepoint comparisons with Sarisha-14. By day 26, Sarisha-14 has much more developed floral buds compared to R-o-18, despite being the same age.

1.3 CANONICAL FLOWERING TIME PATHWAYS

Plants regulate flowering by integrating environmental and endogenous cues to ensure flowering occurs under favourable conditions. Several major regulatory pathways have been identified through extensive studies in the model plant *Arabidopsis thaliana*, known as the photoperiod, vernalization, gibberellin, autonomous and ageing pathways. Many of these pathways converge on the activation of *FLOWERING LOCUS T (FT)*, which is one of the first identified flowering signals (Kardailsky et al. 1999; Kobayashi et al. 1999), and *SUPPRESSOR OF CONSTANS 1 (SOC1)*. These floral integrators then activate downstream floral meristem identity genes, *APETALA1 (AP1)*, *AP2*, *FRUITFULL (FUL)*, *CAULIFLOWER (CAL)*, and *LEAFY (LFY)*. The following section is a brief overview of these pathways as characterized in Arabidopsis.

This core network provides our foundation for understanding flowering regulation in *Brassica rapa*, although gene duplication following the Brassica whole genome triplication means that multiple homologues often exist for key regulators.

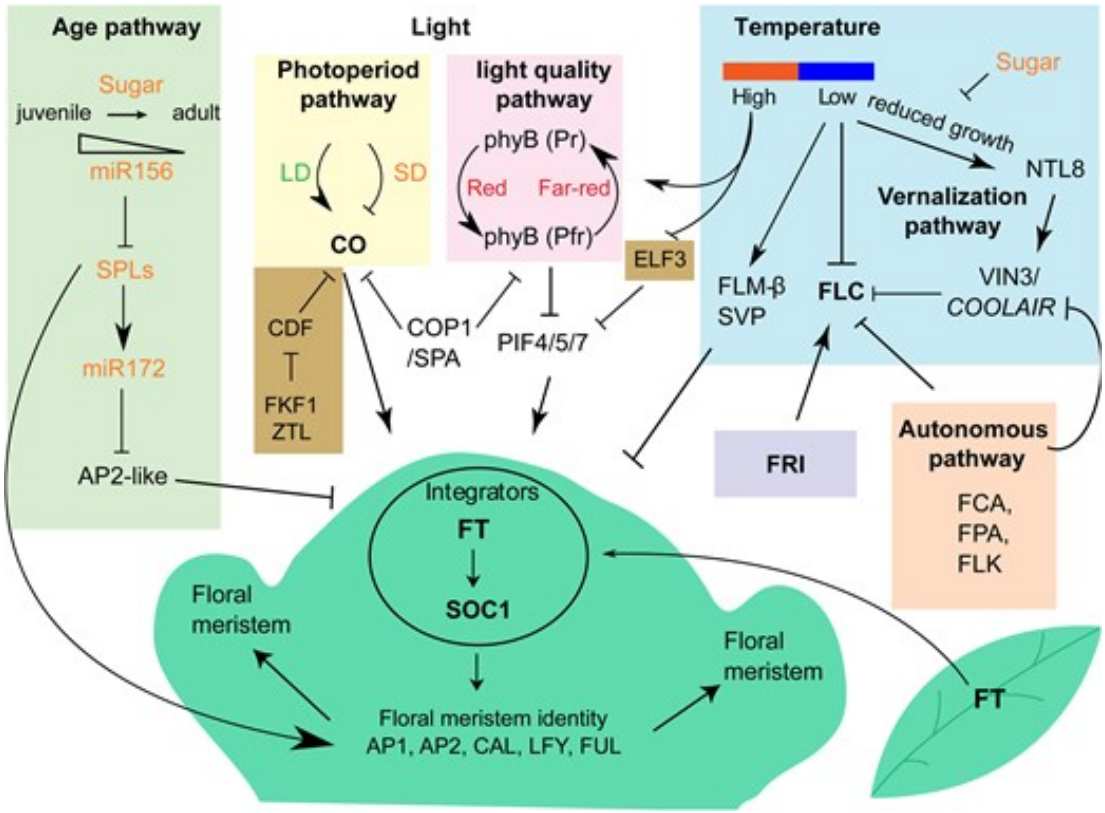


Figure 1.3: **An overview of flowering pathways in Arabidopsis.** This image shows the main pathways and genes that are though to regulate flowering and meristem development in Arabidopsis. The individual pathways are described in more detail in the main text. Figure taken from (Maple et al. 2024) under CC-BY Creative Commons License.

1.3.1 *Photoperiod pathway*

The photoperiod pathway depends on day length to ensure flowering occurs at the appropriate time of year. This pathway employs light signalling in the leaves to send signals to photoperiodic regulators, one of the first identified being *CONSTANS (CO)* (Putterill et al. 1995). During long-day conditions, as experienced during summer months, *CO* activates the expression of *FT* in the leaves, and the FT protein moves to the shoot apex, then promotes flowering via regulation of floral meristem identity genes *AP1*, *AP2*, *FUL*, *CAL* and *LFY* (Suárez-López et al. 2001). In short day-lengths, as experienced during winter months, *CO* accumulation isn't enough to activate *FT*, thus flowering is delayed.

In rapid-cycling *B. rapa* cultivars such as Sarisha-14 and R-o-18, FT is expressed and capable of promoting flowering without a vernalization requirement. However, differences in FT expression alone do not fully explain the divergence in flowering time between these cultivars, suggesting that additional layers of regulation operate downstream or in parallel (Calderwood, Hepworth, et al. 2021).

1.3.2 *Vernalization pathway*

In Arabidopsis, *FLOWERING LOCUS C (FLC)* is a central floral repressor in the vernalization pathway, and is upregulated by *FRIGIDA (FRI)*. In the vegetative stage, highly expressed *FLC* negatively regulates *FT* and *SOC1*, thus reducing the shoot apex's sensitivity to floral inductive signals (Michaels and Amasino 1999; Searle et al. 2006).

Vernalization inhibits *FLC* expression, so plants with a vernalization requirement would require exposure to prolonged cold temperatures in order reduce *FLC* expression and progress to flowering. In the oilseed crop *Brassica napus*, *FLC* paralogues act in a dosage dependent manner which distinguishes spring types from semi-winter and winter type oilseed rape cultivars (Calderwood, Lloyd, et al. 2021). This emphasises the diversity in crop types based on vernalization requirement alone.

In contrast, our rapid-cycling cultivars bypass vernalization requirement, with *SOC1* upregulation occuring prior to decline of *BraFLC.A03a*, the only detected *FLC* paralogue in these cultivars (Calderwood, Hepworth, et al. 2021). This suggests that in these cultivars, flowering time differences are unlikely to be driven by differential cold response, and instead reflect regulation elsewhere in the network.

1.3.3 *Gibberellin and autonomous pathways*

During non-inductive conditions, two pathways come into play to ensure flowering occurs in the absence of favourable environmental signals. The gibberellin pathway is characterised by the gibberellin hormone (GA) which promotes flowering by activating floral integrators while also repressing DELLA growth repressor proteins, allowing plants to flower in non-inductive photoperiods (Langridge 1957; Wilson, Heckman, and Somerville 1992; Cheng et al. 2004).

The autonomous pathway on the other hand, involves RNA-binding proteins and chromatin modifiers to repress *FLC* in the absence of vernalization (Macknight et al. 1997; Schomburg et al. 2001; Lim et al. 2004; Marquardt et al. 2006). Together, GA and autonomous pathways acts as buffers for flowering time control when environmental conditions are not favourable.

1.3.4 *Ageing pathway*

The ageing pathway functions as an endogenous developmental timer, and its key regulators are microRNA families miR156 and miR172. In juvenile plants, high levels of miR156 repress SQUAMOSA Promoter Binding Protein–Like (SPL) transcription factors. As plants age, miR156 declines, leading to SPL upregulation. SPLs directly activate flowering genes such as *LFY*, *AP1*, and *FUL*, and SPL15 has been shown to directly activate miR172 (G. Wu, Park, et al. 2009; J.-W. Wang, Czech, and Weigel 2009).

Meanwhile, miR172 upregulation represses AP2-like floral repressors, thus promoting flowering and floral meristem identity (G. Wu, Park, et al. 2009; Jung et al. 2007). This sequential downregulation of miR156 and upregulation of miR172 provides a conserved molecular cascade linking vegetative phase change to flowering competence.

This regulatory module is highly conserved in the plant kingdom. In cereals, miR172 regulates AP2-like repressors downstream of photoperiod and vernalization signals, thereby influencing heading date and spike architecture (Debernardi, Woods, et al. 2022). In perennial fruit trees, such as apple and citrus, the miR156 module controls the transition from the juvenile to the adult vegetative phase, contributing to the extended juvenility that is a major obstacle for fruit breeding (Zhou et al. 2013; J.-W. Wang, Park, et al. 2011). These examples emphasise the role of the miR156/miR172 regulatory module and how plant age influences flowering time.

1.4 CURRENT KNOWLEDGE ON RAPID-CYCLING *b. rapa* FLOWERING

Although flowering pathways are well characterised in model species such as *Arabidopsis thaliana*, their operation in rapid-cycling *B. rapa* cultivars remains less clearly resolved. Calderwood, Hepworth, et al. (2021) investigated the gene regulatory networks involved in Sarisha-14's early flowering through time series transcriptome profiling of flowering time genes across development between Sarisha-14 and R-o-18. In their time series, they found that Sarisha-14 underwent the floral transition at day 10, a week earlier compared to R-o-18 (day 17).

FT expression in the leaf did not account for the 7-day difference in floral transition timing between R-o-18 and Sarisha-14 (Calderwood, Hepworth, et al. 2021). In both cultivars, floral transition coincided with increased *FT* gene expression; however, Sarisha-14 appeared to initiate flowering at a lower apparent *FT* threshold. Thus, the authors suggested that Sarisha-14 may have increased sensitivity to FT in the shoot apex, since the floral transition occured when FT expression increased past a lower threshold in Sarisha-14, compared to R-o-18.

Unlike temperate Brassica cultivars, these rapid-cycling cultivars do not have a vernalization requirement. *SOC1* upregulation occurred prior to decline of *BraFLC.A03a*, the only detected *FLC* paralogue in these cultivars, indicating that variation in cold-mediated repression of flowering is unlikely to account for their flowering time differences (Calderwood, Hepworth, et al. 2021).

However, the authors also identified that the expression profiles of pri-miR156 and pri-miR172 precursors from the ageing pathway diverged between the two cultivars, hinting that the ageing pathway may be involved in Sarisha-14's earlier flowering. Consistent with the ageing pathway characterised in Arabidopsis, *B. rapa* pri-miR156 expression was high at the seedling stage, then declined with age, while pri-miR172

expression rose antagonistic to pri-miR156 downregulation. Notably, the authors found that at the floral transition developmental stage, pri-miR156 and pri-mir172 abundance were at similar ratios (Calderwood, Hepworth, et al. 2021). This raises the possibility that the developmental timing difference between Sarisha-14 and R-o-18, is due to Sarisha-14 achieving developmental competence earlier than R-o-18, potentially through accelerated progression of age-related regulatory programs.

Collectively, these findings suggest that flowering time divergence in rapid-cycling *B. rapa* cannot be explained solely by transcriptional variation in canonical flowering time genes from the photoperiod or vernalization pathway, and may instead involve differences in regulatory integration or post-transcriptional control.

1.5 MICRORNAS IN DEVELOPMENTAL REGULATION

Although flowering pathways are often described in terms of transcriptional regulators and environmental inputs, this framework captures only part of the regulatory hierarchy controlling developmental transitions. The timing, sensitivity and robustness of these pathways depend on additional layers of post-transcriptional control. The ageing pathway provides a clear illustration of this principle through the action of miR156 and miR172. However, focusing solely on this well-characterised module risks overlooking broader small RNA networks that may modulate flowering time or developmental progression more generally. To fully understand regulatory divergence between cultivars, it is therefore necessary to consider the wider landscape of microRNA-mediated gene regulation.

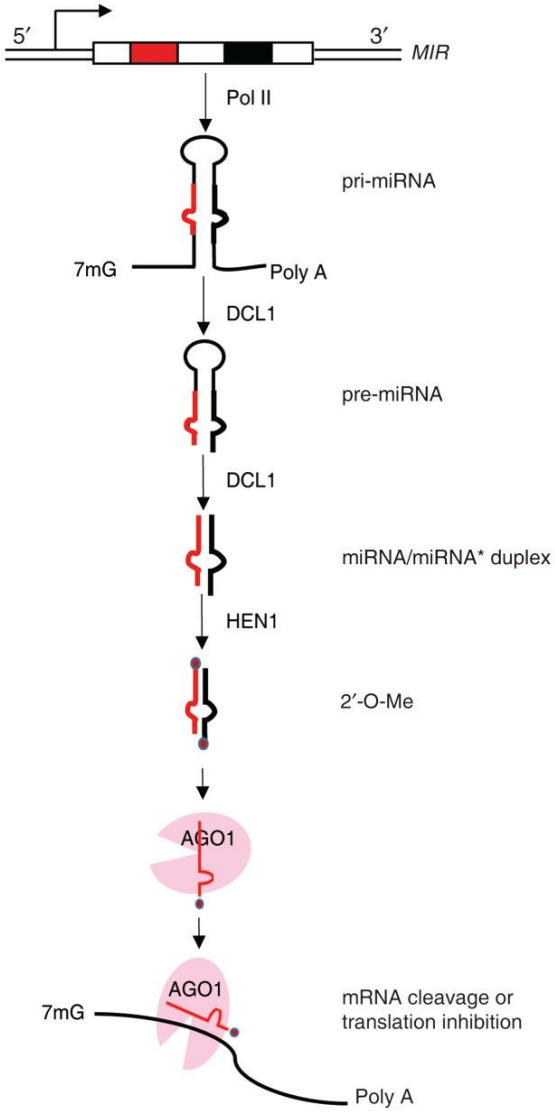


Figure 1.4: **An overview of the key steps in miRNA biogenesis.** This image shows the process and key proteins involved from MIRNA transcription to a mature miRNA. Polymerase II (Pol II) transcribes the MIR encoding loci, producing primary miRNA transcripts (pri-miRNAs) which are processed into a miRNA/miRNA* duplex via DICER-LIKE1 (DCL1). HUA1 ENHANCER1 (HEN1) methylates the 3′ ends of both strands of the duplex to prevent degradation. The mature guide miRNA (in red) is then loaded into ARGONAUTE1 (AGO1) to guide target cleavage or inhibit translation. Figure taken from S. Zhang, Y. Liu, and B. Yu (2015)

.

MicroRNAs (miRNAs) are short, 20–21 nucleotide non-coding RNAs that act as key post-transcriptional regulators of gene expression. Since their discovery in the nematode *Caenorhabditis elegans*, where the small RNA let-7 was shown to control a developmental timing transition, miRNAs have emerged as central regulators of developmental patterning across eukaryotes. The identification of conserved miRNAs in plants soon established that small RNA–mediated regulation represents a deeply conserved layer of gene control.

In plants, miRNAs are processed from stem–loop precursor transcripts and incorporated into ARGONAUTE-containing RNA-induced silencing complexes (S. Zhang, Y. Liu, and B. Yu 2015). Through near-perfect sequence complementarity, they typically guide cleavage of target mRNAs, although translational repression can also occur. Through transcript cleavage or inhibition of translation, miRNAs reduce target gene output, enabling fine-scale tuning of transcription factor abundance and buffering of developmental transitions against environmental or stochastic variation. This mode of regulation enables precise, quantitative control of transcription factor abundance and allows rapid adjustment of developmental programs in response to both intrinsic and environmental signals.

Plant microRNAs are central regulators of developmental transcription factors, forming conserved modules that coordinate phase transitions, organ patterning and hormone responsiveness (Table 1.1). Due to their near-perfect complementarity to target binding sites, plant miRNAs can regulate entire transcription factor families. This means that many miRNA families and their targets are conserved across angiosperms (Qin, P. Xu, and Jiao 2025).

One example is the miRNA regulatory module controlling developmental phase transitions and organogenesis. The miR156–SPL and miR172–AP2-like modules function antagonistically to integrate developmental age with floral competence, forming a conserved ageing pathway (Aukerman and Sakai 2003; Chuck et al. 2007). Other regulatory modules define spatial patterning during organogenesis, such as miR164–NAC which regulates meristem boundaries (Laufs et al. 2004), miR165/166–HD-ZIP III which establishes tissue polarity and vascular identity (L. Williams et al. 2005), and miR390-derived tasiRNAs which reinforce polarity and lateral root development (Fahlgren, Montgomery, et al. 2006). Together, these modules demonstrate how small RNAs act at key transcriptional nodes to stabilise developmental progression.

Other miRNA regulatory modules are involved in hormone signalling. miR160 and miR167 regulate ARF transcription factors during root and reproductive development (M.-F. Wu, Tian, and Reed 2006; Bustos-Sanmamed et al. 2013), while miR393 modulates auxin receptor abundance and auxin homeostasis (Parry et al. 2009). Additional conserved modules, including miR396–GRF which controls cell proliferation (Debernardi, Rodriguez, et al. 2012) and miR169–NF-YA which influences root architecture (Sorin et al. 2014), further illustrate the broad developmental roles of plant miRNAs. The

conservation of these families across angiosperms (as shown in Table 1.1) highlights the evolutionary conservation of these regulatory interactions.

Because these miRNAs target core developmental regulators, shifts in their abundance are likely to reflect changes in developmental timing, tissue identity or physiological state. In polyploid species such as *Brassica rapa*, genome duplication further increases regulatory complexity by expanding both miRNA loci and their potential targets (Sun et al. 2015) . This can lead to differences in regulatory dosage and expression pattern divergence between cultivars, thus contributing to their developmental differences.

For this reason, analysing only the ageing pathway miRNAs could overlook broader changes in gene regulation during floral transition. Thus, a small RNA sequencing approach will give us a wider picture of miRNA expression dynamics throughout the floral transition, and identify both conserved developmental regulators and cultivar-specific variation, before testing our specific ageing pathway hypothesis.

Function	miRNA	Target	Species
Phase change and flowering	miR156	SPL family	Arabidopsis, *Zea mays*
	miR171	SCL	Arabidopsis, barley
	miR172	AP2 family	Arabidopsis, *Zea mays*, *Oryza*, *H. vulgare*, *S. tuberosum*
Auxin-mediated development	miR167	ARFs	Arabidopsis, *Oryza*
Meristem identity	miR164	NAC family	Arabidopsis, *Zea mays*, *Oryza*
	miR165/166	HD-ZIP III	Arabidopsis
	miR394	LCR	Arabidopsis
Leaf and root development	miR160	ARFs	Arabidopsis, *Medicago truncatula*, *Zea mays*
	miR319	TCP family	Arabidopsis, *Solanum lycopersicum*
	miR390	TAS3	Arabidopsis
	miR393	TIR1 and AFB	Arabidopsis, *Oryza*
	miR396	GRF	Arabidopsis, *Medicago*, *Oryza*
	miR847	IAA28	Arabidopsis
	miR169	CBF and NF-YA family	Arabidopsis, *Antirrhinum majus*, *Zea mays*
Stomatal patterning	miR824	AGL16	Arabidopsis
Reproductive development (anthers, seeds)	miR159	GAMYB	Arabidopsis
Grain yield	miR397	OsLAC	*Oryza*
Secondary metabolism and fiber traits	miR828 and miR858	MYBs	Cotton, grapes

Table 1.1: MiRNAs and their roles in plant development. This table was adapted from Dong, B. Hu, and C. Zhang (2022).

1.6 UNDERSTANDING HOW MICRORNAS REGULATE FLOWERING IN RAPID-CYCLING B. RAPA

Since the expression differences of major flowering regulators such as *FT, FLC* or *SOC1* do not fully account for the differences in floral transition timing between R-o-18 and Sarisha-14, this suggests that additional layers of regulatory control may be underlying the differences. Another class of gene regulators are microRNAs, which post-transcriptionally regulate gene expression, and are capable of modulating transcription factor abundance, balancing gene expression dynamics, and coordinating developmental phase transitions.

Previous analyses in Sarisha-14 and R-o-18 relied primarily on mRNA and pri-miRNA measurements (Calderwood, Hepworth, et al. 2021). However, functional repression is mediated by mature microRNAs incorporated into ARGONAUTE-containing complexes, not the pri-miRNA transcribed. Thus, the abundance of pri-miRNAs does not necessarily reflect levels of active mature miRNAs due to variation in processing efficiency and precursor stability. Thus, direct quantification of mature microRNAs is therefore required to accurately capture their regulatory activity.

An exploratory small RNA sequencing approach will allow us to characterise the small RNA landscape during the floral transition, similar to Calderwood, Hepworth, et al. (2021)'s gene expression time series. With this approach, we can detect both conserved developmental regulators (such as the ageing miRNAs, miR156 and miR172) and potentially cultivar-specific or dosage-sensitive miRNAs that may contribute to the flowering time divergence between the cultivars.

Studying this in the diploid *B. rapa* is particularly interesting, since the Brassica whole genome triplication and subsequent gene retention would have expanded regulatory complexity (Sun et al. 2015). Therefore, miRNA-mediated modulation of flowering time can occur not just in the ageing pathway, but also at other nodes of the network or possibly in parallel to the flowering network, such as FT responsiveness or growth signalling.

Thus, by first establishing a comprehensive profile of mature miRNA dynamics, this thesis provides a foundation for how small RNAs can be regulators of flowering time and consequently lead to earlier flowering in Sarisha-14.

1.7 THESIS HYPOTHESIS AND OBJECTIVES

This thesis will investigate whether divergence in flowering time between rapid-cycling *Brassica rapa* cultivars is associated with differences in mature microRNA regulation during vegetative development and the floral transition.

To address this, I performed shoot apex small RNA sequencing across developmental time courses in Sarisha-14 and R-o-18, integrated mature microRNA abundance with transcriptomic data, and analysed a segregating flowering population to identify regulatory differences between the two cultivars.

In this thesis, I hypothesise that differences in mature microRNA abundance during vegetative development and floral transition contribute to flowering time divergence by modulating the activity of key flowering-time gene families. Specifically, differences in microRNA abundance may alter threshold responses within flowering-time gene networks, potentially impacting the timing at which floral transition is initiated in *B. rapa*.

The objectives of this thesis were to:

1. **Adapt a small RNA extraction, sequencing, and bioinformatics workflow** for quantifying microRNA abundance in the shoot apex tissue of *B. rapa*.

2. **Characterise the mature microRNA landscape of Sarisha-14 and R-o-18** during vegetative development and floral transition, identifying differentially abundant miRNAs, including conserved developmental regulators and cultivar-specific abundance profiles.

3. **Integrate microRNA temporal abundance profiles with existing transcriptomic data** to infer potential regulatory relationships and assess whether microRNA dynamics correspond to shifts in target gene expression associated with floral transition.

4. **Investigate the role of ageing-associated miRNAs** (e.g., miR156, miR172) in regulating flowering time differences between cultivars.

5. **Identify candidate genomic loci associated with earlier flowering** in Sarisha-14 using bulked segregant analysis of an F2 cross with R-o-18.

1.8 THESIS OVERVIEW

This thesis will investigate how flowering time is regulated in rapid-cycling *B. rapa*, and the regulatory mechanisms which allow Sarisha-14 to flower rapidly. A key theme in this thesis is that temporal regulation of gene expression leads to an earlier shift in the timing of the floral transition, which I investigate in the following three chapters.

Chapter 2 describes the process of small RNA extraction, sequencing and miRNA identification used for miRNA analysis in subsequent chapters. A small RNA time-series dataset was generated by sampling R-o-18 and Sarisha-14 at regular intervals. This chapter also describes the process of optimising RNA extraction for *B. rapa* shoot apices, testing new sequencing technologies, and most importantly, development of a bioinformatics pipeline to identify *B. rapa* miRNAs. A description of the small RNA timeseries dataset was included.

In Chapter 3, an exploratory analysis was performed to identify miRNAs associated with earlier flowering in Sarisha-14. MiRNA expression profiles were clustered using fuzzy c-means clustering to identify miRNA groups which share similar temporal expression dynamics. Using differential gene expression analysis and miRNA target gene prediction, several case studies of potential miRNAs are presented.

In Chapter 4, I compared miR156 and miR172 expression to test the hypothesis that their expression ratio coincides with the floral transition timepoint. This includes miR156 and miR172 sequence identification, small RNA expression profiling, and target gene prediction and comparisons between miRNA and target gene expression.

In Chapter 5, R-o-18 and Sarisha-14 were crossed to generate an F2 population with segregating flowering times. DNA sequencing was performed on pooled F2 individuals with early flowering phenotype and late flowering phenotype. From this sequencing data, SNPs associated with early flowering phenotype were identified within candidate genomic loci.

Lastly, in Chapter 6, I discuss the findings from the previous chapters to explore how flowering time is regulated in Sarisha-14 and R-o-18. This includes a description of the small RNA timeseries, insights into conserved and cultivar-specific miRNAs, and candidate genomic loci linked to early flowering. I also discuss how these findings contribute to our understanding of miRNA-mediated regulation, the limitations of the study, and the potential applications for crop breeding and improving flowering time in rapid-cycling *B. rapa*.

2

METHODOLOGY FOR MICRORNA EXPRESSION PROFILING

2.1 INTRODUCTION

To study micro RNA expression during *Brassica rapa* development, a small RNA extraction protocol and a miRNA identification method had to be established. The following methods had to be optimised: tissue grinding method, small RNA extraction kit, sequencing technology, and lastly bioinformatic identification of Brassica miRNAs from small RNA sequencing data.

2.1.1 *Description of expected extracted small RNA results*

The aim of this chapter is to describe the optimised small RNA extraction and identification protocol, including the troubleshooting steps and results. Sequencing companies that we worked with (Novogene, Beijing, China; RealSeq Biosciences, California, USA) required total RNA samples with high amount of RNA ($\geqslant$ 2000 ng) and high RNA integrity (RNA Integrity Value (RIN) $\geqslant$ 8.0) to perform small RNA sequencing, thus the small RNA extraction protocol had to be optimised to reduce RNA degradation and produce high quality total RNA input for sequencing.

The sequenced small RNA would have a distribution of small RNA sizes, with micro RNAs between 20-22 nucleotides long, and other small RNA classes such as small interfering RNAs (siRNA) which can be 21, 22 or 24 nucleotides long (Y.-X. Liu, M. Wang, and X.-J. Wang 2014). The length distribution of small RNAs in plants often has a high peak at 24nt, followed by a second peak at 21nt, which is the most common plant miRNA length (Y.-X. Liu, M. Wang, and X.-J. Wang 2014). This is also found in a *B. rapa* miRNA characterization study by B. Kim et al. (2012) and in *B. oleracea* small RNA (Hui Li et al. 2017), thus we would expect a similar result in our small RNA dataset.

2.1.2 *Brassica oleracea accession DH1012 used for method optimisation*

Prior to performing small RNA sequencing of the two rapid-cycling *B. rapa* cultivars (R-o-18 and Sarisha-14), the rapid-cycling *Brassica oleraca* double haploid accession, DH1012, was used for testing RNA extraction methods. A preliminary study of vernal-

ization response in a diversity set of *B. oleracea* found that the *MIR172b* loci was associated with the lack of vernalization requirement in this specific accession (Woodhouse et al. 2021). Thus, we wished to identify other miRNAs expressed in this rapid-cycling cultivar that may be shared with the rapid-cycling *B. rapa* cultivars, thus potentially finding conserved regulatory mechanisms during rapid-cycling Brassica floral development. In addition, this would allow us to test our "crucial ratio" hypothesis regarding miR156 and miR172 expression coinciding with the floral transition, in two Brassica species. This accession also had the advantage of growing larger shoot apex tissue compared to the *B. rapa* cultivars, easing tissue dissection and providing more tissue which should increase the concentration of extracted total RNA. This allowed us to establish a successful protocol for Brassica small RNA extraction and analysis, with minor changes to the pipeline depending on the species used for sequencing.

The objectives of this chapter were to:

1. Establish a small RNA extraction protocol which yielded high quality small RNA for sequencing and miRNA identification
2. Utilise bioinformatic tools to identify miRNA sequences expressed in the samples

2.2 MATERIAL AND METHODS

2.2.1 *Plant material generation*

2.2.1.1 *Plant growth conditions*

Rapid-cycling *Brassica rapa* (R-o-18 and Sarisha-14) and *Brassica oleracea* (DH1012) plants were sown in 24 cell trays in cereals mix (40% medium grade peat, 40% sterilised soil, 20% horticultural grit, 1.3 kg/m^3 PG mix 14-16-18 + Te base fertiliser, 1 kg/m^3 Osmocote Mini 16-8-11 2 mg + Te 0.02% B, wetting agent, 3 kg/m^3 maglime and 300 g/m^3 Exemptor). Material was grown in a Conviron MTPS 144 controlled environment room with Valoya NS1 LED lighting (250 µmol m^{-2} s^{-1}) 18 °C day/15 °C night, 70% relative humidity with a 16 hour day (lights on from 00.30hrs until 16.30hrs).

2.2.1.2 *DH1012 sampling*

To test different tissue grinding methods and RNA extraction kits, DH1012 tissue was used for this trial, due to its larger apex tissue compared to the *B. rapa* samples. 48 DH1012 plants were grown in CER conditions as described above. When the plants reached 14 days old, apex tissue was dissected and pooled (4 apices per pool) in RNAse-free 1.7 mL microtubes over ice, then flash frozen in liquid nitrogen. In total, 12 replicates were collected for RNA extraction protocol optimisation. The samples were stored in -70 °C until RNA extraction was performed.

2.2.1.3 *R-o-18 pilot sampling*

To test new sequencing technologies (Sequence Suppression Probe) on *B. rapa* small RNA, R-o-18 apex and leaf tissue was sampled. 16 R-o-18 plants were grown in CER conditions as described above, and at 11 days after sowing (when the first true leaf had emerged), the apex and first emerged true leaf tissue were sampled. Shoot apices were dissected over ice and pooled (4 apices per pool) in RNAse-free 1.7 mL microtubes over ice, flash frozen with liquid nitrogen, then stored in -70 °C until RNA extraction was performed. The first true leaves from these plants were also sampled (the same 4 pooled plants per replicate), and since they were still small, they were also stored in RNAse-free 1.7 mL microtubes and flash frozen, then stored in -70 °C.

2.2.1.4 *MiRNA timeseries sampling for Sarisha-14 and R-o-18*

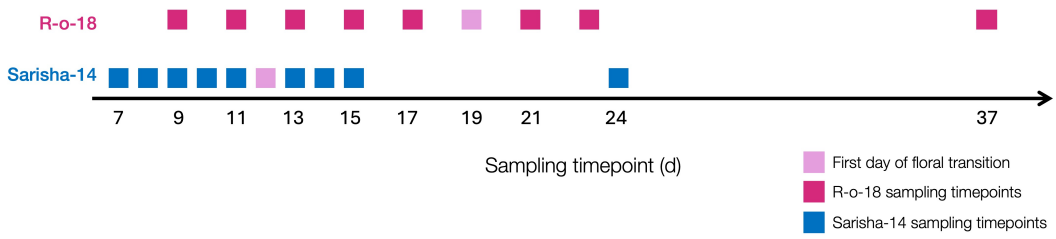


Figure 2.1: **Timeseries**

Sampling of Sarisha-14 and R-o-18 leaf and apex was performed in the 10th hour of the day (ZT10). Sampling was performed on Sarisha-14 every day from 7 days after sowing until 15 days after sowing, with a final sampling timepoint at day 24 which was when the floral buds were visible (developmental stage BBCH51 (Meier et al. 2009))). R-o-18 was sampled every two days from day 9 until day 23, and also at BBCH51 on day 37. These timepoints were chosen to capture small RNA expression prior to and during the floral transition stage, until the plants were considered past the floral transition.

To identify the timepoint when the floral transition occured, 4 plants were inspected at the same time of day as the sampling, by dissection under a light microscope. The floral transition timepoint was marked when at least 3 out of 4 plants showed domed meristems with round floral primordia on the flanks of the apex (Kinoshita et al. 2020), compared to flat vegetative meristems flanked by triangular leaf primordia.

During tissue sampling, apex and leaf tissue were dissected and pooled on ice inside the growth chamber, then flash frozen in liquid nitrogen once all the tissue were pooled for each replicate. Apex tissue were pooled in RNAse-free 1.7 mL microtubes, while leaf tissue was stored in aluminium packets. These samples were then stored in -70 °C storage, until total RNA extraction was performed at a later date.

A total of 9 timepoints were collected from R-o-18, while 10 timepoints were collected for Sarisha-14. For each timepoint, 4 replicates were sampled, with 4 pooled apices/leaf tissue per replicate. Only 3 out of 4 replicates were processed for RNA extraction and sequencing.

2.2.2 *Tissue grinding*

Tissue grinding with steel ball was performed by inserting a fully frozen sterile 5mm steel ball into a microtube containing the tissue to be ground. This was repeated for as many samples as needed, since this method allowed multiple samples to be processed at once. The microtubes were then placed into frozen steel racks packed with dry ice, inserted into a GenoGrinder machine, and set at speed 25 for 1 minute at a time. The samples were considered fully crushed when they appeared as a white powder. The

samples were immediately placed in liquid nitrogen to stay frozen, and were either kept for storage at -70 °C, or used for RNA extraction.

For tissue grinding with a micropestle, a sterile micropestle was attached to a drill. The microtube containing frozen tissue was kept cold by submerging the bottom of the tube in liquid nitrogen, ensuring the sample stayed frozen. The drill attachment was turned on, spinning the micropestle, and it was briefly dipped in liquid nitrogen to keep it cool followed by inserting it into the microtube slowly. The microtube was rotated to ensure the micropestle crushed the tissue fully, and was kept submerged in liquid nitrogen to prevent the sample from thawing. This process took a few minutes, depending on the size of the tissue. When sample grinding was complete, the micropestle was detached from the drill, and left inside the microtube, which was placed in a sterile tube rack immediately followed by addition of lysing reagent (such as TRI reagent ®when using the Direct-zol RNA Microprep kit (kit R2063, Zymo Research https://zymoresearch.eu/), or β-mercaptoethanol with the E.Z.N.A. MicroRNA kit (kit R7034-01, Omega Bio-tek Inc., http://omegabiotek.com/store)) to homogenise the ground tissue before it thawed. RNA extraction was then performed following the respective RNA extraction kit's protocol.

For larger tissues such as leaf tissue, tissue grinding was performed using a ceramic pestle and mortar. Prior to tissue grinding, the ceramic pestle and mortars were wiped down with RNAse-removal spray, wrapped in aluminium foil, and then autoclaved. The sterilised pestle and mortars were then unwrapped and cooled down with liquid nitrogen, until ice crystals formed on the side of the mortar. Frozen leaf tissue was ground using the pestle and mortar mixed with liquid nitrogen to prevent the sample from thawing. Tissue grinding was complete when the tissue was finely ground and pale green in colour. Then, 1 mL of frozen tissue powder was transferred into an RNAse-free 1.7 mL microtube and the excess tissue in a separate 15 mL tube for storage in -70 °C. The 1 mL of frozen tissue was then ready for subsequent RNA extraction.

2.2.3 *Total RNA extraction*

After tissue grinding, RNA extraction was performed using either the E.Z.N.A. MicroRNA kit (kit R7034-01, Omega Bio-tek Inc., http://omegabiotek.com/store), with optional DNAse treatment, or with the Direct-zol RNA Microprep kit (kit R2063, Zymo Research https://zymoresearch.eu/), with TRI reagent and optional DNAse treatment. The E.Z.N.A. MicroRNA kit was used for small RNA protocol testing using *B. rapa* apex and leaf tissue (subsection 2.3.1), and DH1012 leaf and apex RNA extraction (subsection 2.3.3). For the rest of the sequencing using RealSeq, (subsections 2.4.2 and 2.5), the Direct-zol RNA Microprep kit was used.

2.2.4 *Small RNA sequencing*

2.2.4.1 *Novogene small RNA sequencing*

Total RNA samples were sent to Novogene, Beijing, China, where 50 bp single-end reads were generated using the Illumina NovaSeq 6000 platform. An average of 10 M reads were generated per library.

2.2.4.2 *RealSeq Biosciences small RNA sequencing*

Total RNA samples were sent to RealSeq Biosciences, California, US, where 75 bp single-end reads were generated with the Singular Genomics G4 platform. An average of 30 M reads were generated per library.

2.2.5 *Small RNA-seq data processing and miRNA identification*

This section describes the bioinformatic analysis pipeline for identification of *B. rapa* miRNAs during analysis of the small RNA timeseries in R-o-18 and Sarisha-14, as well as R-o-18 samples during RNA sequencing optimisation with RealSeq. This pipeline is adapted from Garg and Varshney (2022)'s small RNA analysis protocol. A visual summary of the pipeline is presented below (Fig. 2.2).

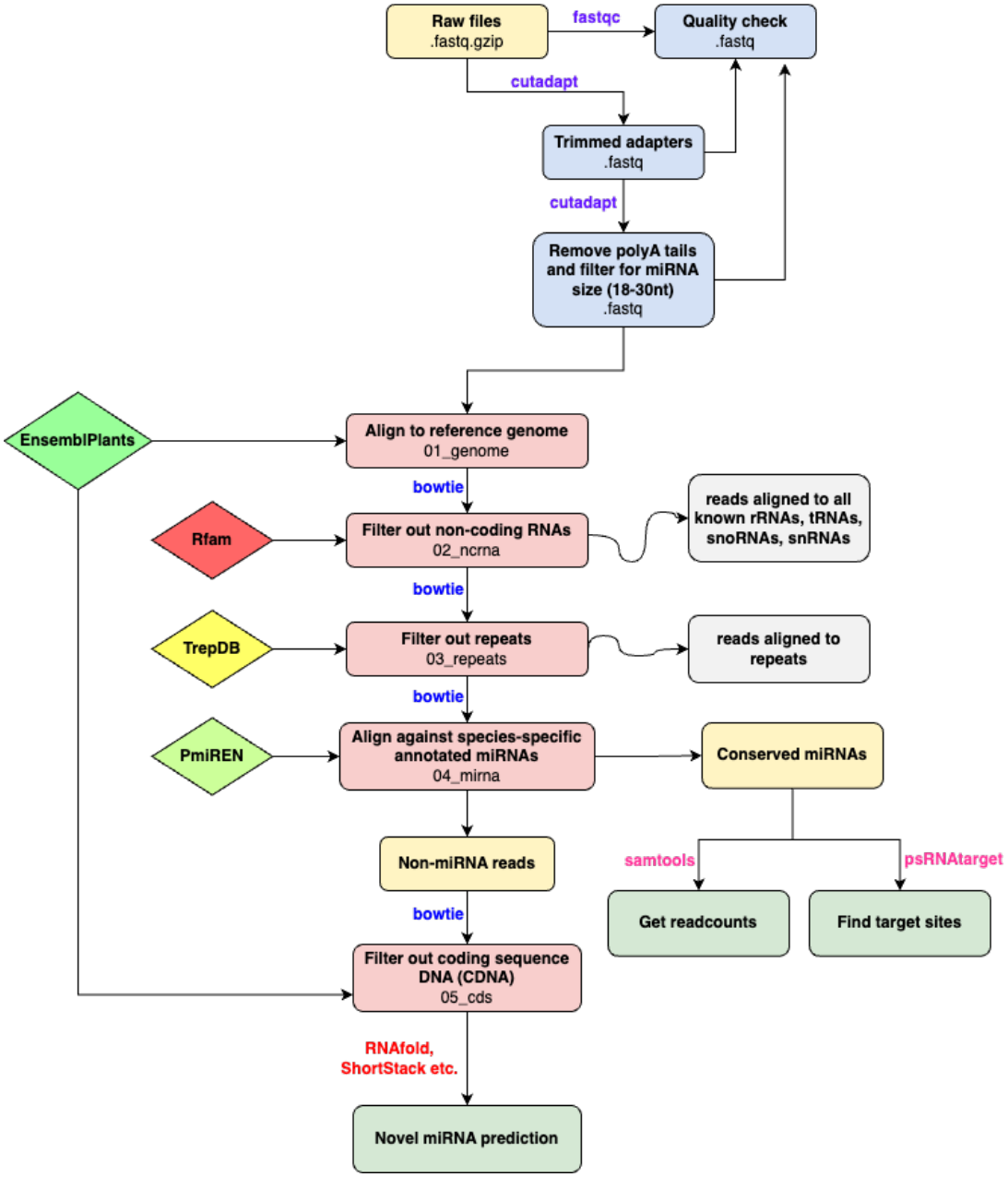


Figure 2.2: **MiRNA identification workflow for identification of *B. rapa* miRNAs.**

High Performance Computing Environment

All analyses were run via batch processing on the JIC High Performance Computing (HPC) cluster with SLURM job scheduler. Jobs were submitted to the `jic-long` and `jic-medium` partitions, requesting 6 CPUs, 40 GB RAM of memory, and 60 GB SSD scratch space per array task. Individual samples were processed as array jobs to parallelise analysis. This HPC setup was referenced from Dr Hugh Woolfenden's BRAVO RNA-seq alignment workflow (https://git.nbi.ac.uk/morris-group/bravo-scripts/-/tree/master/alignment_scripts).

Software used

All analyses were performed in a controlled conda environment to ensure reproducibility. The key tools and their versions are listed below:

- `FastQC v0.11.9` https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ – read quality assessment.
- `Cutadapt v2.10` (Martin 2011) – adapter removal and quality trimming.
- `Bowtie v1.2.2` (Langmead et al. 2009) – alignment of small RNAs to the reference genome. Bowtie1 was chosen over Bowtie2 and other sequence aligners as it is the most suitable for single-end short read alignment.
- `Samtools v1.10` (Danecek et al. 2021) – alignment file manipulation and filtering.
- `HTSlib v1.17` (Bonfield et al. 2021) – core library for high-throughput sequencing file formats.

The versions of these packages were automatically chosen during `conda install` to maintain compatibility within the same conda environment.

Reference databases

The following databases were used for genome alignment, small RNA filtering, and miRNA identification.

- **EnsemblPlants *Brassica rapa* genome (R-o-18)**: Genome assembly and coding sequences for *Brassica rapa* R-o-18 were obtained from EnsemblPlants (release 57; accessed 28th October 2024). Used as the reference genome for read alignment and downstream mapping of small RNA sequences.
  FTP link for reference genome:
  https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/fasta/brassica_rapa_ro18/dna/Brassica_rapa_ro18.SCU_BraROA_2.3.dna.toplevel.fa.gz
  FTP link for coding sequences:
  https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/fasta/brassica_rapa_ro18/cdna/Brassica_rapa_ro18.SCU_BraROA_2.3.cdna.all.fa.gz
- **Rfam**: Rfam database (release 14.10, Kalvari et al. (2021)) was used to identify and filter non-coding RNAs (rRNA, tRNA, snRNA, snoRNA) from sRNA-seq data. The list of non-coding RNAs was manually curated and downloaded by entry type from their webpage:
  https://rfam.org/search#tabview=tab5

- **TrepDB**: Transposable Element database (release 19) was used to filter out small RNA reads derived from all known transposable elements. Download link: https://trep-db.uzh.ch/downloads/trep-db_complete_Rel-19.fasta.gz
- **PmiREN**: PmiREN (release 2.0; Guo, Kuang, and X. Yang (2025)) was used for identification of known *B.rapa* miRNA families from sRNA-seq data. The list of mature *B. rapa* miRNAs was manually downloaded from their webpage: https://www.pmiren.com/browsehref1?wzid=138
  Duplicate mature miRNA sequences were removed, and the final list of miRNAs (available at Appendix A.3) was used to create a bowtie index for *B. rapa* miRNAs.

Bowtie indexes from these databases were generated using the command `bowtie-build`.

Raw data processing

Small RNA sequencing libraries were generated on a Singular Genomics G4 sequencing platform (75bp single-end reads). Raw reads were first assessed for quality using `FastQC`. Quality metrics including per-base quality scores, adapter contamination, and length distribution were inspected both before and after trimming to ensure data integrity. Individual FASTQC reports were collated using `MultiQC` (v1.7) (2016) for visualisation of all samples.

Adapters were removed using `Cutadapt` with the sequencing provider's small RNA adapter sequence (TGGAATTCTCGGGTGCCAAGG). Additional parameters include: quality cutoff of 15 (-q 15) and removal of the first base (-u 1). We set a minimum length requirement of 15nt (-m 15), to avoid missing out on potential regulatory small RNAs, and we weren't sure whether R-o-18/Sarisha-14 miRNAs were between 20-22 nt long. After trimming and size filtering, `FastQC` was rerun to confirm enrichment for small RNA-sized reads (18-30 nt).

Genome alignment

Trimmed reads were converted to FASTA format and aligned to the *B. rapa* reference genome using `Bowtie` with zero mismatches permitted, as R-o-18 miRNAs would align perfectly to the R-o-18 reference genome. Reads successfully aligning to the genome were retained, and Bowtie's output logs were saved for alignment statistics.

Read filtering

To remove not relevant small RNA species, genome-aligned reads were aligned to a non-coding RNA database (Rfam release 14.10) and transposable element database (TrepDB release 19). Reads which mapped perfectly (0 mismatches) to rRNAs, tRNAs, snoRNAs, and snRNAs were discarded, and unaligned reads were carried forward for the next alignment step. This step allows for quicker alignment of *B. rapa* non-coding RNA filtered reads to the bra-mirna bowtie index.

Identification of known miRNAs

The filtered reads were aligned to a custom index of *B. rapa* mature miRNA sequences (based on PmiREN, release 2.0) using `Bowtie`, allowing 0 mismatches, to capture only perfectly aligned canonical miRNAs in Sarisha-14 and R-o-18. Aligned reads were converted into sorted BAM format for downstream analyses.

Alignment to CDS and Secondary Genome Mapping (Steps 05–06)

Reads not mapping to known miRNAs were then aligned against coding DNA sequences (CDS) and the R-o-18 reference genome to capture potential novel small RNA-producing loci.

MiRNA expression quantification

For each sample, miRNA aligned BAM files were indexed, and read counts were extracted with `samtools idxstats`. This produced per-sample miRNA expression tables used as input for downstream expression analyses.

Normalization and Downstream Analyses

For downstream clustering analysis like Principle Component Analysis (PCA), raw read counts were variance-stabilized using the DESeq2 function varianceStabilizing-Transformation (DEseq2 version 1.48.2), with transformation performed in a blind manner to avoid model-dependent bias. This variance stabilising transformation reduces the problem where lowly expressed miRNAs appear much noisier than highly expressed ones, by re-scaling the data so that variance is more comparable across the full expression range. This allows for more reliable comparison of temporal expression patterns between miRNAs, independent of their absolute abundance, and makes the data suitable for downstream exploratory pattern analyses like PCA.

To account for sequencing batch variation in the data, batch effects were removed from the variance-stabilized data using limma's removeBatchEffect (limma version 3.64.3). PCA was then performed on the standardized counts using the PCA function from the FactoMineR R package (version 2.12).

2.3 OPTIMIZATION OF RNA EXTRACTION METHODS

2.3.1 *Testing RNA extraction kits with B. rapa leaf and apex tissue*

For small RNA sequencing services, our sequencing supplier Novogene required RNA samples with over 2 ug total RNA (https://www.novogene.com/eu-en/services/research-services/transcriptome-sequencing/non-coding-rna-sequencing/small-rna-sequencing-srna-seq). Thus, we had to extract enough RNA from each sampling timepoint, and

opted to pool 4 shoot apices per replicate to ensure enough RNA was extracted.

We tested two different RNA extraction kits for highest miRNA yield and quality. Previous Brassica RNA extraction protocols in our lab used Omega Bio-tek's E.Z.N.A Plant RNA kit, but this kit filters out small RNA and captures RNA over 200nt long. Thus, we tested their E.Z.N.A Micro RNA kit, which has two spin columns that can capture small RNA (<200 nt) and longer RNA (>200nt). This was an appealing option, as it allowed us to create small RNA spiked samples, to hopefully enrich the samples with more miRNAs to be detected during small RNA-sequencing.

We also tested Zymo Research's Direct-zol Microprep Total RNA extraction kit, upon recommendation from Dr Shujuan Xu, a postdoctoral researcher from the Feng lab which studied small RNA expression during pollen development, and Sophie Carpenter, a graduate PhD candidate who studied miRNA expression in wheat development. The advantage of this kit was that it extracted total RNA without filtering out small RNA, and claimed to extract RNA sequences over 17nt long, thus saving time for RNA extraction and simplified sending total RNA to sequencing companies. A summary which compares the two kits is described in table 2.1.

(a) (b)

Figure 2.3: **The RNA extraction kits tested for small RNA extraction from Brassica tissue.** (a) An image of Omega Bio-tek's E.Z.N.A MicroRNA extraction kit, which features two spin-columns that captures small RNA (<200nt) and longer RNA (⩾200nt). (b) An image of Zymo Research's Direct-zol Microprep Total RNA extraction kit, with one spin-column that captures total RNA >17nt in length.

2.3.2 *Tapestation analysis of B. rapa extracted RNA*

To assess the RNA quality of RNA extracted between the Direct-zol and E.Z.N.A. kits, Tapestation analysis was performed on *B. rapa* leaf and apex RNA samples (see Table 2.2 for sample descriptions). The extracted RNA samples were run on an electrophoresis gel alongside a reference ladder, and the RNA quality was determined based on its

	E.Z.N.A. Micro RNA	**Direct-zol Microprep Total RNA**
RNA output	Small RNA (<200nt) and long RNA (⩾200nt)	Total RNA (>17nt)
Price	£249.89	£251.00
Pros	This brand has been successful for RNA extraction from Brassica tissue	Straightforward protocol and eluted total RNA can be sent for sequencing immediately
Cons	Multiple spin-columns and elution steps makes protocol longer (>4 hours)	Not yet tested for RNA extraction from Brassica tissue, not sure about RNA quality

Table 2.1: **Comparison between the E.Z.N.A. Micro RNA and Direct-zol Microprep Total RNA extraction kits.** Kits were assessed based on RNA size distribution, RNA yield, protocol complexity, and suitability for small RNA sequencing requirements.

RNA Integrity Number (RINe) was calculated by the Tapestation's software based on electropherogram analysis of ribosomal RNA profiles and RNA degradation ("RNA Quality Assessment with the Agilent Automated Electrophoresis Systems" n.d.).

4 samples were tested, 2 which were leaf RNA extracted using the E.Z.N.A. kit and 2 which were apex and leaf RNA extracted with the Direct-zol kit. For the E.Z.N.A. small RNA spiked total RNA sample (EN_1L), this was created with an approximate 1:18 ratio of small RNA to total RNA extracted from the two spin-columns from the E.Z.N.A. kit.

Lane ID	Sample ID	Kit used	RNA content	RNA conc. (ng/ul)	Tissue	Cultivar	Age (d)
B1	EN_1L (s)	E.Z.N.A	Small RNA only	263	Leaf	R-o-18	12
C1	EN_1L	E.Z.N.A	Spiked total RNA	140	Leaf	R-o-18	12
D1	DZ_1A	Direct-zol	Total RNA	151	Apex	Sarisha-14	15
E1	DZ_1L	Direct-zol	Total RNA	90.4	Apex	Sarisha-14	15

Table 2.2: **Description of RNA samples sent for RNA quality testing.** Lane ID column refers to the lane ID in the Tapestation gel image result (Fig. 2.4). Sample EN_1L is an approximate 1:18 mixture of small RNA to long RNA. RNA quality testing with the Tapestation 4200 was performed by Dr Saleha Bakht from the JIC Molecular Genetics platform.

Sample EN_1L (s) which is the small RNA extracted from R-o-18 leaf tissue using the E.Z.N.A. kit, showed a very strong band between 25-200nt, which confirms that

this kit is able to extract small RNA under 200nt (Fig. 2.4, lane B1). Not surprisingly, this sample had very low RNA Integrity value (RINe) of 2.4, due to the lack of larger RNA bands and an intense small RNA band - in standard RNA extraction cases this would be considered severely degraded RNA (Fig.2.4 and fig.2.5a).

For a more fair comparison between the kits, EN_1L and DZ_1L RNA are the most suitable samples, since they are both total RNA samples from leaf tissue, in contrast with EN_1L(s) which is just small RNA or DZ_1A which is apex total RNA. EN_1L and DZ_1L have similar RNA intensity profiles (Fig.2.5b and Fig.2.5d), and the same RINe value of 7.4 (Fig.2.4, lane C1 and E1). This suggests that both kits extract total RNA of similar quality.

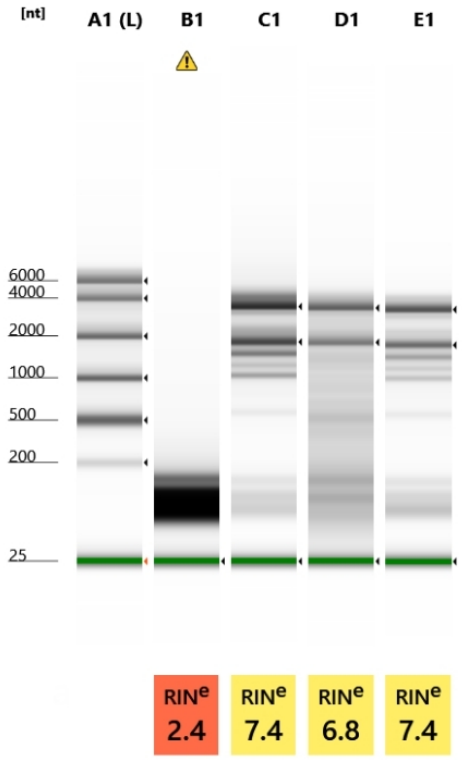
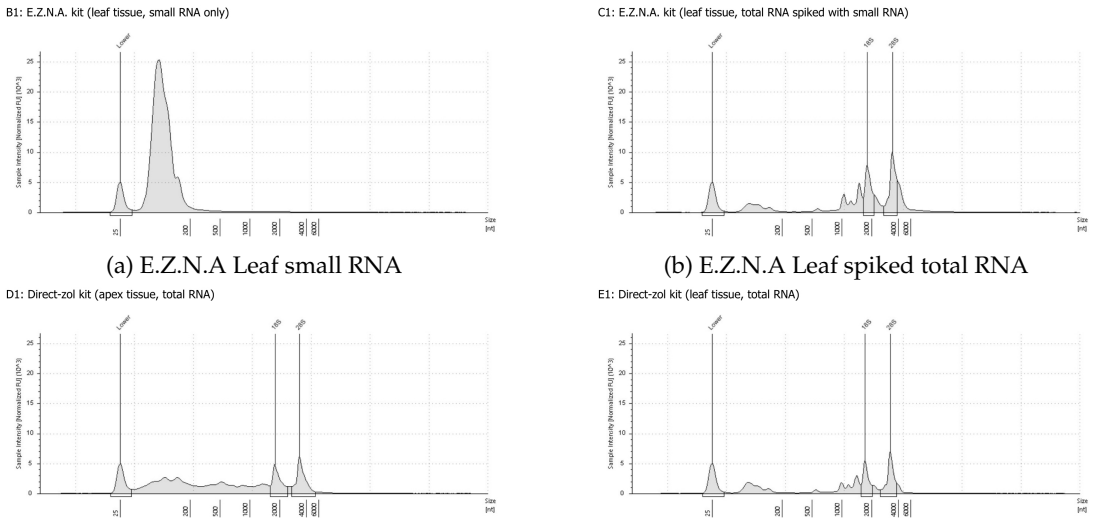


Figure 2.4: **The E.Z.N.A. extraction kit successfully extracts small RNA < 200 nt, but miRNA cannot be detected with this Tapestation.** The image above is the electropherogram of RNA samples extracted with different kits. The values shown at the bottom are the respective RNA quality (RNA Integrity, RINe) determined by the Tapestation software. A good quality RNA sample for small RNA sequencing should have a RINe value above 8.0, with no RNA degradation (seen as a smear under the 200 nt line). Lane A1 is the reference RNA ladder. Lane B1 is small RNA extracted by the E.Z.N.A kit (*EN_1L (s)*), C1 is spiked total RNA from the E.Z.N.A. kit (*EN_1L*), while D1 (*DZ_1A*) and E1 (*DZ_1L*) are apex and leaf total RNA extracted with the Direct-zol kit. Lane B1 shows that the E.Z.N.A. kit successfully extracts small RNA <200nt, with a strong band between 200 and 25 nt, however, miRNAs are between 20-22nt which lies outside the reference ladder. Lane B1 which had just small RNA has the lowest RINe value, which is due to the Tapestation identifying RNA under 200nt as degraded RNA. Both EN_1L and DZ_1L have similar RINe value of 7.4, while DZ_1A which has less distinct RNA bands has a slightly lower RINe value of 6.8. This is likely due to low RNA input from the small volume of apex tissue used for Direct-zol RNA extraction.



(a) E.Z.N.A Leaf small RNA

(b) E.Z.N.A Leaf spiked total RNA

(c) Direct-zol Apex Total RNA

(d) Direct-zol Leaf Total RNA

Figure 2.5: **EN_1L (b) and DZ_1L (d) show similar RNA intensity profiles, with a short and broad peak between 25-200nt.** The x-axis of these plots indicate the RNA size in the sample, and the y-axis indicates the intensity of the RNA band based on the gel image in Fig.2.4. Figure (a) corresponds with sample EN_1L (s) at lane B1 in Fig.2.4, and consistent with the gel image, shows a strong peak between 25-200nt. DZ_1A (c) has less distinct peaks between 25-200nt, but a visibly larger area of RNA intensity compared to EN_1L (b) and DZ_1L (d).

We also tested apex RNA extracted using the Direct-zol kit, since the apex tissue is our main focus for understanding miRNA expression during the floral transition. The RNA quality of apex tissue RNA (DZ_1A) is not as good as the leaf tissue RNA (EN_1L and DZ_1L), as shown in the gel image - it does not have as clearly defined bands as sample EN_1L and DZ_1L (Fig.2.4, lane D1 for DZ_1A, and lanes C1 and E1 for EN_1L and DZ_1L respectively). The peak distributions of DZ_1A (Fig.2.5c) are also not as pronounced as EN_1L (Fig.2.5b) and DZ_1L (Fig.2.5d), especially in the 200-2000 nt range. This suggests either RNA contamination or degradation during the RNA extraction process. It should be noted that the apex tissue is very small compared to leaf tissue, which could contribute to more sensitivity to RNA degradation due to the smaller volume of RNA extracted. Thus, we decided for future apex extractions we would pool 4 individual plants per replicate, to increase the volume of RNA extracted with the advantage of normalising the sampling series.

Plant miRNAs are typically 20–22 nucleotides in length, but the Tapestation analysis used here only detected RNA sequences longer than 25 nucleotides. As a result, this method could not directly assess the efficiency of miRNA extraction, since they escape detection via this method. This is a limitation due to the type of Tapestation RNA analysis kit used, which was not sensitive enough to detect different sizes of small RNAs. A more sensitive kit called the High Sensitivity RNA Kit (https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-rna-screentape-reagents/high-sensitivity-rna-screentape-analysis-228267) would have been more suited for our

analysis, but was not available at the time. Based on the RNA profiles, both kits appeared to perform comparably in isolating small RNAs from leaf tissue. In addition to this, the apex tissue sampled was very small in size, where grinding it in the pestle and mortar was ineffective in extracting all the RNA from the tissue. Thus, we decided to extract more RNA samples using different grinding methods, followed by small RNA sequencing to quantify miRNA recovery across different grinding methods and extraction kits.

2.3.3 *Testing grinding methods and extraction kits with B. oleracea apex tissue*

In this section, I assess the ideal combination of RNA extraction kits (Fig. 2.3) and tissue grinding method which extracts the most miRNA from Brassica apex tissue. The two grinding methods tested are:

1. **Micropestle grinding**: A sterile micropestle is attached to a drill and used to crush frozen apex tissue still in the 1.5 mL microtube it was sampled in, while the tube was immersed in liquid nitrogen to keep the tissue frozen.
2. **Steel ball crushing**: A 5mm sterile steel ball is placed inside the microtube with the frozen sample inside, then the microtube is crushed using a GenoGrinder with dry ice to keep the sample cool. Multiple tubes can be processed in one run, allowing for easier batch processing.

For this trial, I used pooled 14 day old apex tissue from DH1012, a *B. oleracea* double haploid line used for experimental research. We chose DH1012 as it had bigger apex sizes than either *B. rapa* cultivars at the same age, to have better success in extracting more RNA. A combination of either grinding method and extraction kit was used to extract RNA from these tissues.

Nanodrop measurements of extracted RNA had similar RNA yields for either grinding methods (see Appendix Table A.1.) Between the two kits, Direct-zol had higher RNA concentrations compared to the separate elutions from E.Z.N.A extractions. However, the total RNA extracted by E.Z.N.A when combining small RNA and large RNA shows that overall there is not much difference in total RNA amount (ng) extracted between the two kits (Appendix Table A.1).

For the samples extracted with E.Z.N.A., a 50/50 ratio of small RNA and large RNA was combined to make total RNA to send for small RNA-sequencing alongside the Direct-zol extracted samples. It was necessary to send total RNA to the sequencing company (Novogene) as they did not accept small RNA for their sequencing pipeline. A description of the samples sent for sequencing can be found in Table 2.3 below.

Sample IDs	Tissue	Age (d)	Grinding method	Extraction kit
dz1a, dz2a, dz3a	Apex	14	Steel ball	Direct-zol
dz7a, dz8a, dz9a	Apex	14	Pestle and drill	Direct-zol
en4a, en5a, en6a	Apex	14	Steel ball	E.Z.N.A
en10a, en11a, en12a	Apex	14	Pestle and drill	E.Z.N.A

Table 2.3: **Description of pilot DH1012 (*B. oleracea*) samples sent for small RNA sequencing to assess miRNA extraction quantity.**

2.3.4 *RNA yield comparison*

As expected from our earlier Tapestation analysis, during sample quality checks Novogene flagged most of the E.Z.N.A. spiked total RNA samples as low RNA integrity and failed their sample requirements (Figure 2.6). Nevertheless, we moved forward with small RNA-sequencing for all of the samples.

No.	Sample Name	Nucleic Acid ID	Concentration(ng/ul)	Volume(ul)	Total amount(ug)	Integrity value	Sample QC Results	Sample QC Memo	Sample volume for electrophoresis(ul)
1	DZ_DH_1A	FKRN230265420-1A	106.000	36.00	3.81600	6.20	Fail	Unqualified RIN	1.00
2	DZ_DH_2A	FKRN230265421-1A	140.000	37.00	5.18000	7.00	Pass	-	0.50
3	DZ_DH_3A	FKRN230265422-1A	143.000	37.00	5.29100	8.10	Pass	-	1.00
4	DZ_DH_7A	FKRN230265423-1A	71.000	34.00	2.41400	8.50	Pass	-	1.00
5	EN_DH_4A	FKRN230265426-1A	67.000	56.00	3.75200	6.50	Fail	Unqualified RIN	1.00
6	EN_DH_5A	FKRN230265427-1A	55.000	44.00	2.42000	6.20	Fail	Unqualified RIN	1.00
7	EN_DH_6A	FKRN230265428-1A	67.000	47.00	3.14900	6.50	Fail	Unqualified RIN	1.00
8	EN_DH_10A	FKRN230265429-1A	60.000	48.00	2.88000	6.50	Fail	Unqualified RIN	1.00

(a) First round of QC

No.	Sample Name	Nucleic Acid ID	Concentration(ng/ul)	Volume(ul)	Total amount(ug)	Integrity value	Sample QC Results	Sample QC Memo	Sample volume for electrophoresis(ul)
1	EN_DH_11A_mix	FKRN23H001007-1A	47.000	60.00	2.82000	7.40	Pass	-	1.00
2	EN_DH_12A_mix	FKRN23H001008-1A	48.000	49.00	2.35200	7.00	Pass	-	1.00
3	DZ_DH_8A_mix	FKRN23H001009-1A	107.000	59.00	6.31300	8.10	Pass	-	1.00
4	DZ_DH_9A_mix	FKRN23H001010-1A	82.000	56.00	4.59200	8.60	Pass	-	1.00

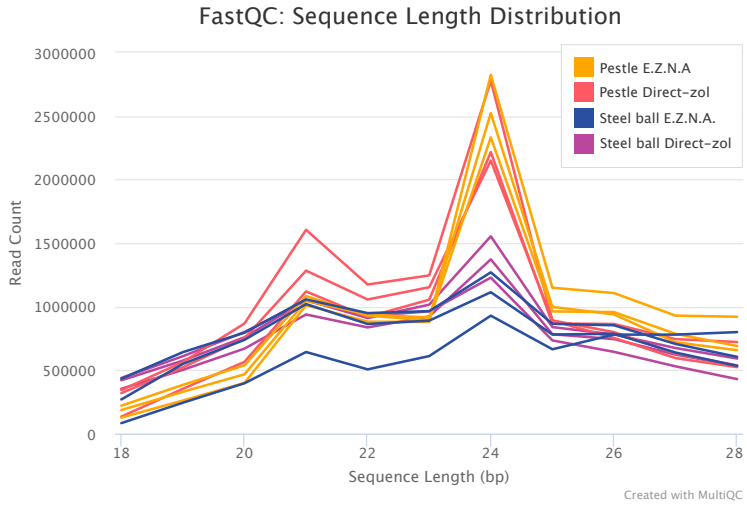
(b) Resent samples

Figure 2.6: **E.Z.N.A. extracted RNA samples failed Novogene's sample quality checks due to spiked small RNA.** The tables above show QC summaries of samples that went through the first round of quality checks (table A), and samples that were resent due to insufficient RNA initially (table B). These summaries were screenshotted directly from Novogene's QC report.

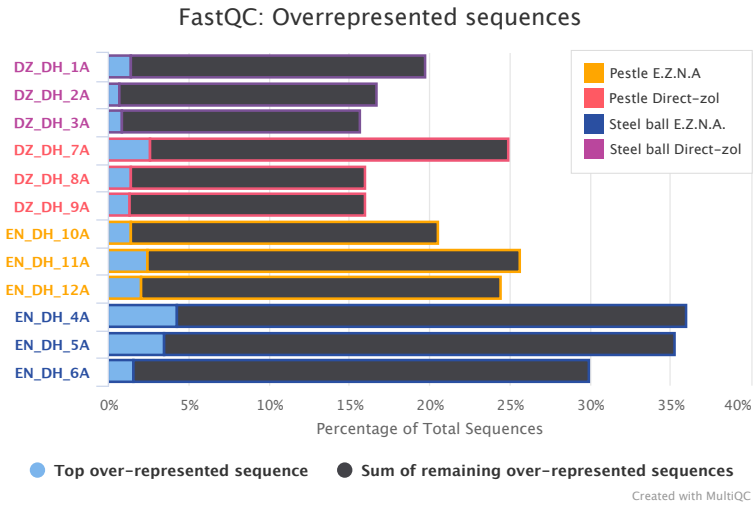
2.3.5 *Bioinformatic analysis of DH1012 samples*

After receiving the raw small RNA-sequencing data of the DH1012 samples, we performed data quality checks using FastQC. Figure 2.7 shows FASTQC results after adapter trimming and size selection to assess the small RNA libraries. Comparison of sequence lengths showed a clear enrichment of 24 nt read lengths across samples (Fig. 2.7a), which is consistent with the expected size distribution of plant small RNA datasets, where 24 nt heterochromatic siRNAs typically represent the most abundant class. The next most abundant size class was 21 nt, which is also consistent with the

predominance of canonical plant miRNAs in this size range. Samples extracted with Direct-zol had lower proportions of overrepresented sequences in the clean library (Fig. 2.7b).



(a)



(b)

Figure 2.7: **Small RNA libraries show the expected enrichment of 24 nt reads, followed by 21 nt reads.** (a) FASTQC Sequence length distribution of all RNA samples. (b) Percentage of overrepresented sequences of all RNA samples. Overrepresented sequences in small RNA libraries occur due to high abundance of non-coding RNA such as ribosomal RNA. Samples extracted with the E.Z.N.A. have a slightly higher percentage of overrepresented sequences.

Our miRNA identification pipeline involves filtering the clean reads, by using bowtie 1.0 to align clean reads against non-coding RNA sequences and repeat sequences to filter out non-miRNA sequences. The filtered reads against known *B. oleracea* miRNA sequences. During the alignments, bowtie 1.0 outputs the total number of reads processed, and the number of aligned and unaligned reads after each filtering step. These values are used for assessing RNA alignment and are shown in Table 2.4.

The majority of the clean reads aligned to non-coding RNA (54% - 81%), while less than 2% of the clean reads are aligned to known *B. oleracea* miRNAs (Fig.2.8, Table 2.4. This low percentage raised our concerns since we were expecting more reads to align to *B. oleracea* miRNAs. In retrospect, this is likely due to a small list of annotated *B. oleracea* miRNAs, where the PmiREN database had 68 annotated *B. oleracea* miRNAs, which is lower than *B. rapa* (290 sequences), *B. napus* (388 sequences) or *A. thaliana* (221) sequences.

Sample	Grind	Clean reads	Genome	NcRNA	miRNA
dz1a	steel_ball	8400207 (100.00%)	7980011 (95.00%)	6513322 (77.54%)	70076 (0.83%)
dz2a	steel_ball	9183204 (100.00%)	8656598 (94.27%)	6603881 (71.91%)	82090 (0.89%)
dz3a	steel_ball	8289074 (100.00%)	7705201 (92.96%)	6040717 (72.88%)	59900 (0.72%)
dz7a	pestle_drill	9504817 (100.00%)	8866794 (93.29%)	5696063 (59.93%)	95779 (1.01%)
dz8a	pestle_drill	11356945 (100.00%)	10523043 (92.66%)	6160773 (54.25%)	200207 (1.76%)
dz9a	pestle_drill	10105887 (100.00%)	9476204 (93.77%)	6616026 (65.47%)	100530 (0.99%)
en4a	steel_ball	6348202 (100.00%)	6057678 (95.42%)	5109061 (80.48%)	13683 (0.22%)
en5a	steel_ball	8182454 (100.00%)	7839765 (95.81%)	6677676 (81.61%)	29545 (0.36%)
en6a	steel_ball	9147449 (100.00%)	8730325 (95.44%)	7293083 (79.73%)	42524 (0.46%)
en10a	pestle_drill	9765414 (100.00%)	9199720 (94.21%)	6329468 (64.82%)	67286 (0.69%)
en11a	pestle_drill	10415237 (100.00%)	9742999 (93.55%)	6152970 (59.08%)	58416 (0.56%)
en12a	pestle_drill	9691277 (100.00%)	9075323 (93.64%)	5808765 (59.94%)	65582 (0.68%)
dz7a *	pestle_dril	9504817 (100.00%)	9155956 (96.33%)	2678378 (28.18%)	15844 (0.17%)
dz8a *	pestle_dril	11356945 (100.00%)	10817437 (95.25%)	3033711 (26.71%)	33781 (0.30%)

Table 2.4: **Almost all samples have very low miRNA content (>2% of clean reads.)** The table shows read alignment summary by grinding method. Raw read counts are shown with its percentage of clean reads in smaller font. Clean reads are raw reads that have been adapter trimmed and size filtered(18-30nt). Samples extracted with the Direct-zol kit have sample ID's starting with "dz", while E.Z.N.A. extracted samples start with "en". NcRNA stands for non-coding RNA sequences. * These two dz7a and dz8a samples are read alignment summaries from Novogene's internal bioinformatic analysis from the same samples, which identified less miRNA than our pipeline.

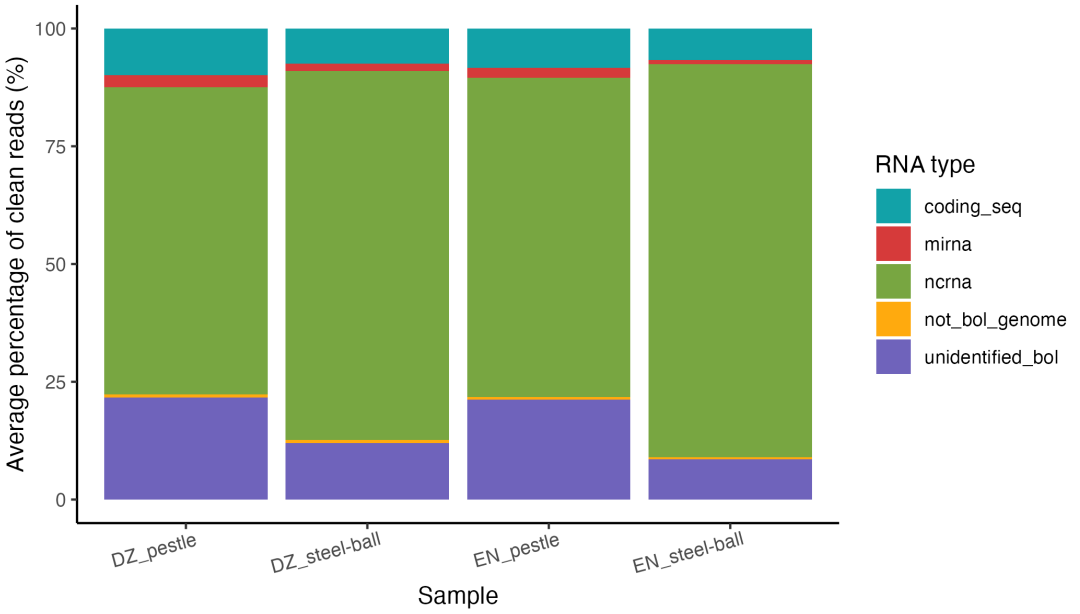


Figure 2.8: **Non-coding RNA makes up a majority of small RNA sequenced in all samples.** This figure shows the average proportion of RNA types within the samples based on grinding and extraction method. RNA types include *B. oleracea* coding sequences, miRNA, non-coding RNA, unannotated *B. oleracea* reads (unidentified_bol), and a small percentage of reads that didn't align to the reference *B. oleracea* genome (not_bol_genome). "DZ" stands for Direct-zol extraction, while "EN" stands for E.Z.N.A. extraction. On average, the pestle grinding method had a higher proportion of miRNA, but not much difference between the extraction kits.

A closer look at the miRNA content in these samples revealed that samples that were ground with the micropestle and extracted with Direct-zol (DZ_pestle) had higher percentages of miRNAs than the other combination of grinding and extraction methods (Fig. 2.9). 2-way ANOVA identified a significant difference between grinding methods (pestle vs steel ball)($p=0.00168$), but no differences between the extraction kits. This suggests that both kits performed equally well in extraction miRNA for small RNA-sequencing and bioinformatic analysis, and higher miRNA quantity depended on the grinding method.

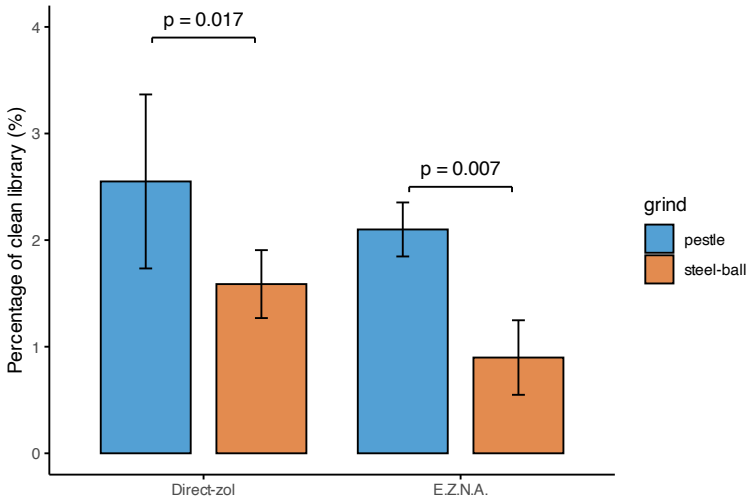


Figure 2.9: **Pestle grinding method extracted the most miRNA compared to steel-ball crushing method.** Each bar represents the extraction kit and grinding method used (4 replicates per bar), and the height of the bar represents the average percentage of the clean library which make up miRNAs. Error bars indicate standard deviation. Two-way ANOVA identified significant effects of kit ($p < 0.01$), but not for grind ($p = 0.95$) and no significant interaction between these two ($p = 0.75$). Post hoc pairwise comparisons (Tukey-adjusted emmeans) indicated that within the Direct-zol kit, the pestle grinding method yielded a significantly higher % of miRNA than the steel-ball crushing method ($p = 0.017$), and it is also significantly higher in the E.Z.N.A. kit ($p = 0.007$).

To conclude, we assessed the quality and ability to extract miRNAs from *B. oleracea* apex samples by a combination of different extraction methods, and found that the pestle and drill grinding combined with Direct-zol extraction kit had the best miRNA yields. We decided to move forward with Direct-zol as it had marginally better miRNA yields, the protocol was more straightforward and it reduced the risk of RNA degradation compared to E.Z.N.A's dual spin-column protocol.

2.4 SEQUENCING MIRNAS

2.4.1 *Concern regarding overrepresented sequences*

We were concerned by the low percentage of identified miRNAs from Novogene's sRNA-seq data of DH1012 apex samples (Table 2.4, Fig. 2.9), so Novogene offered to perform in-house bioinformatic analysis on two samples, dz7a and dz8a. Unfortunately, they identified even less miRNAs than our pipeline (0.17-0.30% of the clean library), so we decided to try an alternative sequencing provider which could improve miRNA detection.

The relatively low percentage of identified miRNAs in Novogene's sRNA-seq data from DH1012 apex samples (Table 2.4, Fig. 2.9) prompted further evaluation of library composition. Novogene performed in-house bioinformatic analysis on two

representative samples (dz7a and dz8a); however, this yielded an even lower proportion of miRNAs (0.17–0.30% of the clean library) compared to our pipeline. Given these results, we considered alternative sequencing strategies to improve miRNA detection.

In previously published Brassica small RNA studies, a substantially higher proportion of reads were annotated as miRNAs. For example, Ahmed et al. (2020) reported 10–33% miRNAs within clean read libraries, and Kong et al. (2022) reported 6–9% of their libraries aligned to known miRNAs. Based on these reports, we expected a higher percentage of miRNAs in the DH1012 libraries.

Since the majority of the library mapped to non-coding RNA sequences besides miRNA (Fig. 2.8), we hypothesised that these could be overrepresented ribosomal RNA sequences which are known to be highly abundant in RNA samples. We therefore selected a sequencing provider employing Selected Suppression Probes (SSP) technology, which targets highly abundant small RNA species during library preparation and enriches for less abundant classes such as miRNAs (Fig. 2.10).

2.4.2 *Testing miRNA enrichment using Sequence Suppression Probes*

For this trial SSP study, we sent 4 pooled apex and 4 pooled leaf RNA samples extracted from 11 day old R-o-18 to RealSeq Biosciences, where they did an initial sequencing round to identify overrepresented sequences in all the samples, and then designed Selected Suppression Probes to remove the overrepresented sequences. They then did two rounds of sequencing on the samples, one without the SSPs, and one with the SSP treatment to assess its impact on miRNA detection.

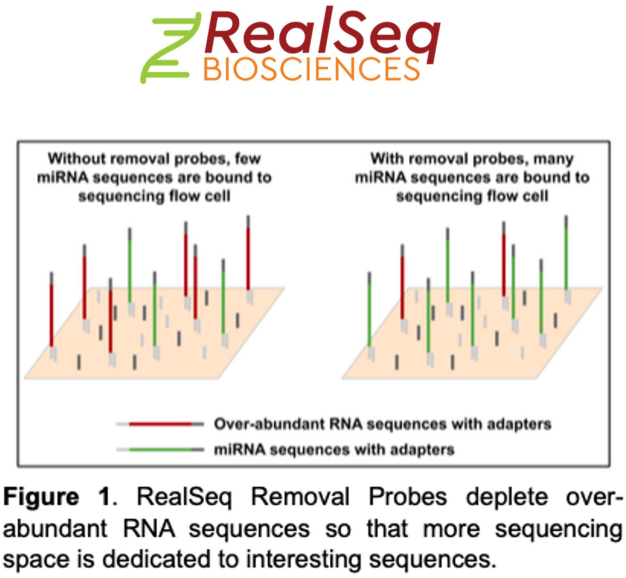


Figure 2.10: **An illustration of RealSeq Biosciences Sequence Suppression Probe technology, which capture overrepresented RNA sequences to enrich miRNAs in small RNA libraries.** Illustration sourced from RealSeq Bioscience's website.

Sample	Genotype	Tissue	Age (days)	Replicate ID
ro1a	R-o-18	Apex	11	1
ro2a	R-o-18	Apex	11	2
ro3a	R-o-18	Apex	11	3
ro4a	R-o-18	Apex	11	4
ro1l	R-o-18	Leaf	11	1
ro2l	R-o-18	Leaf	11	2
ro3l	R-o-18	Leaf	11	3
ro4l	R-o-18	Leaf	11	4
dz1l *	R-o-18	Leaf	14	1
en1l *	Sarisha-14	Leaf	14	1

Table 2.5: **Summary of samples sent to RealSeq Biosciences for SSP method testing to improve miRNA detection in *Brassica* tissue samples.** Apex and leaf tissue were collected from four independently pooled plants (n = 4 biological replicates per tissue type, pool size = 4 individual plants). * dz1l and en1l were not sent to RealSeq Biosciences, but were included in the later analysis to compare Novogene sequencing with RealSeq sequencing.

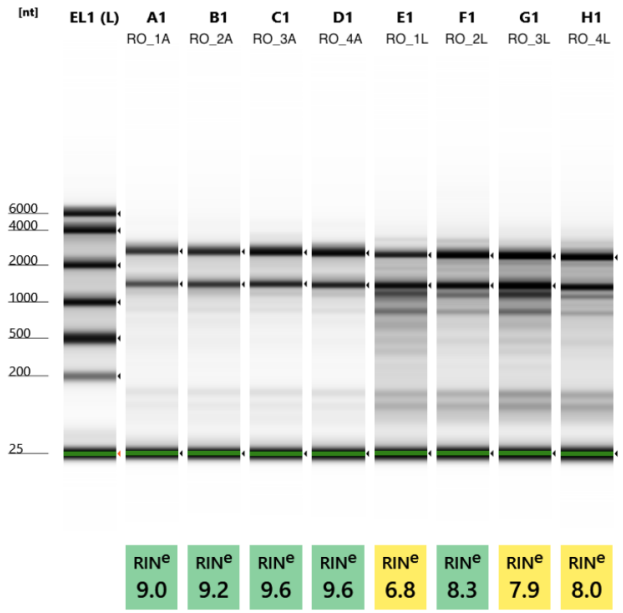


Figure 2.11: **All RNA samples sent to RealSeq Biosciences passed their sample quality checks.** The gel image above shows the RNA band sizes of all the RNA samples compared to the reference ladder in lane EL1. The leaf RNA samples had lower RINe values than the apex RNA samples, but this can be attributed to stronger small RNA bands (under 200nt).

All the RNA samples had passed RealSeq's QC, with high RNA integrity values ranging between RINe 6.8 - 9.6 (Fig. 2.11). The leaf samples had slightly lower RINe than the apex samples, with the electropherogram displaying more RNA fragmentation

in the leaf samples (Fig. 2.11, last 4 columns). This is not unexpected, as the sampled leaf tissue (which was larger in volume than the shoot apices) would have more secondary metabolites and RNAse enzymes compared to the smaller shoot apices, which could contribute to more RNA degradation in the leaf tissue. With this in consideration, we continued with performing small RNA sequencing with all apex and leaf samples.

2.4.3 *Bioinformatic analysis of SSP treated R-o-18 apex libraries*

Sequencing data quality

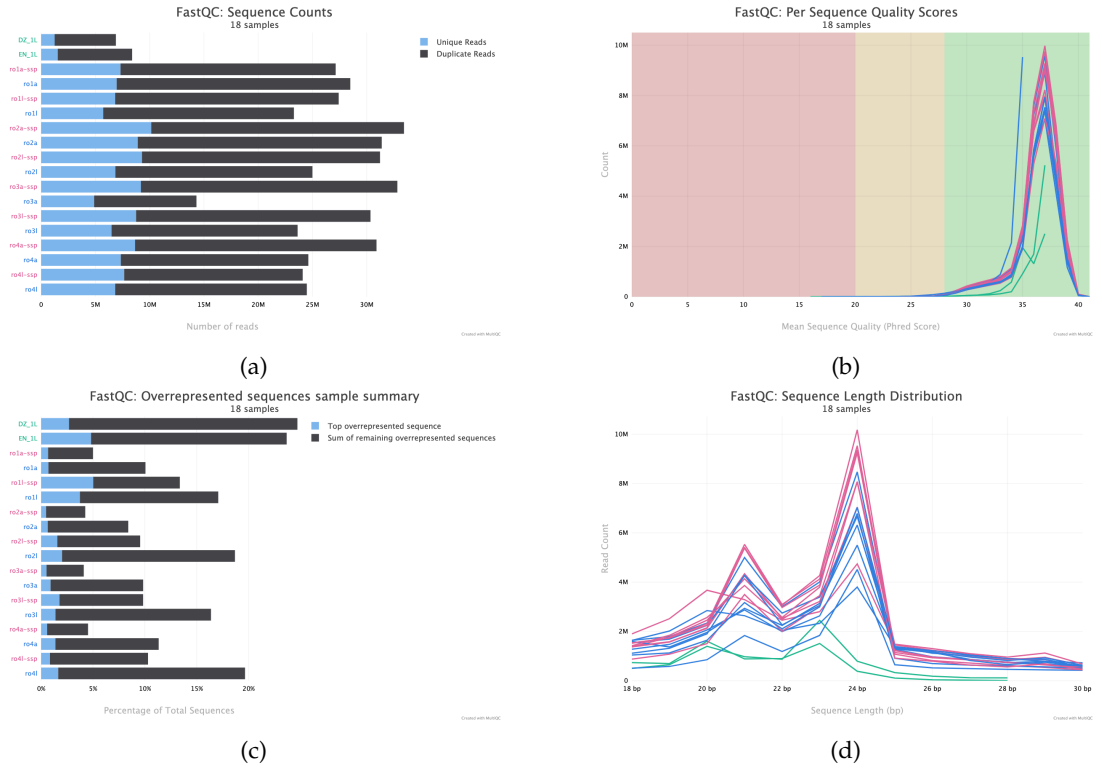


Figure 2.12: **RNA samples sequenced after SSP treatment had lower overrepresented sequences than RNA samples without SSP treatment.** (a) Total sequence counts of all samples including dz11 and en11. (b) Per sequence quality scores of all samples. (c) Summary of overrepresented sequences in all samples. Notice how dz11 and en11 had more overrpresented sequences in the library compared to all samples sequenced by RealSeq. (d) Sequence length distribution of all samples. All RealSeq samples showed a peak at 21nt and 24nt read length, which is consistent with plant miRNA (21nt) and siRNA (24nt) read lengths. For all of these plots, RealSeq samples without SSP treatment are highlighted in blue, samples with SSP treatment are in pink, while Novogene leaf samples are highlighted in green.

After two rounds of sequencing of the 8 samples (first round without SSP treatment, and the second round after SSP treatment), we received 15 raw small RNA-sequencing data files. 1 sample, ro3a, had insufficient RNA so only the second round of sequencing

was performed (ro3a-ssp). Instead, for this analysis I used the initial ro3a sequencing data prior to the first round of sequencing without SSP treatment.

After adapter trimming and size filtering (kept reads between 18-30nt), we ran FASTQC checks on the sequencing data. In this step, we also included two sequencing datasets from *B. rapa* samples that were sequenced by Novogene (dz1l and en1l, see Table 2.5). Please note that Novogene provided 10M sequencing read depth while RealSeq did 30M sequencing read depth.

The FASTQC analysis reported a reduction of total overrepresented sequences in SSP treated libraries (average of 7.6% of total reads) vs 13.9% in non-SSP libraries (average of 13.9% of total reads) (Fig. 2.12c). The novogene sequenced DZ_1L and EN_1L samples had the highest proportion of overrepresented sequences (24.7% and 23.7% respectively), while the apex SSP treated libraries had consistently low over-represented sequences (<4.5%). The length distribution profiles (Fig. 2.12b) showed the expected small RNA profile with two dominant peaks at 21nt (miRNAs) and 24nt (small interfering RNAs, siRNAs) across both the SSP and non-SSP treated libraries, with a slight increased abundance of both peaks in the SSP treated libraries. Thus, we considered all of the RealSeq sequenced samples to have passed FASTQC checks, with no adapters and good sequence quality scores (Fig. 2.12).

sRNA-seq alignment

The SSP treated samples had 25-40% of clean reads aligning to ncRNAs, which is less than non-SSP treated samples (34-40% aligned to ncRNAs), or the Novogene samples (46-51% aligned to ncRNAs)(see Fig.2.13 and Table 2.6). The SSP treated samples also had a small increase in miRNA percentage compared to non-SSP treated samples and Novogene samples, with the exception of ro4a, which had 4.74% reads aligned to *B. rapa* miRNAs compared to ro4a-ssp which had 3.05% reads aligned to miRNAs (Table 2.6).

SSP treated samples had slightly higher miRNA content compared to non-SSP treated samples and Novogene sequenced samples, although the difference is negligible (see Fig. 2.14).

Sample	Clean reads	Genome aligned	NcRNA aligned	miRNA aligned
ro1a	26967627 (100.00%)	22392539 (83.03%)	12465387 (46.22%)	656160 (2.43%)
ro2a	30111660 (100.00%)	24169511 (80.27%)	11081056 (36.80%)	785390 (2.61%)
ro3a	13447728 (100.00%)	11134540 (82.80%)	5026841 (37.38%)	254547 (1.89%)
ro4a	23545159 (100.00%)	19057880 (80.94%)	8212054 (34.88%)	1116906 (4.74%)
ro1l	21860149 (100.00%)	18328808 (83.85%)	9019818 (41.26%)	390866 (1.79%)
ro2l	23610207 (100.00%)	19983580 (84.64%)	9182105 (38.89%)	627067 (2.66%)
ro3l	22079577 (100.00%)	18438930 (83.51%)	7994007 (36.21%)	586193 (2.65%)
ro4l	23128954 (100.00%)	19772504 (85.49%)	8435356 (36.47%)	775092 (3.35%)
ro1a-ssp	25925654 (100.00%)	20994387 (80.98%)	10500427 (40.50%)	822028 (3.17%)
ro2a-ssp	32222361 (100.00%)	25452761 (78.99%)	9884671 (30.68%)	924085 (2.87%)
ro3a-ssp	31652219 (100.00%)	25050139 (79.14%)	11099630 (35.07%)	963240 (3.04%)
ro4a-ssp	29918103 (100.00%)	23263873 (77.76%)	9234941 (30.87%)	911834 (3.05%)
ro1l-ssp	26018492 (100.00%)	21918534 (84.24%)	9703590 (37.29%)	533388 (2.05%)
ro2l-ssp	29658223 (100.00%)	24733745 (83.40%)	8285814 (27.94%)	1003740 (3.38%)
ro3l-ssp	28547728 (100.00%)	23701007 (83.02%)	8168723 (28.61%)	890806 (3.12%)
ro4l-ssp	22954407 (100.00%)	19193052 (83.61%)	5810599 (25.31%)	997902 (4.35%)
dz1l	7593238 (100.00%)	5560316 (73.23%)	3495952 (46.04%)	194826 (2.57%)
en1l	8637565 (100.00%)	7277259 (84.25%)	4407360 (51.03%)	226912 (2.63%)

Table 2.6: **Summary of read alignments from small RNA sequencing of R-o-18 leaf and shoot apex RNA before and after SSP treatment.** Two samples (dz1l and en1l) which were sequenced by Novogene were also included in this analysis. Counts are shown with percentages in smaller font.

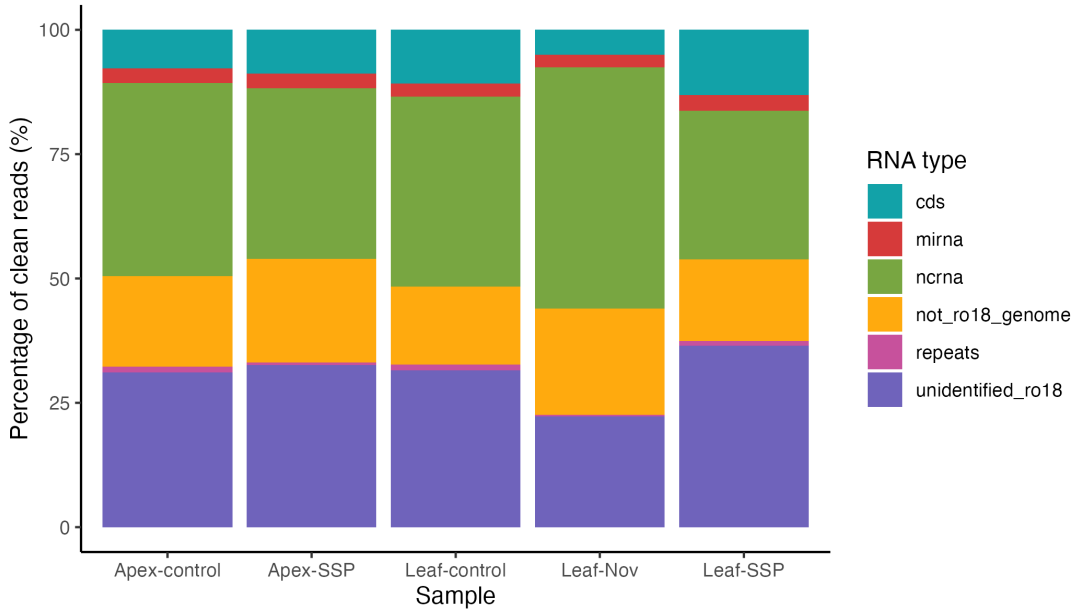


Figure 2.13: **Small RNA libraries had lower proportions of non-coding RNA after SSP treatment.** Each bar represents the average proportion of RNA types of R-o-18 apex and leaf samples before and after SSP treatment, with an additional comparison with Novogene's leaf sequencing data. RNA types include R-o-18 coding sequences (cds), miRNA, non-coding RNA (ncrna), unannotated R-o-18 aligned reads (unidentified_ro18), and a small percentage of reads that didn't align to the reference R-o-18 genome (not_ro18_genome).

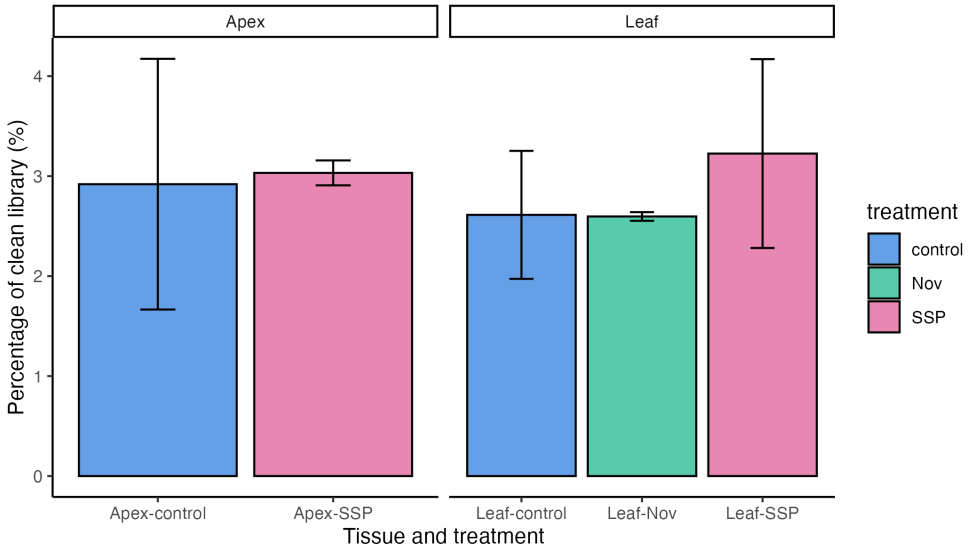


Figure 2.14: **SSP treatment slightly improved miRNA content, but not significantly.** Each bar represents the average percentage of clean reads which aligned to miRNA sequences, depending on tissue type (apex, leaf) and treatment (no SSP treated (control, blue), SSP treated (pink)), with an additional comparison with two Novogene sequenced leaf samples (en1l, dz1l, in green).

SSP treatment does not alter proportion of expressed miRNAs in each sample

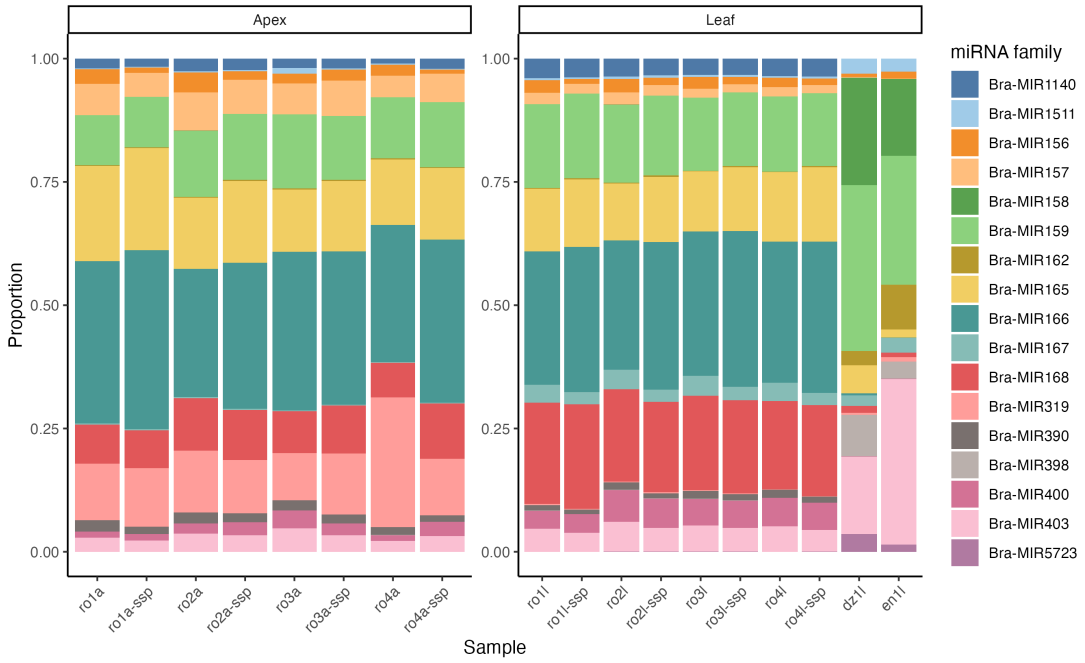


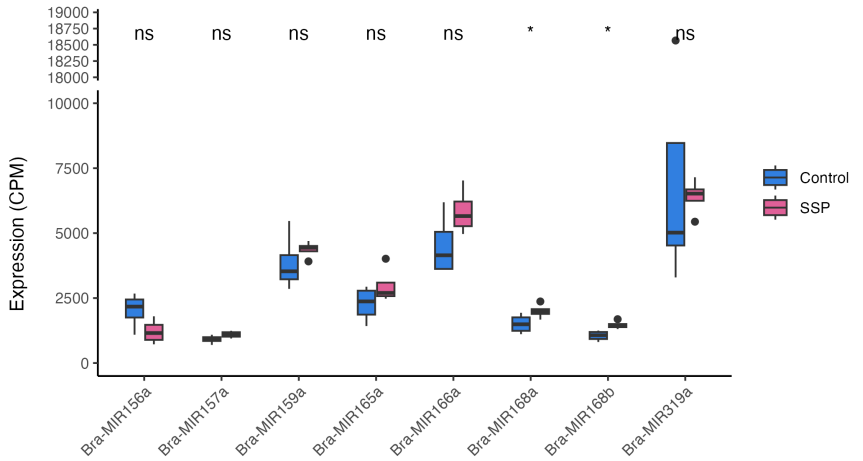
Figure 2.15: **SSP treatment does not skew the proportion of the top 10 expressed miRNAs within the sample.** A list of top 10 expressed miRNAs from each sample was combined, and each bar represents the proportion of those top expressed miRNAs per sample. Overall, SSP treated samples had consistent proportions of top 10 miRNA with its corresponding sample. Sample ro4a is slightly skewed due to extremely high expression of Bra-miR319a compared to other apex samples (see Fig.2.16 Bra-miR319a expression in apex tissue), but this is not reflected in ro4a-ssp.

To confirm that the SSP treatment did not alter the proportions of miRNAs expressed in the samples, the top 10 miRNA families expressed in each sample were selected and we compared the proportions of those miRNA families in all the samples.

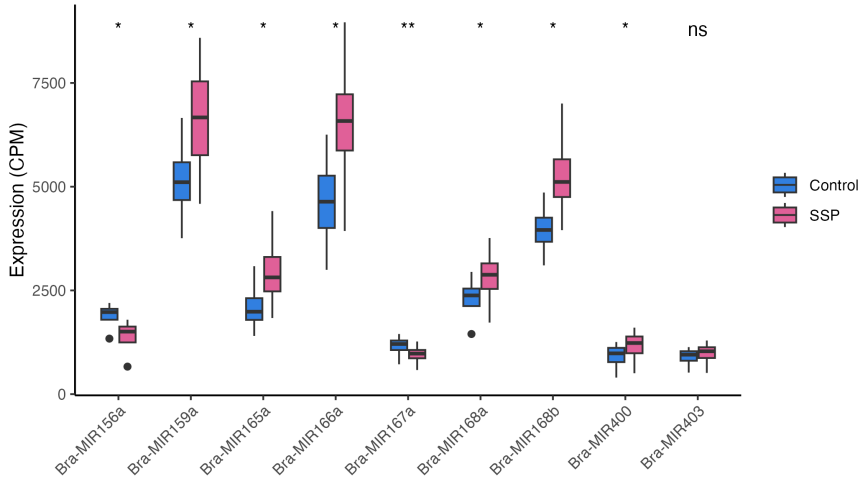
Not all samples had the same top 10 miRNAs, for example, the en1l and dz1l samples from Novogene had Bra-miR158, Bra-miR162, Bra-miR398, and Bra-miR5723 in their top 10 expressed miRNAs, which were not in the top 10 miRNAs for any of the RealSeq samples (Fig. 2.15). In the Realseq sequenced leaf samples (samples ro1l, ro2l, ro3l, ro4l and their SSP counterparts), there was a higher proportion of Bra-miR167 and Bra-miR400 expression compared to the apex samples, while in the apex samples, there was a higher proportion of Bra-miR157 and Bra-miR319. Only ro4a had the highest Bra-miR319 proportion compared to all the leaf and tissue samples, but its SSP treated sample had a similar proportion to the rest of the apex samples (Fig. 2.15).

Overall, there is no difference in the top 10 expressed miRNA families between SSP treated and untreated samples respective to leaf and apex tissue, which confirms that SSP treatment does not alter the proportion of highly expressed miRNAs. This means that the SSP treatment should not impact the miRNA makeup of the samples, thus

preserving the biological snapshot of miRNA expression and allowing for accurate biological interpretation.



(a) Highly expressed apex miRNAs (>1000cpm)



(b) Highly expressed leaf miRNAs (>1000cpm)

Figure 2.16: **SSP treatment improved detection of most highly expressed miRNAs above 1000 CPM (counts per million).** The plot shows expression levels of detected miRNAs in (a) apex and (b) leaf tissue, before and after SSP treatment. Note that the y-axis in plot (a) has a break between 10000-18000 CPM for ease of visualisation. SSP improvement of miRNA detection was more significantly pronounced in leaf miRNAs. In both apex and leaf tissue, Bra-miR168a/b show significantly higher expression in SSP treated samples, alongside Bra-miR159a, Bra-miR165a and Bra-miR166a which also increases but not significantly in the apex. Meanwhile, Bra-miR156a expression is negatively impacted by SSP treatment in apex and leaf tissue, and is significantly reduced in leaf tissue. The high outlier in Bra-miR319a apex is from sample ro4a. Paired t-test was performed between SSP vs Control treatments for each miRNA, and significance is annotated on the plot (*: $p \leqslant 0.05$; **: $p \leqslant 0.005$).

After we confirmed that the SSP treatment did not alter the proportion of detected miRNAs, we were interested in whether it could increase the detection of miRNAs. Although there was not much improvement in miRNA content (Fig. 2.14), we did

observe an increase in detection of highly expressed miRNAs (Fig. 2.16).

Bra-miR156a, Bra-miR159a, Bra-miR165a, Bra-miR166a and Bra-miR168ab were highly expressed in both apex and leaf tissues at over 2000CPM. In the apex, other miRNAs were also highly expressed (Bra-miR157a and Bra-miR319a), and SSP treated apex samples had slightly more detection of these miRNAs such as Bra-miR168ab and Bra-miR166a, although the difference was negligible except for Bra-miR168a and Bra-miR168b (Fig. 2.16a).

In the leaf, Bra-miR167a, Bra-miR400 and Bra-miR403 were also highly expressed, and SSP treated leaf samples increased detection of almost all of these highly expressed miRNAs, except for Bra-miR403 which showed no difference (Fig. 2.16b).

SSP treatment negatively impacts Bra-miR156 detection

Surprisingly, Bra-miR156a had significantly reduced detection in SSP treated leaf samples (Fig. 2.16b). This reduction was also seen in the apex samples (Fig. 2.16a). Expression analysis of other members of the Bra-miR156 family showed a reduction in Bra-miR156i, the second most highly expressed member of the Bra-miR156 family after Bra-miR156a (Fig. 2.17). Although the reduction in Bra-miR156 detection is not statistically significant, we didn't want to take any risks with altering the detection of a key miRNA in this thesis, thus we decided not to use SSP treatment for future sequencing.

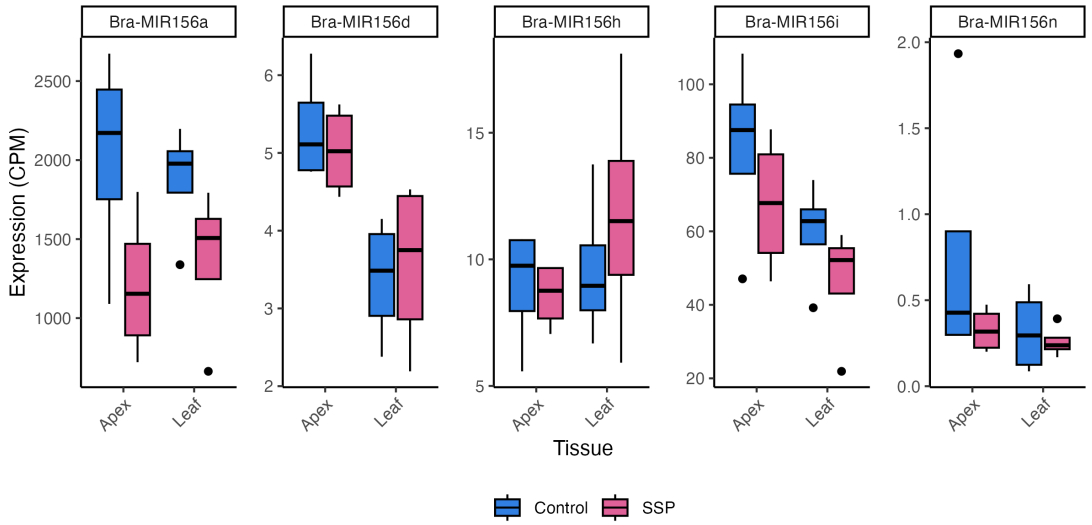


Figure 2.17: **SSP treatment negatively impacts detection of Bra-miR156 members, particularly Bra-miR156a and Bra-miR156i.** Expression of each Bra-miR156 member is represented with a barplot for each tissue and treatment. Wilcoxon tests revealed no statistically significant differences between control and SSP treatment for all Bra-miR156 individuals.

Conclusion

From our Tapestation analysis and bioinformatic analysis of miRNA content in DH1012 and R-o-18 apex and leaf samples, we decided to use the micropestle tissue grinding method due to its ability to extract more miRNA, then the Direct-zol kit for its straightforward protocol and ability to extract total RNA as short as 18nt, followed by small RNA-sequencing with RealSeq Biosciences without SSP treatment. RealSeq Biosciences provided a better price for sequencing >70 samples for the full miRNA timeseries in Sarisha-14 and R-o-18 (£240 per sample) compared to Novogene (£333 per sample), and their sequencing quality was consistent across all samples.

2.5 B. RAPA MIRNA TIMESERIES DATA DESCRIPTION

In this final section, I will provide an overview of the small RNA sequencing dataset generated from the timeseries sampling of Sarisha-14 and R-o-18 (sampling methods described in earlier Methods section).

Sarisha-14 and R-o-18 were grown in CER conditions, and shoot apex/leaf tissue were sampled in regular intervals throughout their development, from seedling stage until developmental stage BBCH51. For Sarisha-14, a total of 10 timepoints was sampled, while for R-o-18, a total of 9 timepoints was sampled (see Fig. 2.1 for more detail on sampling strategy). For each timepoint, 3 replicates per tissue were sent for sequencing. This means that in this timeseries, we have a total of 60 samples from Sarisha-14, and 54 samples from R-o-18, which adds up to 114 samples.

To assess miRNA detection in our timepoints, I sent apex and leaf samples from 3 timepoints per cultivar for small RNA-sequencing, 1 timepoint for the vegetative stage (Sarisha-14 day 7, R-o-18 day 9), another timepoint for the latest vegetative stage before the floral transition (Sarisha-14 day 11, R-o-18 day 17) and the last timepoint for post-floral transition (Sarisha-14 day 15, R-o-18 day 23). The preliminary results from this trial sequencing run were satisfactory, with successful detection of Bra-miR156 and Bra-miR172, so we continued with sequencing the rest of the timepoints. However, this meant that there was a sequencing batch effect that needs to be considered during data analysis, and I will later discuss how this is corrected while performing Principle Component Analysis in subsection 2.5.2.

2.5.1 *Sequencing data quality*

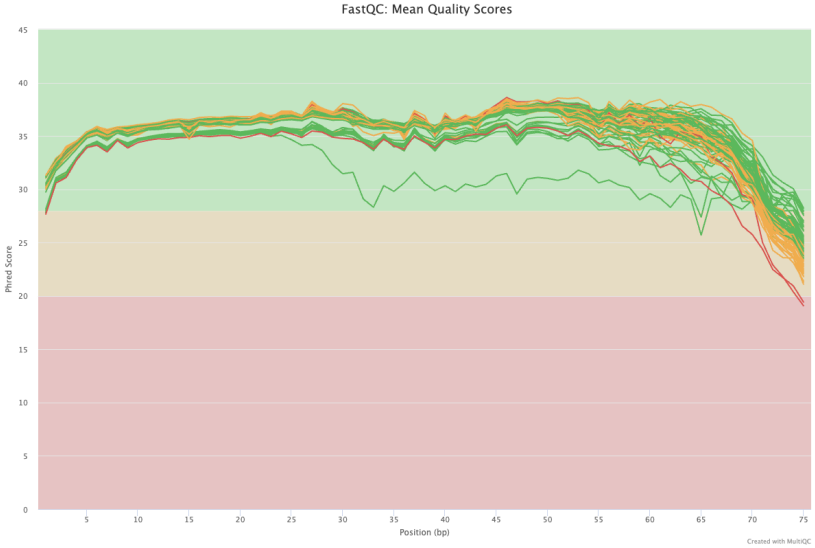
The raw small RNA-sequencing data was quality checked with FastQC (ver 0.11.9), then adapter trimmed with cutadapt (Martin 2011).

Per-base quality profiles of the adapter-trimmed reads showed that most reads under 30 nt long retained high-quality sequence quality, while reads longer than 30

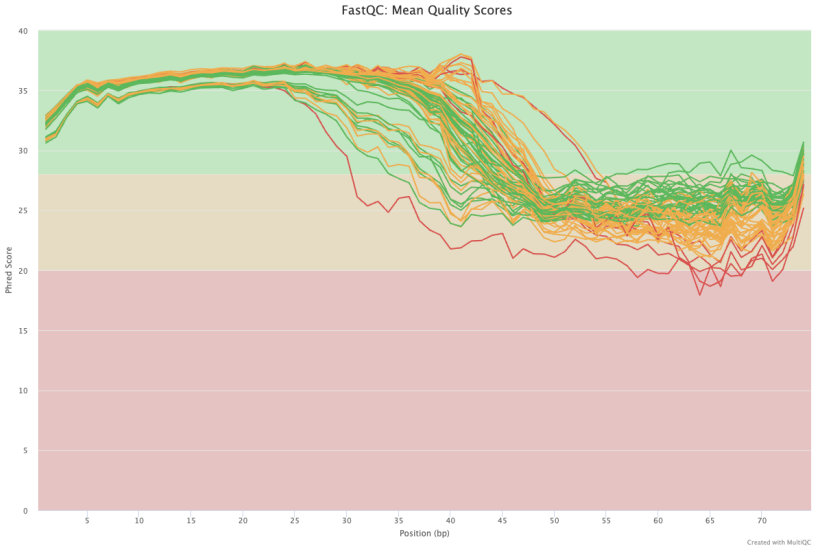
nt began to show a drop in sequence quality (Fig. 2.18). Since these reads are filtered out during subsequent size selection (retaining only reads between 18-28 nt), and they are longer than plant miRNAs (20-22 nt), this decrease does not impact downstream analysis.

Adapter content was also assessed before and after trimming (Fig. 2.19). The raw libraries contained high proportions of adapter sequences, which were successfully removed after adapter trimming, which completely removed all the adapter sequences (Fig. 2.19b).

Lastly, the read length distribution of the adapter-trimmed reads from all the samples revealed enrichment of 20 nt reads, and more abundantly, 24 nt reads (Fig. 2.20). These read lengths are consistent with plant miRNAs (20-22 nt), and small interfering RNAs (siRNAs, 24 nt). This confirms that miRNAs are detected in the samples, and hints that siRNAs may also be involved in development.

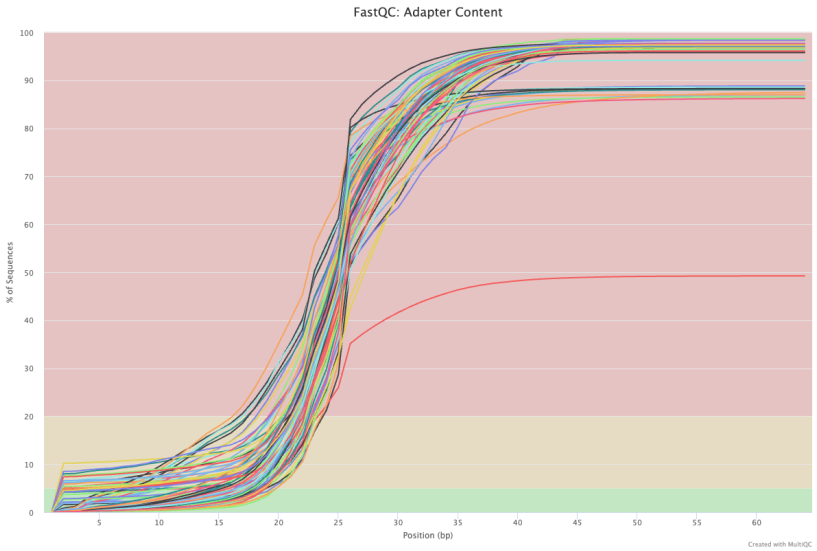


(a)

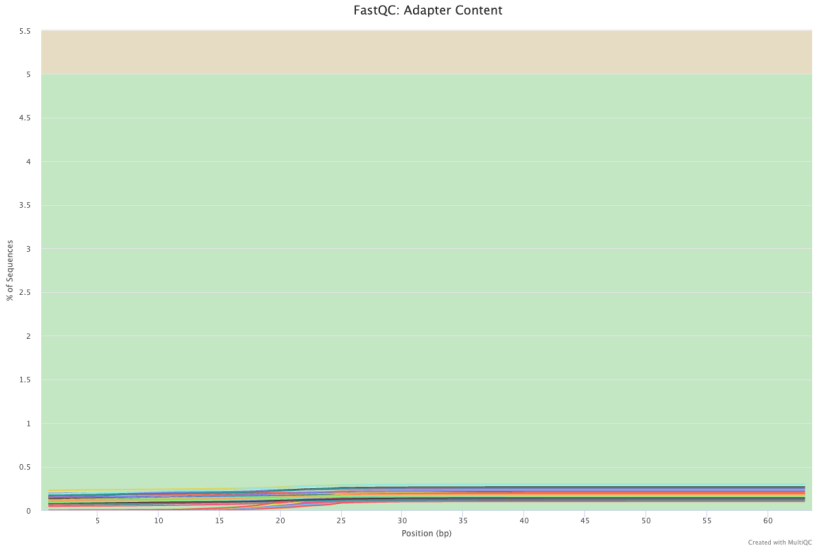


(b)

Figure 2.18: **Per base sequence quality drops in reads longer than 30 nt after adapter trimming.** The figures above show the mean sequence quality scores across the length of sequencing reads in the raw data (a) and after adapter trimming (b). Each line represents a single sample. Most adapter-trimmed reads under 30 nt long have high sequence quality, which includes miRNA reads (20-22 nt). Meanwhile, longer reads (> 30 nt) have lower quality bases near the end of the reads, but this is not concerning since they are filtered out during miRNA size selection. These figures were generated using MultiQC (Ewels et al. 2016).



(a)



(b)

Figure 2.19: **Adapter trimming step fully removes all adapter sequences.** The figures above show the percentage of sequences in each sample (represented in each line) which have adapters across the sequence lengths in the raw data (a) and after adapter trimming (b). Figure (b) shows that the adapter trimming step was successful in removing adapters from all the sequences. These figures were generated using MultiQC (Ewels et al. 2016).

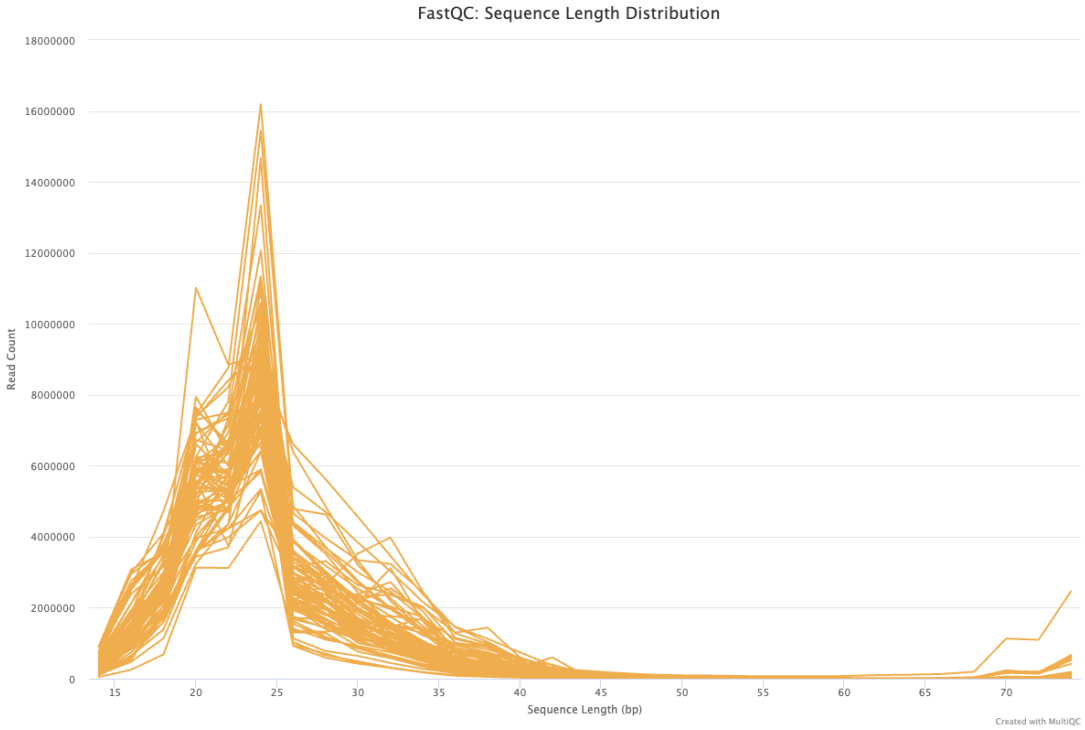


Figure 2.20: **20 nt and 24 nt small RNA reads are abundant in all samples.** The figure above shows the distribution of read lengths in each sample (represented with lines) after adapter trimming, but before size selection. A distinct peak is seen at the 24 nt sequence length, followed by 20 nt sequence length, which shows that these are the two most abundant small RNA read lengths in all the samples. Plant miRNAs are usually between 20-22 nt, while small interfering RNAs (siRNAs) are 24 nt long, which suggests that siRNAs are more abundant than miRNAs in these samples. This figure was generated using MultiQC (Ewels et al. 2016)

2.5.2 *Overview of identified small RNA*

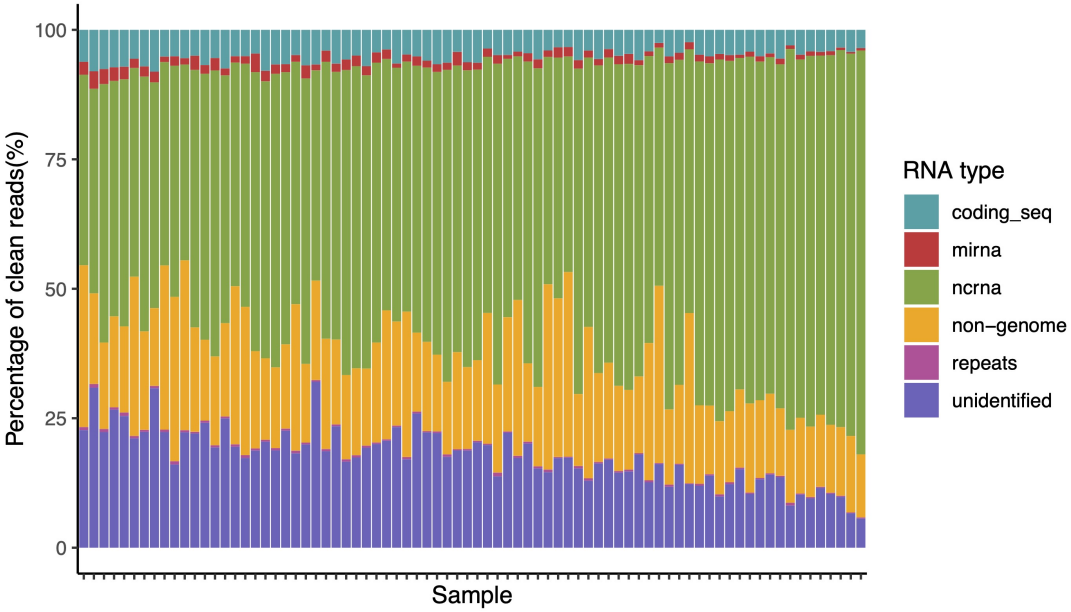


Figure 2.21: **Summary of small RNAs identified in the timeseries dataset.** Each bar represents the average proportion of RNA types identified in each sample. The miRNA identification pipeline aligns the adapter-trimmed and size-selected small RNA reads (clean reads) against the reference R-o-18 genome, non-coding RNA sequences (ncrna), repeat sequences (repeats), miRNA sequences (mirna), and reference R-o-18 coding sequences (coding_seq). Small RNA reads that did not align to the reference genome are labelled as "non-genome", while other genome-aligned small RNA reads that did not align to any known miRNA/ncRNA/repeat/-coding sequences were labelled as "unidentified". On average, non-coding RNA make up the majority of the small RNA libraries (~50-75%), followed by unidentified *B. rapa* reads (~15-30%) and non-genome reads (~10-25%). Around 5-8% of the small RNA libraries aligned to R-o-18 coding sequences, while miRNAs make up around 1-3% of reads. Less than 1% of all the samples aligned to known repeat sequences.

Following the miRNA identification pipeline in Section2.2.5, different groups of small RNAs were identified (Fig. 2.21). Clean small RNA reads (adapter-trimmed and size-filtered for 18-28 nt reads) were first aligned against the R-o-18 reference genome. Around 80 % of the reads successfully aligned against the reference genome, leaving roughly 20 % of reads that are not *B. rapa* derived. This is likely from environmentally derived small RNA which contaminated the tissues sampled. After that, non-coding RNAs (rRNAs, tRNAs, snRNAs and snoRNAs) made up the majority of genome-aligned reads, around 50–75% of the libraries. In contrast, only a small percentage (< 1%) of reads were derived from repeat sequences, followed by small RNAs from degraded coding mRNA transcripts (5–8% of mapped reads). *B. rapa* canonical miRNAs make up around 1–3% of the small RNA reads, leaving around 15-30% of the libraries as unannotated *B. rapa* small RNAs. These unidentified small RNAs could potentially be isoforms of canonical miRNAs which were filtered out due to strict alignment parameters (0 nucleotide mismatch), but it is also possible that they could be small

interfering RNA sequences that are yet to be annotated in *B. rapa*. This distribution of small RNAs shows that the libraries captured a diverse set of small RNAs, with structural non-coding RNAs making up the majority of the libraries, and miRNAs consistently being in low abundance across all samples.

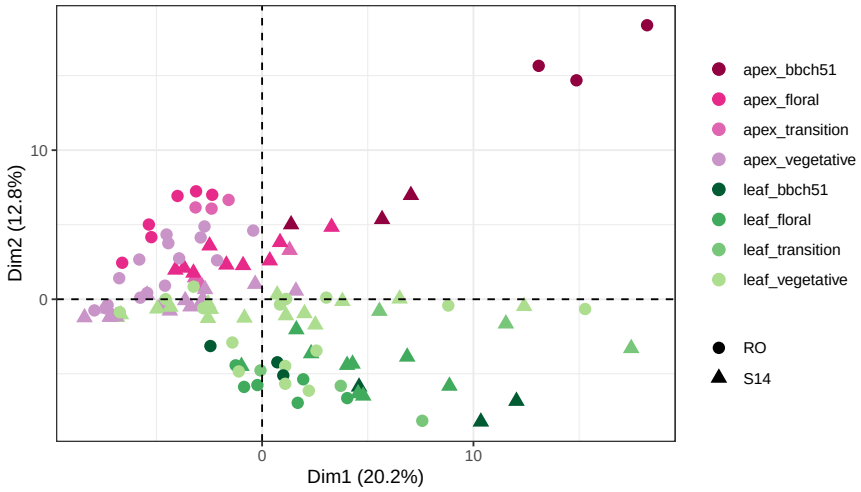
Data standardization for clustering analysis

To investigate the robustness of the miRNA expression profiles in the timeseries, I performed principal component analysis (PCA) on the dataset. PCA is an unsupervised machine-learning technique which clusters samples that have similar miRNA expression profiles together, and achieves this by reducing high-dimensional data (in this case, read counts of all the miRNAs in individual samples) into a few principal components that captures the main sources of variations. Each point in a PCA plot represents a sample, and ideally samples from the same genotype, tissue type and developmental stage should cluster together if their miRNA expression profiles are similar.

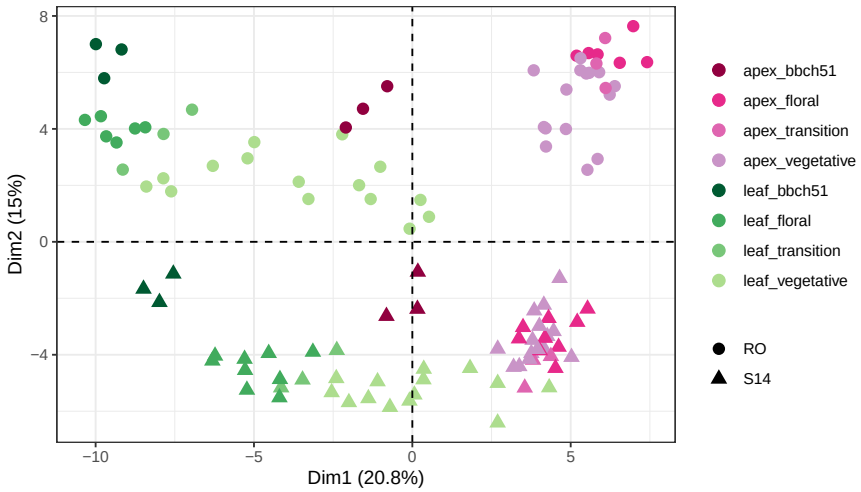
When PCA was performed on the whole dataset with all 114 samples using library-size normalized counts (counts per million, CPM), the samples grouped together by tissue type, but there was no clear separation in samples from different genotypes or developmental stages (Fig. 2.22a). However, this is due to two factors: 1) the normalization method against library size, that doesn't standardize variation in miRNA proportions per library, and 2) as alluded earlier, there is a batch effect due to two separate sequencing runs. Thus, variance stabilizing transformation (VST) (from the DESeq2 R package) was applied to raw read counts using DESeq2 to rescale the data, reducing noise in lowly expressed miRNAs and making variance comparable across expression levels for reliable PCA analysis. Batch effects from separate sequencing runs were then removed from the VST data using the R package limma's removeBatchEffect function, preserving biological variation. This revealed tighter clustering of the samples (Fig. 2.22b), with samples from similar genotypes and tissue types clustered closer together. Here, samples separated primarily by genotype: R-o-18 samples had positive Dim2 values, while Sarisha-14 samples clustered at the opposite side of the plot (negative Dim2). Within each genotype, leaf samples showed a strong trajectory corresponding to developmental stage progression, moving across the plot from early to late stages. In contrast, apex samples did not show the same stage-resolved progression, and different developmental stages overlapped each other, except for BBCH51 stage apex samples which clustered separately from the apex samples. These results show that there is a clear separation in the samples between genotypes and tissue types, confirming robust sampling methods.

To better assess the apex sampling timeseries, PCA was performed separately on all the apex samples, excluding BBCH51 stage apices, because it was more developmentally advanced compared to the other vegetative and early floral timepoints (Fig.

2.23). Similar to the above analysis, PCA with only library-normalized counts showed samples clustering by genotype, but no clear progression in the timepoints (Fig. 2.23a). In contrast, PCA with the batch corrected and variance stabilised data revealed a clear pattern of developmental progression (Fig. 2.23b). R-o-18 samples progressed consistently from early to late timepoints along Dim2, while Sarisha-14 floral samples (days 12–15) overlapped, indicating less separation between these timepoints in this cultivar, possibly because they are very similar in developmental stage. Overall, this analysis confirms that the apex samples were collected at biologically similar developmental stages.

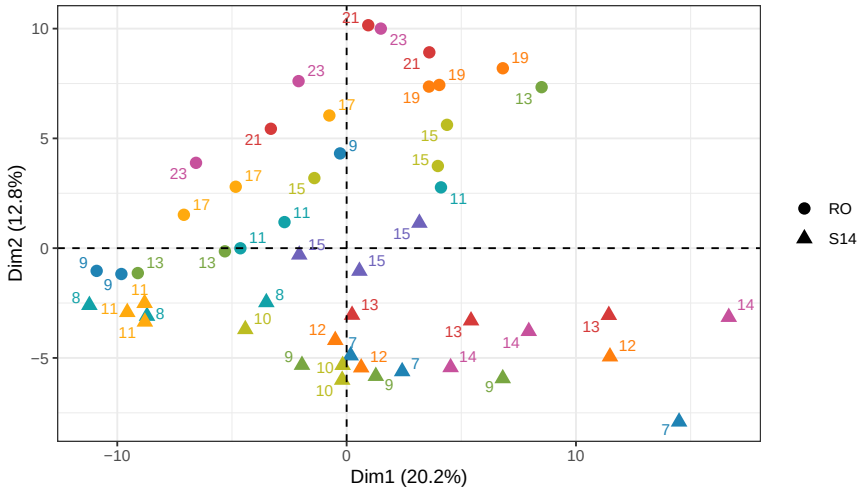


(a) Original data

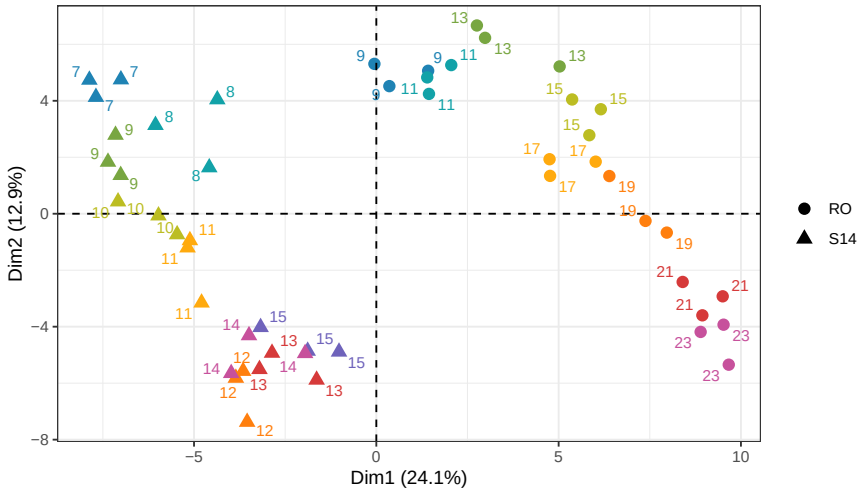


(b) Batch corrected and variance stabilised data

Figure 2.22: **Principle Component Analysis using miRNA read counts from all 114 leaf (green) and apex (pink) samples from R-o-18 (circular points) and Sarisha-14 (triangular points).** PCA was performed to assess tissue/stage/cultivar specific clustering in the samples. Each point represents a single replicate coloured by tissue and stage, and genotype (shape). Each developmental stage has a different colour shade depending on the tissue. Initially, library-size normalized counts (counts per million, cpm) were used for PCA (a), which showed some overlap between leaf and apex groups, but no clear genotype or stage-specific clustering was observed. After applying variance stabilizing transformation (VST) and batch correction to miRNA read counts, the PCA was repeated on the corrected samples (b). Here, the samples cluster based on genotype, with R-o-18 samples at the upper half of the plot (positive Dim2 values) while Sarisha-14 samples occupy the lower half of the plot (negative Dim2 values). Samples in the same developmental stage and tissue type also cluster together, and developmental stage progression is observed in the leaf tissue (developmental stage progresses further from the right to the left of the plot). However, the apex samples do not show a similar progression, which could be due to leaf samples contributing more to Dim2 differences. These plots show that VST and batch correction improve clustering results.



(a) Original data



(b) Batch corrected and variance stabilised data

Figure 2.23: **Principle Component Analysis of apex samples from R-o-18 (circular points) and Sarisha-14 (triangular points) using miRNA read counts.** To assess whether the apex samples showed a similar developmental progression like the leaf timepoints in Fig. 2.22, a separate PCA was performed on the apex samples only (except for samples from the BBCH51 stage, which skewed the data due to being more developmentally advanced than the earlier vegetative/floral timepoints). Each point represents a single replicate, with its timepoint labelled and coloured. Similar to Fig. 2.22, plot (a) is PCA performed with library-size normalized counts, which shows that the samples grouped by genotype, but there was no clear clustering by timepoint. In plot (b), the miRNA read counts are variance stabilized and batch corrected, then PCA was repeated. The apex samples now show clear progression throughout the timeseries, with the youngest samples at the top of the plot (high positive Dim2 value), and the oldest samples at the bottom of the plot (low negative Dim2 value). The progression is more defined in the R-o-18 samples on the right side of the plot, while the Sarisha-14 floral samples (days 12-15) overlapped with each other and no longer separated by timepoints. This confirms that the apex sampling timeseries correctly represents the developmental stages that the plants progress through.

2.6 DISCUSSION

Although the primary objective of this thesis was to characterise miRNA abundance in *B. rapa* cultivars, the initial optimisation of small RNA extraction and sequencing workflow was conducted using the DH1012 rapid-cycling cultivar (*B. oleracea*) as a technical test subject. Once the protocol was streamlined, all subsequent small RNA analysis was focused on the *B. rapa* cultivars R-o-18 and Sarisha-14. And with this, I have generated an extensive small RNA-sequencing dataset which profiles small RNA expressed in apex and first true leaf tissue throughout *B. rapa* floral development. We aim to upload this dataset to the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra?db=sra) or European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) by the end of 2026.

2.6.1 *Challenges in Brassica miRNA identification*

While RNA extraction and refinement of the small RNA sequencing strategy took time, the real challenge was in identifying Brassica miRNAs from small RNA-seq data.

An initial difficulty was the low proportion of reads aligning to annotated miRNAs, with less than 1% of the total clean library corresponding to miRNAs (Table 2.4). In comparison, previously published Brassica studies reported substantially higher proportions of miRNA reads between 6-33% of the clean reads (Ahmed et al. 2020; Kong et al. 2022), which suggested that we could increase our miRNA detection during sequencing. In addition, early Arabidopsis studies similarly reported 10–29% of small RNA reads corresponding to miRNAs in wild-type backgrounds (C. Lu et al. 2006; Rajagopalan et al. 2006). However, closer examination revealed that these studies calculated miRNA percentages only after filtering structural non-coding RNAs, rather than reporting relative to total clean reads (C. Lu et al. 2006; Rajagopalan et al. 2006). This reporting strategy inflates the apparent miRNA proportion and complicates direct comparison across datasets. Furthermore, some Brassica studies aligned their small RNA data to all known plant miRNAs in miRBase rather than restricting analysis to species-specific annotations, which may also increase apparent miRNA detection rates. Thus, these differences in reporting likely contributed to our initial expectation of higher miRNA representation, although in hindsight, our results were actually common in plant small RNA sequencing.

Another limitation as hinted earlier, was the available reference sequences from current miRNA databases. The most widely used, miRBase (https://www.mirbase.org/), has limited annotations for Brassica miRNAs - there are only 10 miRNA entries for *B. oleracea* and 96 entries for *B. rapa*, compared to 326 annotated miRNA sequences in *Arabidopsis thaliana*. This is likely because miRBase relies on user submissions to expand its database.

Thus, due to limited Brassica miRNA resources, published Brassica studies often use different identification strategies. For instance, some rely purely on homology-based searches such as BLASTn against miRNAs from other species (Jiang et al. 2021), while others combine sequence homology with RNA secondary structure prediction to identify precursor stem-loop structures (Hui Li et al. 2017). More recent studies have integrated both approaches: mapping reads to all plant miRNAs and validating predictions with structural analysis (Kong et al. 2022). A wide range of computational tools also exists, as reviewed by Q. Li et al. (2021), from command-line pipelines like miRDeep-P2 (which can predict novel loci as well as quantify known miRNAs (Kuang et al. 2019)) to web-based tools such as sRNAbench (which primarily uses miRBase for miRNA identification (Aparicio-Puerta et al. 2022)).

For a dataset of this scale (114 samples), pipeline parallelisation and automation was essential. I therefore based my approach on the detailed protocol by Garg and Varshney (2022), which has the same strategy as the other Brassica studies of filtering out non-coding RNAs, and then aligning the rest of the sRNA reads to miRNA sequences. I adapted the protocol for my data, such as using cutadapt for adapter trimming and length selection, changing bowtie parameters to ensure perfect alignments to canonical miRNA sequences, and replaced miRBase with a more comprehensive database, PmiREN (Plant microRNA Encyclopedia, Guo, Kuang, and X. Yang (2025)). I chose PmiREN because it integrates published small RNA datasets and miRDeep-P2 predictions, effectively outsourcing the prediction step and expanding the reference miRNA sequences to 206 known and 84 novel *B. rapa* miRNAs (compared to 96 in miRBase).

However, this pipeline is limited by the list of reference miRNA sequences. For example, in DH1012 samples the apparent low proportion of miRNAs likely reflects the restricted *B. oleracea* reference miRNA list, rather than actual biological absence. The same caveat applies to *B. rapa*, so although we did detect miRNA expression, a significant proportion (~20%) of the small RNA libraries remain "unidentified". This pipeline can be tested using publically available data and the miRNA abundance could be compared with published results. Unfortunately, initial attempts to test with Brassica datasets were unsuccessful due to different sequencing strategies (paired-end instead of single-end sequencing, Kong et al. (2022)) or unreported miRNA abundance (Jiang et al. 2021). In retrospect, this pipeline could be tested with existing Arabidopsis small RNA-seq datasets such as German et al. (2008) or Jeong et al. (2013).

To identify more Brassica miRNAs, de novo miRNA prediction (e.g. miRDeep-P2) can be used to predict precursor miRNA loci using the unannotated sRNA reads against the Brassica reference genome. Other miRNA databases such as sRNAanno (http://plantsrnas.org/index.html), which not only provides precursor coordinates, but also includes sequences of phased siRNAs (phasiRNAs), could extend the analysis further. Given the high proportion of 24 nt small RNAs in this dataset, exploring

phasiRNA expression during floral development could provide additional insights into small RNA–mediated regulation beyond miRNAs, however this is currently beyond the scope of this thesis.

In conclusion, I established an optimised protocol for RNA extraction from Brassica shoot apices, and provide an adaptable pipeline for miRNA identification. This work generated an extensive small RNA time series dataset from the vegetative to floral stages in rapid-cycling *B. rapa* cultivars, R-o-18 and Sarisha-14. Beyond miRNAs, FastQC read length distributions reveal an abundance of 24 nt potential siRNAs, highlighting the broader regulatory potential of small RNAs during floral development. Finally, PCA confirms that the samples are representative of their respective tissue, genotype and developmental stages, making it a robust dataset for analysing miRNA expression dynamics during *B. rapa* shoot apex and leaf development. We plan to upload these datasets to the NCBI Sequence Read Archive in the future.

3

IDENTIFICATION OF CANDIDATE MICRORNAS INVOLVED IN FLOWERING

3.1 INTRODUCTION

In Chapter 2, a small RNA time series dataset was generated for rapid-cycling *B. rapa* cultivars R-o-18 and Sarisha-14, providing a high-resolution view of miRNA abundance dynamics across vegetative and reproductive development. This dataset enables a systematic investigation of how temporal regulation of miRNAs contributes to differences in flowering time between cultivars. Notably, Sarisha-14 undergoes the floral transition substantially earlier than R-o-18, suggesting that shifts in regulatory timing may underlie its rapid-cycling phenotype.

The floral transition represents a major developmental phase change requiring coordinated reprogramming of gene expression. While transcription factors have traditionally been emphasised in flowering regulation, miRNAs act as key post-transcriptional regulators capable of fine-tuning gene expression networks (Dong, B. Hu, and C. Zhang 2022). Across plant species, numerous miRNAs have been implicated in flowering control (Teotia and Tang 2015; Waheed and Zeng 2020). These include miRNAs involved in hormone signalling pathways such as auxin and gibberellin signalling, floral organ identity, and abiotic stress responses, forming an interconnected regulatory network that integrates developmental and environmental cues (Fig. 3.1).

Given the important role of miRNAs in regulating developmental timing, differences in miRNA abundance dynamics between Sarisha-14 and R-o-18 may contribute to their contrasting flowering phenotypes. Earlier or altered temporal shifts in miRNA abundance could accelerate the activation of floral promoters or relieve repression of flowering pathways. Conversely, cultivar-specific miRNA expression patterns may reflect distinct regulatory strategies underlying Sarisha-14's earlier flowering phenotype.

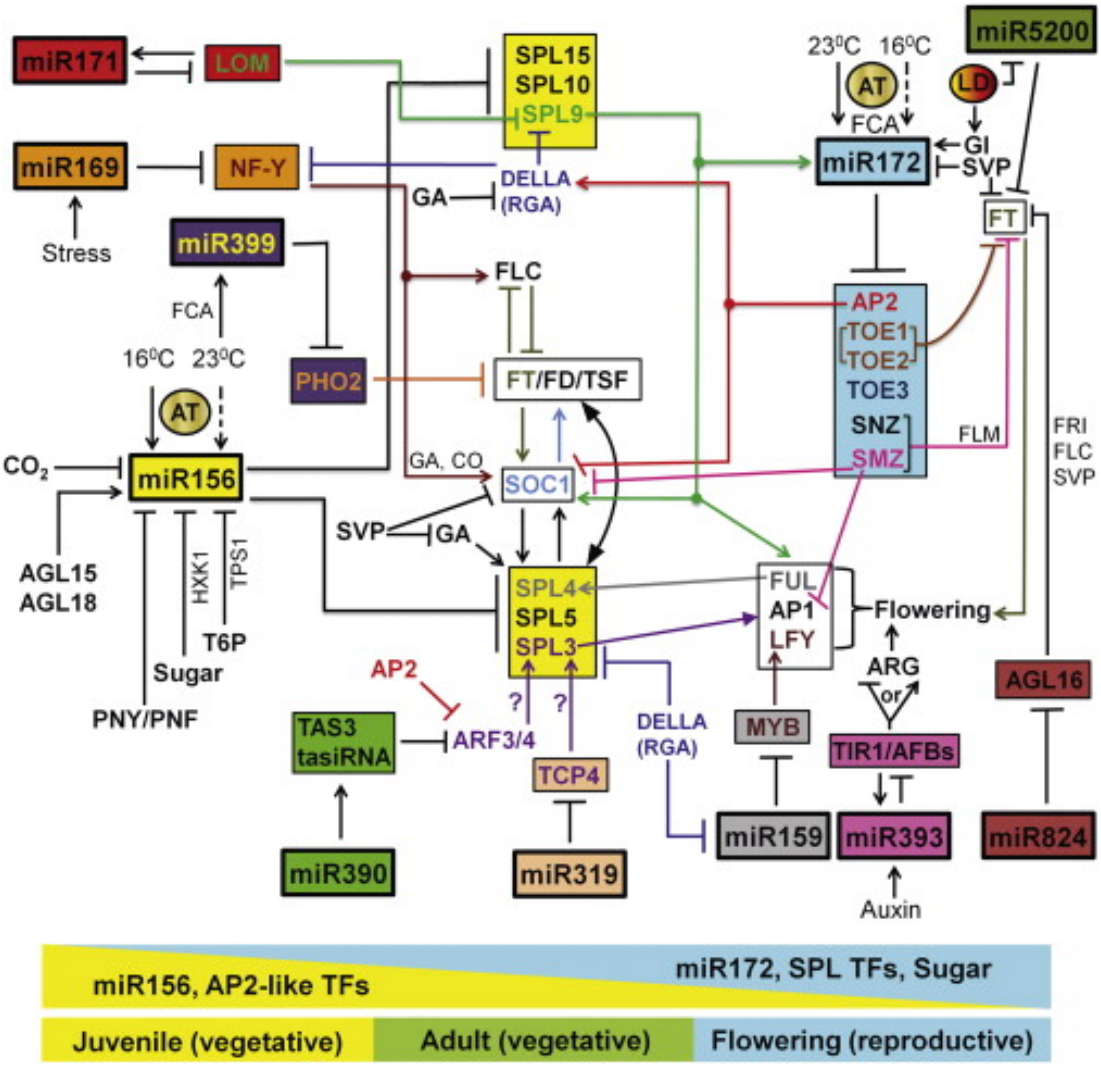


Figure 3.1: **miRNA-mediated regulatory networks controlling flowering time.** Overview of microRNA pathways involved in the regulation of the floral transition. While the miR156–SPL and miR172–AP2 modules form the central age-dependent regulatory axis, multiple additional miRNA modules (e.g., miR159, miR169, miR319, miR390, miR393, miR399, miR824, and miR5200) also contribute to flowering time control through interactions with key integrators such as *FT*, *SOC1*, and *FLC*. These pathways integrate developmental, environmental, hormonal, and stress-related signals. Arrows indicate activation; bars indicate repression. This figure was taken from a review by Teotia and Tang (2015).

Thus, this chapter is an exploratory analysis of other miRNAs besides miR156/miR172, that also influence flowering time in rapid-cycling *B. rapa*. The main goal is to procure a list of candidate flowering miRNAs which contribute to earlier flowering time in Sarisha-14, either by an earlier temporal shift in abundance, or by different abundance dynamics or overall abundance abundance. Identifying other miRNAs beyond miR156 and miR172 would provide new targets for plant breeding not only for regulating flowering time, but also stress resilience and growth rate.

In this chapter, I applied a combination of clustering, curve registration, and differential expression analyses to identify candidate flowering-associated miRNAs. Fuzzy c-means clustering (Matthias E. Futschik and Carlisle 2005) was used to group miRNAs by shared temporal abundance patterns in each cultivar, allowing the identification of potentially conserved flowering regulators based off their abundance patterns. Curve registration (2024) aligned the abundance profiles of these miRNAs between Sarisha-14 and R-o-18, assessing whether Sarisha-14's earlier flowering is due to an earlier shift in abundance, or unique abundance patterns not shared between the cultivars. Spline-based differential expression analysis (2018) also identified miRNAs with genotype-specific abundance trajectories. Lastly, I predicted the target genes of candidate flowering miRNAs, and used current knowledge of conserved miRNAs, or reciprocal miRNA-target gene expression dynamics (due to miRNA-mediated repression of target genes) to present a selection of candidate miRNAs which are both conserved and cultivar-specific that contribute to Sarisha-14's earlier floral transition. This systems-level analysis establishes a foundation for subsequent focused investigation of specific regulatory modules.

3.1.1 *Hypotheses*

There are two hypotheses for this chapter:

1. **Hypothesis 1**: Some miRNAs will display conserved temporal abundance patterns similar to miR156 and miR172, with Sarisha-14 showing earlier shifts in dynamics consistent with its earlier floral transition.

2. **Hypothesis 2**: Other miRNAs will exhibit cultivar-specific abundance patterns or abundances, potentially contributing to Sarisha-14's accelerated flowering.

3.1.2 *Objectives*

The objectives of this chapter are to:

1. Identify conserved flowering-associated miRNAs by clustering temporal abundance profiles in Sarisha-14 and R-o-18 and comparing them with canonical flowering regulators (miR156, miR172).

2. Evaluate temporal differences in abundance dynamics between Sarisha-14 and R-o-18 using curve registration and spline-based DE analysis, testing whether Sarisha-14 shows earlier or altered trajectories.

3. Characterize cultivar-specific miRNAs and their potential targets, highlighting candidates that may underlie Sarisha-14's early flowering phenotype.

3.2 METHODS

3.2.1 *Plant growth conditions, sampling, and small RNA sequencing*

Brassica rapa cultivars Sarisha-14 and R-o-18 plants were sown in 24 cell trays in cereals mix (40% medium grade peat, 40% sterilised soil, 20% horticultural grit, 1.3 kg/m^3 PG mix 14-16-18 + Te base fertiliser, 1 kg/m^3 Osmocote Mini 16-8-11 2 mg + Te 0.02% B, wetting agent, 3 kg/m^3 maglime and 300 g/m^3 Exemptor). Material was grown in a Conviron MTPS 144 controlled environment room with Valoya NS1 LED lighting (250 µmol m^{-2} s^{-1}) 18 °C day/15 °C night, 70% relative humidity with a 16 hour day (lights on from 00.30hrs until 16.30hrs). Sampling of Sarisha-14 and R-o-18 leaf and apex was performed in the 10th hour of the day (ZT10). Shoot apex and first true leaf samples were taken throughout development from the vegetative stage, the floral transition, post-floral transition stage and until floral buds were visible (developmental stage BBCH51 (Meier et al. 2009). A total of 9 timepoints were collected from R-o-18, while 10 timepoints were collected for Sarisha-14. For each timepoint, 4 replicates were sampled, with 4 pooled apices/leaf tissue per replicate. Only 3 out of 4 replicates were processed for sequencing.

To identify the timepoint when the floral transition occured, 4 plants were inspected at the same time of day as the sampling, by dissection under a light microscope. The floral transition timepoint was marked when at least 3 out of 4 plants showed domed meristems with round floral primordia on the flanks of the apex (Kinoshita et al. 2020), compared to flat vegetative meristems flanked by triangular leaf primordia.

During tissue sampling at hour 10 of the day, apex and leaf tissue were dissected and pooled on ice inside the growth chamber, then flash frozen in liquid nitrogen once all sampled were pooled for each replicate. These samples were then stored in -70 °C storage, until total RNA extraction was performed at a later date.

Apex tissue samples were ground in the original microtube using a micropestle attachment on a drill while immersed in liquid nitrogen until a fine powder was formed. Leaf tissue was ground in a pestle and mortar while liquid nitrogen was added to keep the tissue frozen. Total RNA extraction was performed following the manufacturers standard protocol from Zymo Research's Direct-zol Microprep Total RNA extraction kit, supplemented with TRI Reagent for tissue lysis and DNAse treatment (https://zymoresearch.eu/products/direct-zol-rna-microprep-kits).

The total RNA samples were sent for sequencing with RealSeq Biosciences, California, United States of America. RNA libraries were constructed using Singular Genomics G4 F3 Sequencing Kit and sequencing was performed on the Singular Genomics G4 sequencer, at an average depth of 30 million single-ended reads.

3.2.2 *MiRNA identification and abundance quantification*

Raw small RNA sequencing reads (75 bp single-end) were quality-checked with FastQC (ver. 0.11.9, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapters trimmed using Cutadapt (ver. 2.10, Martin (2011)). After size filtering for reads between 15-30 nt, reads were aligned to the *Brassica rapa* R-o-18 reference genome (EnsemblPlants release 59) with Bowtie (ver. 1.2.2,Langmead et al. (2009)). Reads mapping perfectly (no mismatch) to ribosomal, transfer, small nuclear, or small nucleolar RNAs (sequences downloaded from Rfam release 14.10, Kalvari et al. (2021)) and to transposable elements (TrepDB release 19) were discarded. The remaining reads were aligned to a deduplicated set of mature *B. rapa* miRNA sequences from the Plant microRNA Encyclopedia (PmiREN, release 2.0, Guo, Kuang, and X. Yang (2025)) to identify known miRNAs, only keeping miRNAs perfectly mapping to canonical *B. rapa* miRNA sequences with no mismatch. Raw miRNA read counts for each sample were extracted using Samtools-idxstats (ver. 1.10, Danecek et al. (2021)). Read count normalization or transformation was performed depending on the requirements of the analysis, and are described below. For miRNA abundance abundance plots, library size normalised counts (CPM) were scaled with Trimmed Mean of M factors (TMM) to allow more accurate sample-to-sample gene expression abundance comparisons (Smid et al. 2018).

3.2.3 *Fuzzy c-means clustering for grouping miRNAs with similar abundance patterns*

To identify groups of miRNAs with similar temporal abundance profiles in Sarisha-14 and R-o-18 respectively, fuzzy c-means clustering was performed using the Mfuzz R package (version 2.68.0, Kumar and Matthias E Futschik (2007)).

First, raw count data were filtered to remove very low-abundance features (keeping only miRNAs with a total count $\geqslant 10$ reads across samples). Counts were then variance-stabilized using the DESeq2 function varianceStabilizingTransformation (DESeq2 version 1.48.2, Love, Huber, and Anders (2014)), with transformation performed in a blind manner to avoid model-dependent bias. This variance stabilising transformation reduces the problem where lowly expressed miRNAs appear much noisier than highly expressed ones, by re-scaling the data so that variance is more comparable across the full abundance range. This allows for more reliable comparison of temporal abundance patterns between miRNAs, independent of their absolute abundance, and makes the data suitable for downstream exploratory pattern analyses like principle component analysis (PCA) or fuzzy c-means clustering where the shape of the abundance profile is more important than raw counts.

To account for sequencing batch variation in the data, batch effects were removed from the variance-stabilized data using limma's removeBatchEffect (limma version 3.64.3, Ritchie et al. (2015)), with timepoint included in the design matrix. After batch

correction, abundance values were averaged across biological replicates for each timepoint, producing a matrix of mean abundance values per miRNA across the developmental series, one for each genotype.

This processed abundance matrix was standardized by Mfuzz, and the fuzzification parameter m was estimated using the mestimate function. The optimal number of clusters was evaluated using the Dmin() function, which calculates minimum centroid distances across repeated fuzzy c-means runs (2-25 clusters, 5 repeated runs). revealing elbows at c=8-10 for R-o-18 (Dmin plateau 0.81-0.86) and c=5-7 for Sarisha-14 (Dmin plateau 0.81-0.96). Fuzzy c-means clustering was then performed with the mfuzz function. Clustering was performed across a range of cluster numbers (c = 3-8) to visualise shared cluster patterns between the genotypes. Lower cluster numbers failed to capture the diversity of miRNA temporal abundance patterns, while higher cluster numbers increased the resolution of within genotype abundance patterns, however, the increased cluster patterns were not directly comparable between genotypes. Thus c=5 was selected to enable direct cross-cultivar trajectory comparisons while capturing major developmental profiles. The clustering results were visualised using the mfuzz package's mfuzz.plot2() function. Only the timepoints from vegetative stage until post-floral transition were used in this clustering, and the final sampled timepoint in both timeseries (BBCH51) was excluded to avoid mis-clustering due to abundance changes at the final timepoint.

3.2.4 *Curve registration of miRNA abundance profiles between R-o-18 and Sarisha-14*

Curve registration was used to directly compare temporal abundance dynamics between Sarisha-14 and R-o-18 with the greatR R package (version 2.0.0.9, loaded on R version 4.1.0, Kristianingsih (2024)). Library size normalised (counts per million, CPM) values from apex tissue were provided as input, with Sarisha-14 set as the query acession and R-o-18 as the reference accession. Abundance profiles were standardised using z-score scaling to allow shape-based comparisons independent of absolute abundance. Several optimization methods (Nelder-Mead, L-BFGS-B, simulated annealing) were tested for curve registration. Nelder-Mead optimisation was chosen because it provided the most consistent and biologically plausible alignments across a relatively small number of timepoints ($\leqslant 10$), whereas gradient-based or stochastic approaches were less stable for this dataset. I also set the parameter for at least 70% overlap required to consider aligned miRNA abundance profiles as "registered".

3.2.5 *Timeseries differential expression analysis with DESeq2*

To test for differential expression (miRNA abundance) in the miRNA timeseries between Sarisha-14 and R-o-18, I used the DESeq2 R package (version 1.48.2, Love, Huber,

and Anders (2014)). Due to the nature of our sampling timeseries, where Sarisha-14's faster progression meant that developmental stages such as the floral transition were not chronologically similar with R-o-18's, and an unequal number of sampling timepoints (total of 10 timepoints for Sarisha-14, and 9 timepoints for R-o-18), I chose not to do pairwise timepoint-to-timepoint comparisons between the two genotypes.

Instead, I fitted smooth expression trajectories across chronological time using natural cubic splines with 3 degrees of freedom (ns() function in the base R package splines (version 4.5.0)). This is a simpler implementation similar to what is described in Fischer, Theis, and Yosef (2018) and Spies et al. (2017). The code to build a DESeqDatasetFromMatrix (dds) object to model genotype x time interaction between the two miRNA abundance timeseries was:

```
dds_interaction <- DESeqDataSetFromMatrix(
    countData = counts,
    colData   = coldata,
    design    = ~ batch + genotype * ns(time, df = 3)
)
```

where counts was raw miRNA counts for each sample, coldata had the metadata for each sample's batch, genotype and sampling timepoint (labelled as "time" in this model), and design was the model I used for differential expression analysis.

Two likelihood ratio tests (LRTs) were used for differential expression analysis:

1. **Global time effect:** The full model included batch, genotype and spline(time) effect, compared to a reduced model without spline(time). This identified miRNAs which had significant differential expression throughout time, irrespective of genotype.
2. **Genotype specific trajectories:** The full model included batch, genotype, spline(time) and genotype x spline(time) interaction, compared to a reduced model without the genotype x spline(time) interaction. This identified miRNAs with genotype specific expression trajectories, suggesting a shift in pattern or abundance.

Batch effects were controlled for by including sequencing run as a covariate in all models.

3.2.6 *MiRNA target gene prediction*

MiRNA target gene prediction was performed locally with TargetFinder (https://github.com/carringtonlab/TargetFinder) (Fahlgren and Carrington 2010). *B. rapa* miRNA sequences were downloaded from PmiREN (release 2.0) (Guo, Kuang, and X. Yang 2025), and used as small RNA sequence input for TargetFinder. *B. rapa* R-o-18

reference coding DNA sequences (EnsemblPlants release 59) were used as the target database input. The prediction score cutoff was kept at the default value (-c 4).

3.2.7 *Target gene mRNA expression quantification and normalisation*

For gene expression profiling from Sarisha-14 and R-o-18 timeseries, I used sequencing data generated from (Calderwood, Hepworth, et al. 2021). The Illumina sequencing reads are available from the NCBI Sequence Read Archive, project ID PRJNA593493. The raw sequencing reads were quality checked, adapter trimmmed and aligned against the *B. rapa* R-o-18 reference genome (EnsemblPlants release 59), then gene expression counts were calculated using stringtie. Read alignment and gene expression quantification was performed using an automated script on the High Performance Computing cluster, provided by Dr Hugh Woolfenden (script available on gitlab : https://git.nbi.ac.uk/morris-group/bravo-scripts/-/tree/master/alignment_scripts). For gene expression abundance plots, library size normalised counts (CPM) were scaled with Trimmed Mean of M factors (TMM) to allow more accurate sample-to-sample gene expression abundance comparisons.

3.2.8 *Plotting results*

The heatmap in Figure 3.4 was plotted using the pheatmap R package (ver. 1.0.13, R. Kolde and M. R. Kolde (2015)). To visualise miRNA abundance dynamics (not raw miRNA abundance) over development, miRNA read counts were variance stabilised using the DESeq2 function varianceStabilizingTransformation (DEseq2 version 1.48.2), with transformation performed in a blind manner (blind = TRUE), accounting for sequencing batch, genotype and a spline model of time (ns(time), df = 3) in the initial design model. Mean abundance was calculated per timepoint for each miRNA, before performing z-score scaling across development for R-o-18 and Sarisha-14 respectively. The z-score values were used as input for pheatmap, and the miRNA rows were manually grouped based off their Sarisha-14 cluster, and then annotated by cluster. The developmental stages for each timepoint column were also annotated.

Comparison of miRNA and target gene expression was plotted per genotype. For each miRNA presented in a case study, one target gene was chosen out of the list of predicted target genes based off current literature for known miRNAs, or if it displayed reciprocal abundance dynamics with miRNA abundance. The TMM scaled counts from miRNA timeseries and gene expression timeseries were log2-transformed to allow for comparison of abundance dynamics, and anchored to the floral transition timepoint (for miRNA timeseries, this is day 12 in Sarisha-14 and day 19 in R-o-18; for target gene timeseries, this is day 10 in Sarisha-14 and day 17 in R-o-18). MiRNA and target gene

expression across days to floral transition was plotted with the ggplot2 R package (ver. 3.5.2, Wickham (2011)).

For miRNA and target gene expression abundance plots, TMM scaled counts were used from the respective miRNA and target gene expression timeseries. MiRNA and gene expression against time was plotted separately using the ggplot2 R package (ver. 3.5.2).

3.3 RESULTS

This chapter analyses the small RNA-seq timeseries data produced from the two *B. rapa* rapid-cycling cultivars, R-o-18 and Sarisha-14, across vegetative development and the floral transition. The data quality, small RNA library composition and high-level patterns of abundance variation (using PCA) was assessed in the previous chapter (Chapter 2), confirming the suitability of the dataset for downstream comparative analysis. The small RNA-seq data was aligned against a list of 169 total known and predicted novel *B. rapa* miRNA sequences sourced from PmiREN (2025) to measure the change in abundance of miRNAs throughout development. For downstream clustering analysis, the final timepoint that corresponded to floral bud emergence (BBCH51) was excluded from both the R-o-18 and Sarisha-14 timeseries datasets, as the distinct miRNA abundance at this stage impacted the clustering of abundance profiles during the earlier developmental stages.

To identify shared temporal miRNA abundance patterns between R-o-18 and Sarisha-14, the miRNA abundance temporal profiles were clustered in each genotype using fuzzy c-means clustering (2005). This method clusters the abundance profiles based on their similarity in their abundance dynamics over time, and assigns membership values to each miRNA based on its similarity to the cluster's mean abundance pattern (centroid). The membership values range from 0 to 1, and each miRNA in the dataset would have a total membership value of 1 across all the clusters. MiRNAs were primarily associated with the cluster where it has the highest membership value ($\geqslant$ 0.6), but this fuzzy clustering method also allows us to consider miRNAs with mixed temporal patterns, which would have lower membership values shared across more than one cluster.

3.3.1 *5 distinct abundance pattern clusters are identified in Sarisha-14 and R-o-18*

Of the 169 miRNAs, 123 and 128 miRNAs had detectable abundance (above a set minimum threshold of total 10 counts throughout the timeseries) in Sarisha-14 and R-o-18 respectively. Fuzzy c-means clustering using the R package Mfuzz (2005) was performed in each genotype, to assess whether their miRNAs shared temporal abundance patterns across genotypes. To determine the optimal cluster number, Dmin() analysis was performed (Supplementary Fig. XX), revealing genotype-specific optima at c=5-7 for Sarisha-14 and c=8-10 for R-o-18. While higher resolution clustering was possible, five clusters (c=5) were selected as this number effectively recapitulated key miRNA developmental trajectories while enabling direct comparison of abundance patterns across genotypes. The clustering revealed five major abundance patterns enabling direct comparison between R-o-18 and Sarisha-14. Only 79 miRNAs were strongly associated with a specific cluster (membership value $\geqslant 0.6$) in both genotypes, while the rest of the miRNAs did not have a strong association with any cluster pattern (membership value $\leqslant 0.6$). MiRNAs with low cluster membership values suggest that these miRNAs did not have a strong degree of association with any particular cluster, due to its profile having partial membership with more than one cluster. Thefollowing analyses will focus on high-membership miRNAs with distinct abundance profiles. The numeric labels for the clusters did not align in both genotypes since clustering was performed separately, so for ease of comparison, Table 3.1 is provided.

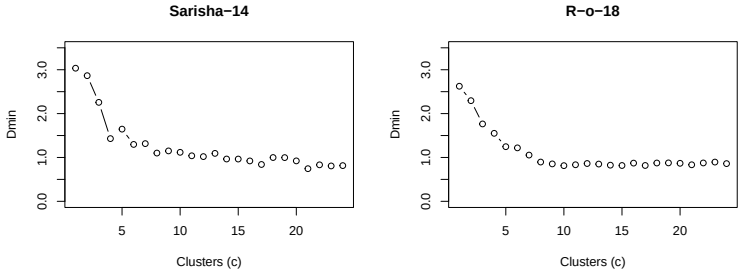


Figure 3.2: **Dmin() elbow analysis for fuzzy c-means cluster optimization.** Minimum centroid distances calculated across c=2-25 (repeats=5) for Sarisha-14 (left) and R-o-18 (right). Optimal resolution indicated at c=5-7 (Sarisha-14, red points) and c=8-10 (R-o-18, red points), but c=5 was selected for direct cross-genotype trajectory comparison while capturing major developmental patterns.

Fuzzy c-means clustering with $c = 5$ revealed five distinct temporal miRNA abundance patterns that could be directly compared between R-o-18 and Sarisha-14. In R-o-18, cluster 1 describes an abundance pattern consistent with early miRNA downregulation, peak during the floral transition (cluster 2), late downregulation (cluster 3), consistent upregulation (cluster 4), and a last cluster with miRNAs that are upregulated

late after the floral transition (cluster 5), but has no miRNAs strongly associated with this last cluster (Fig. 3.3).

In Sarisha-14, the clusters also reflect similar abundance patterns as the R-o-18 clusters, and instead of one cluster for constant upregulation like in cluster 4 in R-o-18, the miRNAs were grouped into two clusters, early upregulation (cluster 1), where the miRNAs were upregulated immediately after seedling stage, and late upregulation (cluster 3), where the miRNAs were at constant abundance and then increased right before the floral transition (Fig. 3.3.)

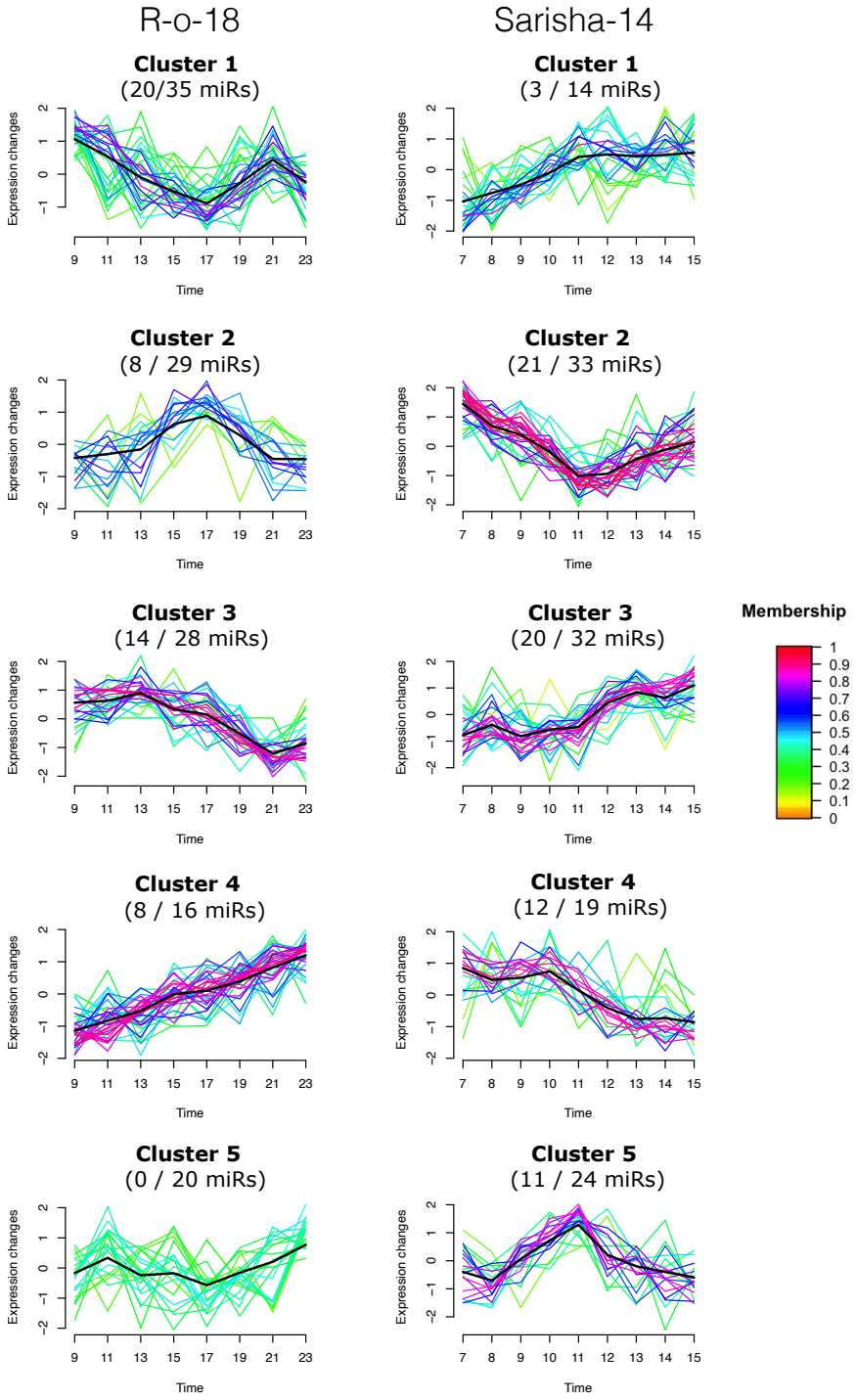


Figure 3.3: **Fuzzy c-means clustering of *B. rapa* miRNA reveals similar abundance patterns between R-o-18 and Sarisha-14 clusters.** miRNA abundance dynamics during the floral transition were clustered separately for each genotype using Mfuzz. Each line represents the normalized trajectory of a single miRNA within a cluster, colored by its membership strength (higher membership = red, low membership = green; see colour bar). The number of miRNAs with high membership (value $\geqslant 0.6$) is indicated against the total miRNAs assigned to each cluster (high/total miRs). The solid black line represents the cluster centroid, i.e. the trajectory estimated by fuzzy c-means, which summarizes the overall abundance trend within that cluster. Although cluster numbers are not directly comparable between genotypes, several clusters show similar dynamics (e.g. early upregulation, early/late downregulation, abundance peak). The final developmental stage (BBCH51) was excluded from the clustering analysis, as its distinct abundance profile could overshadow the earlier developmental abundance patterns.

Cluster pattern	R-o-18 cluster	Sarisha-14 cluster
Early downregulation (down_early)	1	2
Late downregulation (down_late)	3	4
Early upregulation (up_early)	4	5
Late upregulation (up_late)	4	3
Mid up, late down (peak)	2	5
Low confidence (NA)	5	1

Table 3.1: **Comparison of abundance patterns between R-o-18 and Sarisha-14 clusters.** Cluster patterns identified by Mfuzz were matched between the two genotypes. Note that R-o-18's cluster 4 upregulation pattern corresponds to Sarisha-14's cluster 5 (up_early) and cluster 3 (up_late). This table is a guide for comparing shared miRNA abundance dynamics.

3.3.2 *Identifying miRNAs shared between clusters in Sarisha-14 and R-o-18*

Figure 3.4 groups the miRNAs based on the Sarisha-14 cluster its most associated with, and visualizes its z-score scaled abundance pattern in R-o-18 and Sarisha-14. 58 of these miRNAs were in clusters with similar abundance patterns in both genotypes, while 21 miRNAs were found in different clusters with differing abundance patterns. In addition to comparing cluster memberships, differential expression analysis was performed using DESeq2 to identify miRNAs with significant divergent abundance between R-o-18 and Sarisha-14 across development. Curve registration, a mathematical technique which transforms the Sarisha-14 abundance profile using shift and stretch parameters to align Sarisha-14's abundance curve to R-o-18's abundance curve, was also applied to identify miRNAs which successfully aligned between both genotypes, suggesting abundance pattern alignment but with timing shifts in Sarisha-14. This also identified miRNAs that failed to register due to different abundance curves that cannot align between the two genotypes, which suggests that those miRNAs have cultivar-specific abundance patterns, and thus, could be associated with Sarisha-14's rapid floral development.

An example of miRNAs with shared abundance between R-o-18 and Sarisha-14 is Bra-miR159c (see section 3.3.4.1 for its target gene expression), which is found in R-o-18 cluster 4 and Sarisha-14 cluster 3 (Fig. 3.4), that both display upregulation patterns (Table 3.1), is successfully curve registered and is differentially expressed between genotypes (see Table 3.2).

Thus, with additional evidence from differential expression analysis and curve registration analysis, we've identified high-confidence miRNAs which share similar

abundance patterns with floral promoter Bra-miR172 and floral repressor Bra-miR156, as well as miRNAs with cultivar-specific abundance profiles which were associated with different clusters. The sections below present a few case studies of potential miRNAs which could have a regulatory role in flowering time.

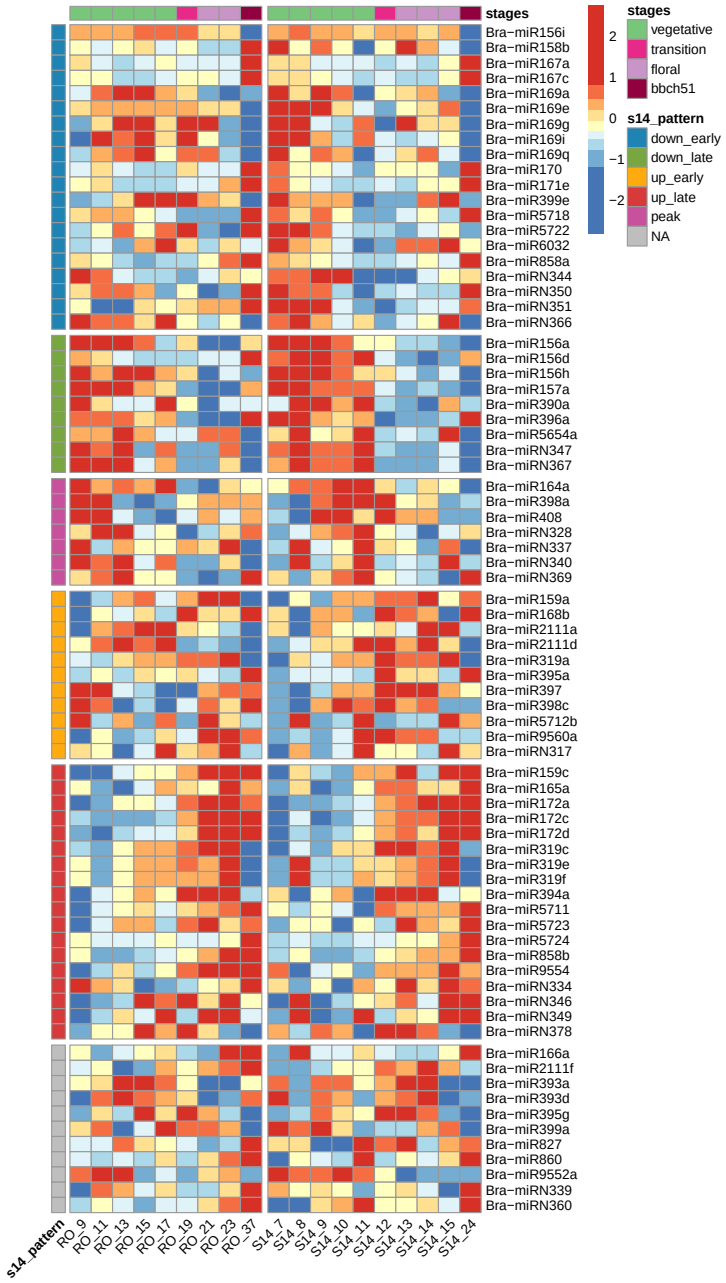


Figure 3.4: **Out of 79 clustered miRNAs, 31 miRNAs showed upregulation in Sarisha-14 before the floral transition, while 29 miRNAs were downregulated.** Rows represent individual miRNAs and columns represent developmental timepoints, with its corresponding stage annotated on the top. Abundance values were variance-stabilized and standardized to Z-scores per miRNA in each genotype. miRNAs are grouped according to their cluster in Sarisha-14, allowing direct comparison between R-o-18 and Sarisha-14 abundance dynamics. The abundance patterns of each cluster is annotated on the far left column.

miRNA	padj	curve registered	s14 pattern	ro pattern	category
Bra-miR156a	ns	TRUE	down_late	down_late	repressor
Bra-miR156d	p<0.005	TRUE	down_late	down_early	repressor
Bra-miR156h	ns	TRUE	down_late	down_late	repressor
Bra-miR157a	p<0.005	FALSE	down_late	down_late	repressor
Bra-miR390a	p<0.005	TRUE	down_late	down_late	repressor
Bra-miR396a	p<0.005	TRUE	down_late	down_late	repressor
Bra-miR5654a	ns	TRUE	down_late	down_early	repressor
Bra-miRN347	ns	TRUE	down_late	down_late	repressor
Bra-miRN367	p<0.005	TRUE	down_late	down_late	repressor
Bra-miR159c	ns	TRUE	up_late	up	promoter
Bra-miR165a	ns	TRUE	up_late	up	promoter
Bra-miR172a	p<0.05	TRUE	up_late	up	promoter
Bra-miR172c	p<0.005	TRUE	up_late	up	promoter
Bra-miR172d	ns	TRUE	up_late	up	promoter
Bra-miR319c	p<0.005	TRUE	up_late	up	promoter
Bra-miR319e	p<0.005	TRUE	up_late	up	promoter
Bra-miR319f	p<0.005	TRUE	up_late	up	promoter
Bra-miR394a	p<0.005	TRUE	up_late	up	promoter
Bra-miR5711	ns	TRUE	up_late	up	promoter
Bra-miR5723	ns	TRUE	up_late	up	promoter
Bra-miR858b	p<0.005	TRUE	up_late	up	promoter
Bra-miR9554	ns	TRUE	up_late	up	promoter
Bra-miRN334	ns	TRUE	up_late	down_early	promoter
Bra-miRN346	ns	TRUE	up_late	up	promoter
Bra-miRN349	ns	TRUE	up_late	up	promoter

Table 3.2: **List of candidate miRNAs with potential floral promoter/repressor roles based on abundance dynamics during the floral transition.** (caption continued in next page)

Table 3.2 (continued)
The padj column is the statistical significance from differential expression analysis of the miRNAs with genotype x time interaction, highlighting miRNAs with genotype-specific abundance patterns ($p<0.05$). The curve registration column is a summary of whether the abundance profiles of these miRNAs are aligned in both cultivars when taking into account time/abundance level (curve registered = TRUE); or not aligned (curve registered = FALSE). The two columns "s14 pattern" and "ro pattern" are the abundance patterns corresponding with the cluster which the miRNAs are assigned to with Mfuzz clustering (see Fig. 3.3 and Table 3.1 for visualisation of the clusters and their abundance pattern description). The category column assigns a potential regulatory role "repressor" or "promoter", based off the miRNA's abundance dynamics during the floral transition (miRNAs that are upregulated during the floral transition would be potential "promoters", while miRNAs that are downregulated in contrast would be potential "repressors").

3.3.3 *Case studies of potential miRNAs that repress flowering*

In this section, I will focus on 5 case studies of candidate miRNAs whose abundance profiles suggest a repressive role, repressing flowering-associated gene expression until the floral transition. Similar to Bra-miR156 abundance in the "late downregulation" clusters, these miRNAs are highly expressed during the vegetative stage, then rapidly downregulation at the floral transition stage. This is reflected by the upregulation of their predicted target genes, which are related to flowering, growth or stress response. The target genes for these miRNAs were predicted using TargetFinder (*carringtonlab/-TargetFinder* 2024) (Fig. 3.5). For each case study, abundance dynamics of the miRNA and its predicted targets were compared across both genotypes, aligned to the floral transition (day 0).

```
Candidate floral repressors
(a) Bra-miR157a – SPL9 (score = 2.5)     A04p003490.1_BraROA
    Target  5′ GUGUUCUCUCUCUUCUGUCAA 3′
              :::.::::: ::::::::::
    miRNA   3′ CACGAGAGAUAGAAGACAGUU 5′

(b) Bra-miR390a – AL7 (score = 3.5)      A08p040520.1_BraROA
    Target  5′ GACGAGGUCCCUCCUGAGCUU 3′
              : ::  .:::::::::::::::
    miRNA   3′ CCGCGAUAGGGAGGACUCGAA 5′

(c) Bra-miR396a – GRF4 (score = 4)       A04p008130.1_BraROA
    Target  5′ CCGUUCAAGAAAGCCUGUGGAA 3′
              :::::::::::: :::::::
    miRNA   3′ UUCAAGUUCUUUCG-ACACCUU 5′

(d) Bra-miRN347 – IPT2 (score = 4)       A04p026570.1_BraROA
    Target  5′ GAGAUCAUAAACGCCGACGC 3′
              :::::::::::: :::::
    miRNA   3′ CUCUAGUAUUUGAGGCUGUU 5′

(e) Bra-miRN367 – VDAC1 (score = 4)      A05p055140.1_BraROA
    Target  5′ GACCA-AAACUCUGGCAAGAU 3′
              :: :: :::::::.:.:::::::
    miRNA   3′ CUAGUAUUUGAGGCUGUUCUA 5′
```

Figure 3.5: **Predicted target sites for floral repressor miRNAs** (Bra-miR157a, Bra-miR390a, Bra-miR396a, Bra-miRN347, Bra-miRN367). Each panel shows the miRNA–target duplex discussed in the case studies, with target (5′→3′) above and miRNA (3′→5′) below. Colons (:) indicate matched bases, dots (.) mismatches, and dashes (–) gaps. Lower targetFinder scores indicate stronger predicted interactions. These miRNAs are downregulated at floral transition, correlating with target gene upregulation.

3.3.3.1 *Bra-miR157a*

miR157 is considered a homologue of miR156, sharing similar repressive roles and target genes (Gandikota et al. 2007). Bra-miR157a showed clear conservation of abundance pattern with Bra-miR156, with high abundance at the vegetative stage (between 50,000 TMM in R-o-18 and 75,000 TMM in Sarisha-14), steadily decreasing until the floral transition (Fig. 3.6). Its downregulation coincided with the induction of one of its predicted targets, *A04p003490.1_BraROA*, an orthologue of *SPL9*, a well-established floral activator (M. Xu, T. Hu, Zhao, et al. 2016). Bra-miR157a downregulation also occurred earlier in Sarisha-14, consistent with this genotype's earlier progression to flowering.

3.3.3.2 *Bra-miR390a*

Bra-miR390a was predicted to target *A08p040520.1_BraROA*, an orthologue of *ALFIN-LIKE 7 (AL7)*, a chromatin-associated protein which reads H3K27me3 repressive marks (Liang et al. 2018; Molitor et al. 2014) (Figure 3.7). Interestingly, although Bra-miR390a is clustered in similar patterns in Sarisha-14 and R-o-18 (late downregulation, see Table 3.2), their abundance patterns early in development seem distinct from each other, with Bra-miR390a increasing during the vegetative stage, while Bra-miR390a decreases at the same stage in R-o-18. Bra-miR390a abundance declined before the floral transition, while its target *AL7* increased sharply during the same period. This reciprocal abundance pattern is consistent with a potential repressive interaction, where miR390a downregulation may allow upregulation of AL7 during the floral transition.

3.3.3.3 *Bra-miR396a*

Bra-miR396a in R-o-18 and Sarisha-14 similarly decreased in abundance prior to the floral transition (Fig. 3.8), with a rebound in the late floral stage (BBCH51, floral buds visible). Its predicted target, *A04p008130.1_BraROA*, a putative orthologue of *GROWTH REGULATING FACTOR 4 (GRF4)*, is a transcription factor implicated in growth regulation (J. H. Kim and Lee 2006). The opposing trajectories of Bra-miR396a and its target suggest that growth regulation may have cross-talk with flowering time regulation.

3.3.3.4 *Bra-miRN347*

The PmiREN database provides a list of potential novel miRNAs in *B.rapa*, and some were identified in this analysis as potential regulators associated with flowering time. Bra-miRN347 is predicted to target an isopentenyltransferase orthologue (*IPT2*), a key regulator of cytokinin biogenesis (Nguyen et al. 2021) (Fig. 3.9). Bra-miRN347 abundance declined before the floral transition, coinciding with increased abundance of its target in both genotypes, suggesting a potential regulatory role in cytokine signalling during the vegetative-to-floral transition.

3.3.3.5 *Bra-miRN367*

Lastly, Bra-miRN367 displayed rapid 3-fold downregulation before the floral transition (Fig. 3.10) in both cultivars. Its predicted target, *A05p055140.1_BraROA*, a putative orthologue of *Voltage Dependent Anion Channel 1 (VDAC1)*, increased sharply during the transition, particularly in Sarisha-14, where abundance nearly doubled (Fig. 3.10 (bottom right panel)). This strong inverse relationship highlights Bra-miRN367 as a candidate regulator whose repression may enable target gene activation at the transition stage.

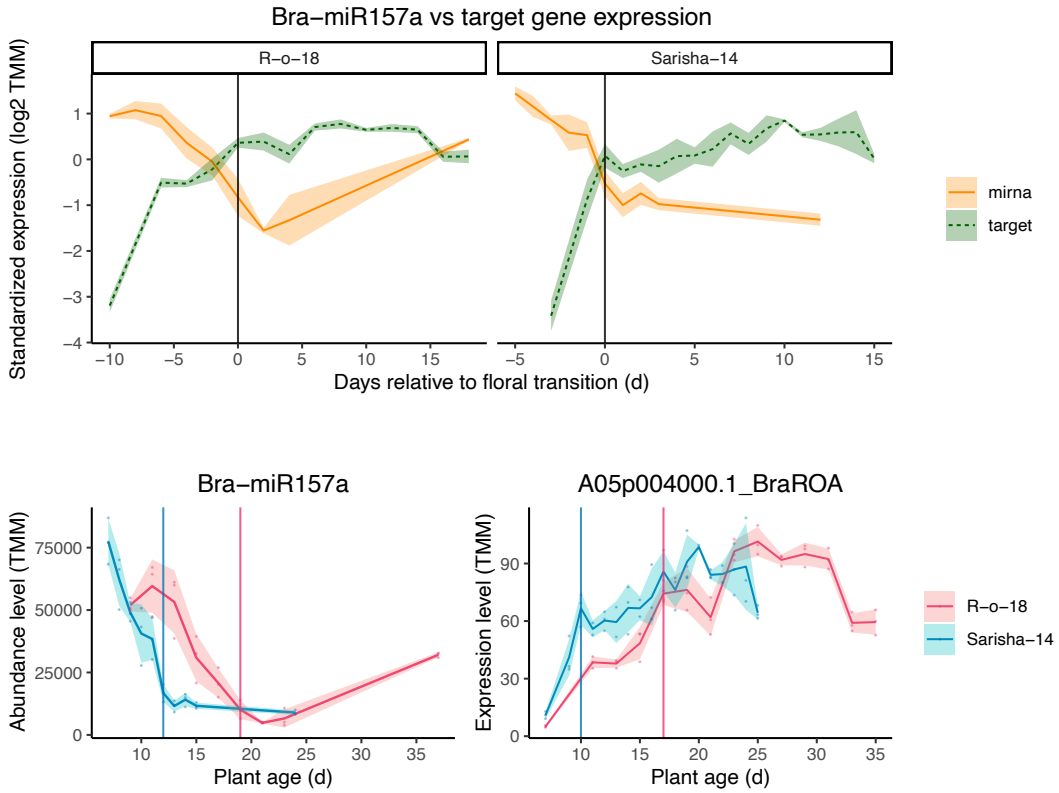


Figure 3.6: **Bra-miR157a downregulation aligns with increased abundance of its predicted target gene *A04p003490.1_BraROA*, an orthologue of *SPL9*.**
(Top) Overlay of normalized abundance profiles for Bra-miR157a and *A04p003490.1_BraROA*, aligned to the floral transition (day 0).
(Bottom left) abundance pattern of Bra-miR157a, which shows an earlier shift in downregulation in Sarisha-14.
(Bottom right) abundance pattern of *A04p003490.1_BraROA*, which increases during the floral transition, consistent with Bra-miR157a downregulation.

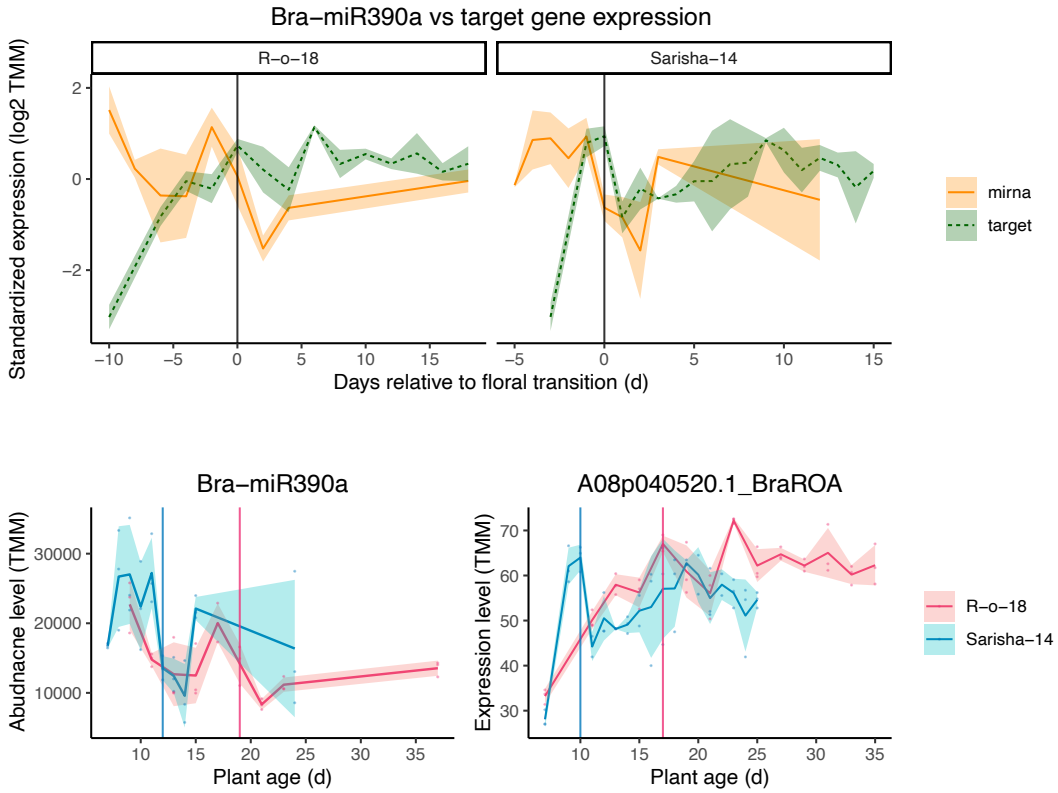


Figure 3.7: **Bra-miR390a targets *A08p040520.1_BraROA*, a predicted orthologue of ALFIN-LIKE 7 (AL7), a member of PHD domain family proteins which reads H3K27me3 marks.** **(Top)** Overlay of normalized abundance profiles for Bra-miR390a and *A08p040520.1_BraROA*, aligned to the floral transition (day 0).
(Bottom left) abundance pattern of Bra-miR390a, which shows downregulation prior to the floral transition (timepoint marked with vertical lines for each genotype).
(Bottom right) abundance pattern of target gene *A08p040520.1_BraROA*, which rapidly increases as the cultivars approach the floral transition.

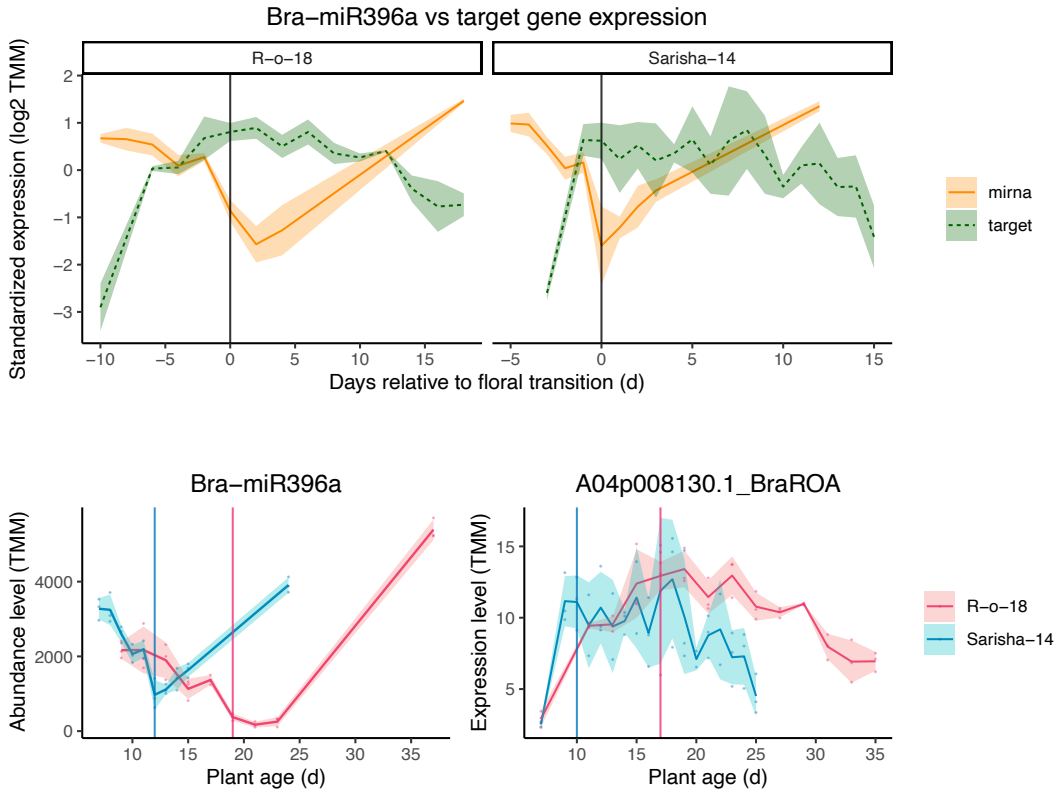


Figure 3.8: **Bra-miR396a is predicted to target *A04p008130.1_BraROA*, a potential orthologue of Arabidopsis *Growth Regulating Factor 4 (GRF4)*.**
(Top) Overlay of normalized abundance profiles for Bra-miR396a and *A04p008130.1_BraROA*, aligned to the floral transition (day 0).
(Bottom left) abundance pattern of Bra-miR396a, which shows downregulation prior to the floral transition (timepoint marked with vertical lines).
(Bottom right) abundance pattern of target gene *A04p008130.1_BraROA*.

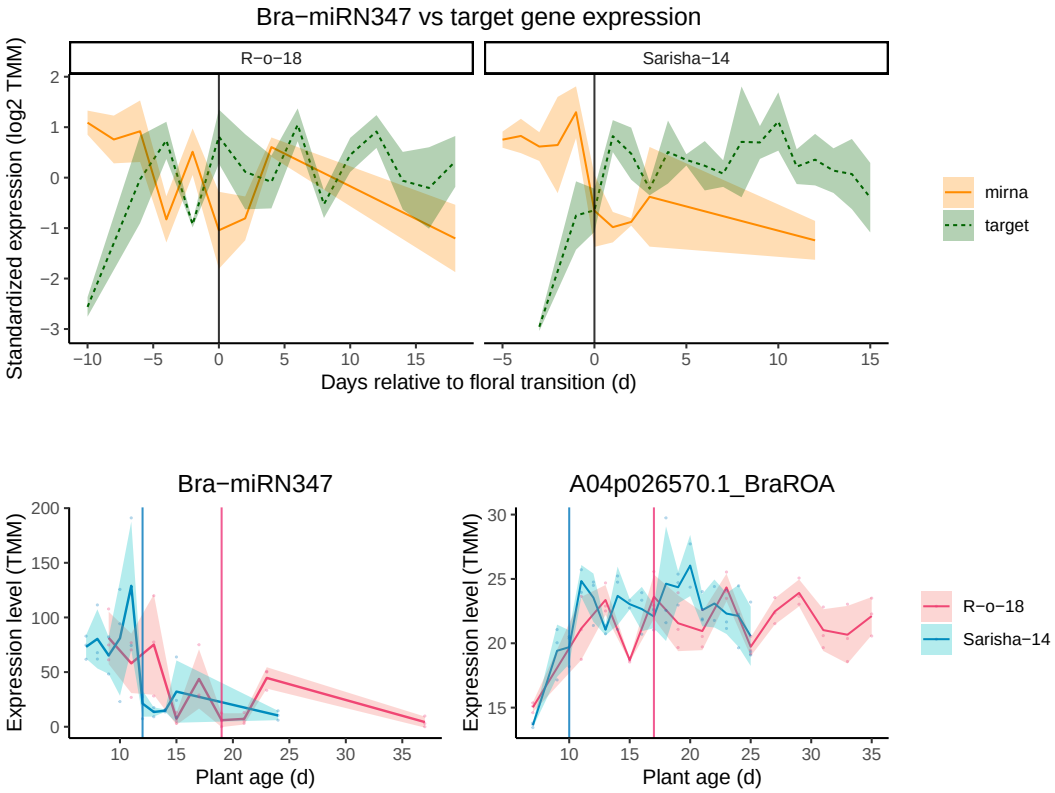


Figure 3.9: **Novel predicted miRNA Bra-miRN347 targets an isopentenyltransferase orthologue *IPT2*, possibly regulating cytokinin biogenesis.**
(Top) Overlay of normalized abundance profiles for Bra-miRN347 and *A04p026570.1_BraROA*, aligned to the floral transition (day 0). Comparison of their relative abundance patterns suggests that Bra-miRN347 abundance represses *A04p026570.1_BraROA*, especially since its variation is also reflected in the target gene expression.
(Bottom left) abundance pattern of Bra-miRN347, which is downregulated during the floral transition (this stage is marked with vertical lines).
(Bottom right) abundance pattern of target gene *A04p026570.1_BraROA*, which increases at the same rate in Sarisha-14 and R-o-18, possibly reflecting derepression by Bra-miRN347 downregulation.

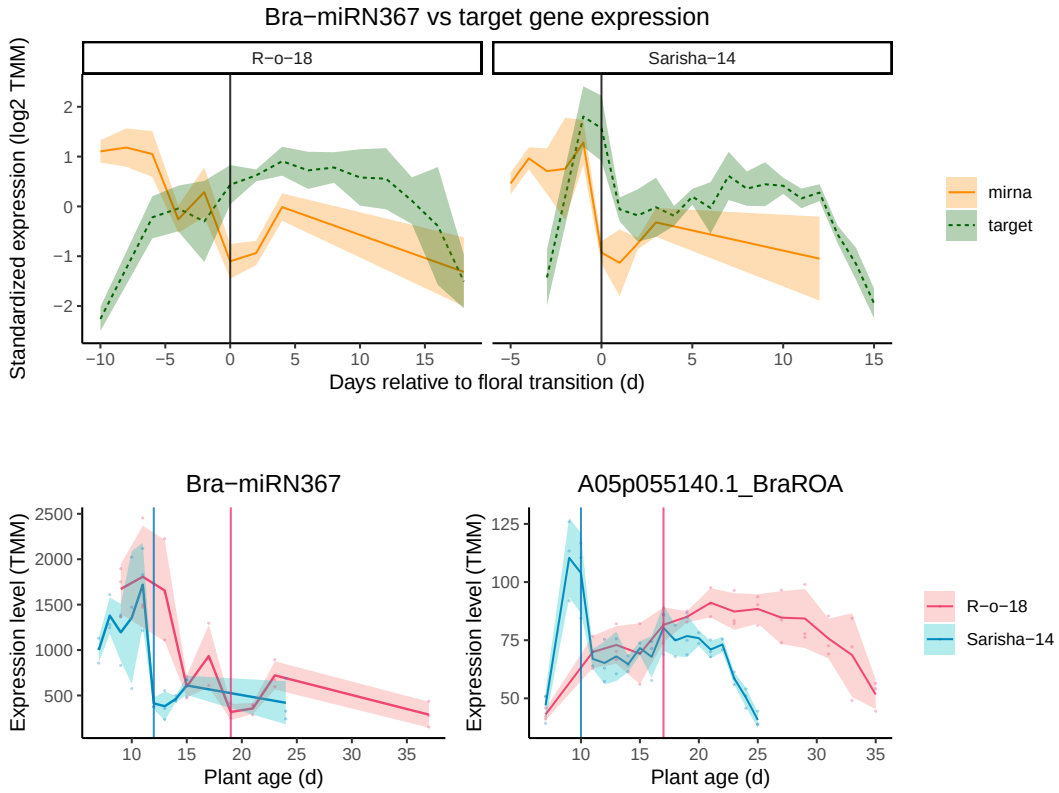


Figure 3.10: **Bra-miRN367 is predicted to target an orthologue of *Voltage-Dependent Anion Channels 1 (VDAC1).***
(Top) Overlay of normalized abundance profiles for Bra-miRN367 and *A05p055140.1_BraROA*, aligned to the floral transition (day 0).
(Bottom left) abundance pattern of Bra-miRN367, which shows rapid downregulation prior to the floral transition (this stage is marked with vertical lines).
(Bottom right) abundance pattern of target gene *A05p055140.1_BraROA*, which increases almost 2-fold in Sarisha-14 and 1.5-fold in R-o-18 at the floral transition stage.

To summarise, Bra-miR157a, Bra-miR390a, Bra-miR396a, Bra-miRN347, and Bra-miRN367 are 5 candidate miRNAs with potential repressive role in regulating flowering. Similar to known floral repressor Bra-miR156, these miRNAs are highly expressed during vegetative growth and downregulated during the floral transition, leading to the upregulation of target genes linked to developmental and hormonal pathways as plants approach reproductive competence.

3.3.4 *Case studies of potential miRNAs that promote flowering*

Alongside miRNAs with potential floral repressor roles, I identified three case studies with miRNAs that are potential floral promoters. These miRNAs are upregulated leading up to the floral transition, and come from the same abundance cluster as Bra-miR172 (late upregulation). The target genes for these candidate floral promoters were predicted using TargetFinder (*carringtonlab/TargetFinder* 2024) (Fig. 3.11). For each case,

abundance dynamics of the miRNA and its predicted targets were compared across both genotypes, aligned to the floral transition (day 0).

```
Candidate floral promoters
(a) Bra-miR159c - TIE4 (score = 4)      A03p019540.1_BraROA
    Target  5′ AGAGCCCCCUUCAAACCAAA 3′
              :::: :::::::: :::::
    miRNA   3′ CCUCGAGGGAAGUUAGGUUU 5′

(b) Bra-miR159c - TIE4 (score = 4)      A03p019540.1_BraROA
    Target  5′ AGAGCCCCCUUCAAACCAAA 3′
              :::: :::::::: :::::
    miRNA   3′ CCUCGAGGGAAGUUAGGUUU 5′

(c) Bra-miR319c - TCP4 (score = 3)      A05p044300.1_BraROA
    Target  5′ AGAGGGGUCCCCUUCAGUCCAG 3′
              :: ::.:: ::::::::::::.
    miRNA   3′ UC-CCUCAAGGGAAGUCAGGUU 5′

(d) Bra-miR394a - LCR (score = 1)       A09p050010.1_BraROA
    Target  5′ GGAGGUUGACAGAAUGCCAA 3′
              :::::: ::::::::::::
    miRNA   3′ CCUCCACCUGUCUUACGGUU 5′
```

Figure 3.11: **Predicted target sites for floral promoter miRNAs** (Bra-miR159c, Bra-miR319c, Bra-miR394a). Each panel shows the miRNA–target duplex discussed in the case studies, with target (5′→3′) above and miRNA (3′→5′) below. Colons (:) indicate matched bases, dots (.) mismatches, and dashes (–) gaps. Lower targetFinder scores indicate stronger predicted interactions. These miRNAs are upregulated leading up to floral transition, correlating with target gene downregulation.

3.3.4.1 *Bra-miR159c*

Bra-miR159c is predicted to target two genes: *A03p019540.1_BraROA*, a putative orthologue of *TCP Interactor containing EAR motif protein 4 (TIE4)*, and *A03p040620.1_BraROA*, a predicted
metallo-hydrolase/oxidoreductase encoding gene (Figure 3.12). Bra-miR159c abundance increased prior to the floral transition in both cultivars, with consistently higher levels in R-o-18 compared to Sarisha-14. Both predicted targets were progressively downregulated over time, consistent with Bra-miR159c upregulation. Interestingly, the abundance of the metallo-hydrolase/oxidoreductase encoding gene (*A03p040620.1_BraROA*) remained slightly higher in Sarisha-14, consistent with the lower abundance of Bra-miR159c in this genotype, suggesting potential miRNA dosage level repression.

3.3.4.2 *Bra-miR319c*

miR319 is part of the miR159/miR319 family which is known to target TCP transcription factors as part of a flowering regulatory module (Fang et al. 2021). This regulatory module seems to be conserved in *B. rapa* as well, with Bra-miR319c targeting TCP transcription factors including the predicted *TCP4* orthologue *A05p044300.1_BraROA* (Figure 3.13). Bra-miR319c abundance increased as both cultivars approached the floral transition, while *TCP4* expression declined at the floral transition, which points to *TCP4* repression by Bra-miR159c. Interestingly, *TCP4* expression increased again at the final timepoint (BBCH51) in both cultivars, consistent with Bra-miR319c downregulation at the same timepoint.

3.3.4.3 *Bra-miR394a*

Finally, Bra-miR394a is predicted to target *A09p050010.1_BraROA*, a putative orthologue of *LEAF CURLING RESPONSIVENESS (LCR)* (Figure 3.14). Bra-miR394a increased leading up to the floral transition, while *LCR* expression was downregulated towards this stage, possibly due to Bra-miR394a repression.

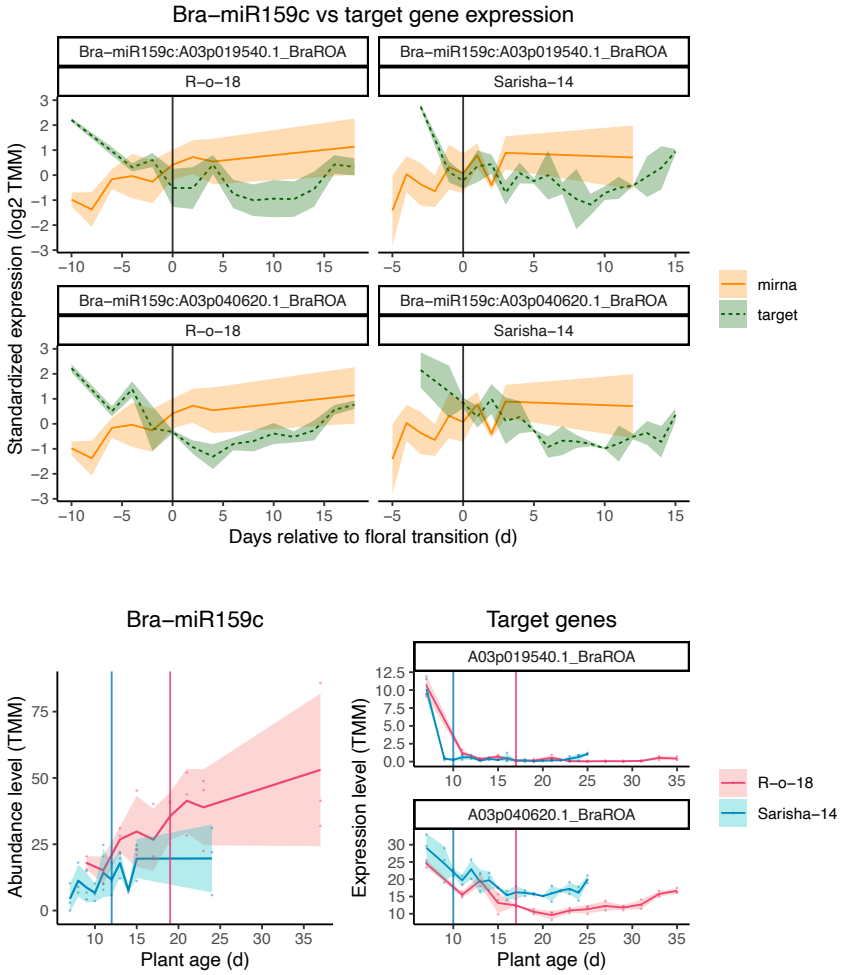


Figure 3.12: **Bra-miR159c is predicted to target *A03p019540.1_BraROA*, a predicted orthologue of *TCP Interactor containing EAR motif protein4 (TIE4)* and *A03p040620.1_BraROA*, a predicted Metallo-hydrolase/oxidoreductase orthologue.**
(Top) Overlay of normalized expression profiles for Bra-miR159c and its target genes, aligned to the floral transition (day 0).
(Bottom left) Abundance pattern of Bra-miR159c, which increases in abundance as both cultivars approach the floral transition. This miRNA is expressed higher in R-o-18 than Sarisha-14.
(Bottom right) Abundance pattern of *A03p019540.1_BraROA* and *A03p040620.1_BraROA*, which decreases in abundance throughout development. *A03p040620.1_BraROA* is expressed slightly higher in Sarisha-14, consistent with Bra-miR159c's lower abundance in Sarisha-14.

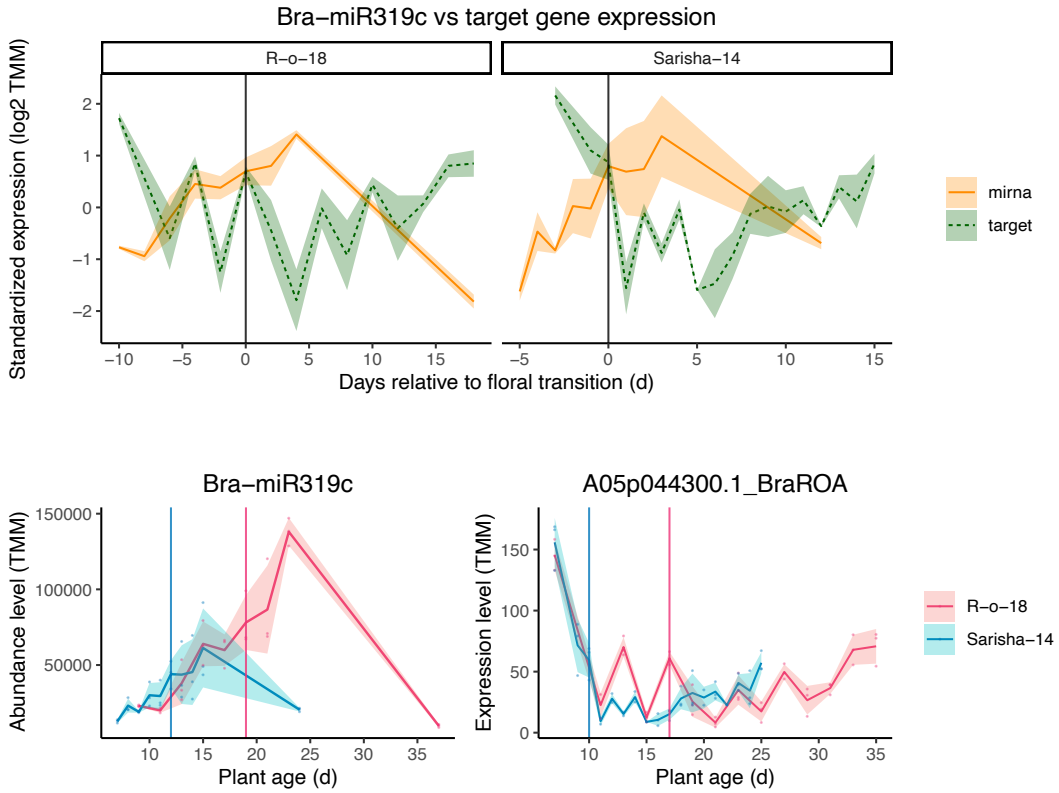


Figure 3.13: **Bra-miR319c targets several TCP transcription factors, including *TCP4* predicted orthologue *A05p044300.1_BraROA*, suggesting a miR319-TCP4 regulation of floral development.**
(Top) Overlay of normalized expression profiles for Bra-miR319c and its target *A05p044300.1_BraROA*, aligned to the floral transition (day 0).
(Bottom left) Abundance pattern of Bra-miR319c, which increases in abundance as both cultivars approach the floral transition.
(Bottom right) Expression pattern of *A05p044300.1_BraROA*, which decreases in expression in line with Bra-miR319c upregulation, and then increases again as the cultivars approach the final timepoint (BBCH51).

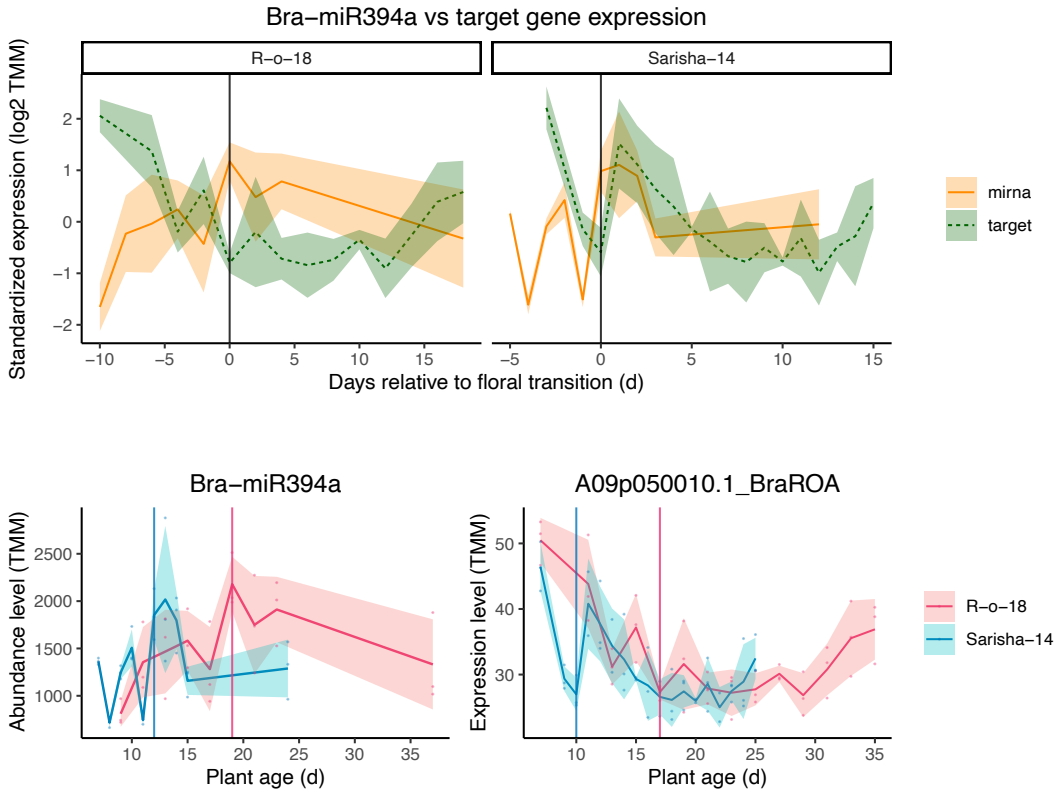


Figure 3.14: **Bra-miR394a targets a potential orthologue of *LEAF CURLING RESPONSIVENESS (LCR)*, repressing it before the floral transition.**
(Top) Overlay of normalized expression profiles for Bra-miR394a and *A09p050010.1_BraROA*, aligned to the floral transition (day 0).
(Bottom left) Abundance pattern of Bra-miR394a, which is upregulated leading up to the floral transition (timepoint marked with vertical lines for each genotype).
(Bottom right) Expression pattern of target gene *A09p050010.1_BraROA*, which rapidly increases as the cultivars approach the floral transition.

In summary, these case studies highlight a set of candidate miRNAs: Bra-miR159c, Bra-miR319c, and Bra-miR394a that exhibit late upregulation prior to the floral transition, similar to the known floral promoter Bra-miR172. In each case, their abundance dynamics were consistent with repression of predicted targets, including transcription factors and stress-associated genes, suggesting potential roles as flowering promoters. Although the abundance levels of these miRNAs are not always similar at the floral transition stage, the earlier upregulation of these miRNAs support our hypothesis that there are conserved floral promoter modules for which a temporal shift (or lower abundance threshold) may underlie Sarisha-14's earlier flowering phenotype.

3.3.5 *Cultivar-specific miRNAs may have a role in Sarisha-14's early flowering phenotype, but their target genes need more investigation*

The Mfuzz clustering analysis allowed us to identify miRNAs assigned to different clusters in Sarisha-14 and R-o-18 respectively. Curve registration confirmed that their abundance profiles failed to align between the two cultivars, which indicates genuine differences in abundance dynamics and not merely due to clustering thresholds. Differential expression analysis also found that these miRNAs had significant genotype x time interactions, further supporting that this subset of miRNAs display cultivar-specific temporal abundance patterns (Table 3.3).

mirna	padj	curve registered	s14 pattern	ro pattern
Bra-miR169a	p<0.005	FALSE	down_early	down_late
Bra-miR169e	p<0.05	FALSE	down_early	down_late
Bra-miR169q	p<0.005	FALSE	down_early	peak
Bra-miR393a	p<0.05	FALSE	NA	NA
Bra-miR393d	p<0.005	FALSE	NA	NA
Bra-miR397	p<0.005	FALSE	up_early	down_early
Bra-miR398a	p<0.005	FALSE	peak	down_early
Bra-miR399a	p<0.005	FALSE	NA	NA
Bra-miR399e	p<0.005	FALSE	down_early	peak
Bra-miR408	p<0.005	FALSE	peak	down_early
Bra-miR860	p<0.005	FALSE	NA	NA
Bra-miR2111d	p<0.005	FALSE	up_early	down_late
Bra-miR5712b	p<0.05	FALSE	up_early	down_early
Bra-miR6032	p<0.05	FALSE	down_early	peak
Bra-miRN340	p<0.005	FALSE	peak	down_late
Bra-miRN360	p<0.05	FALSE	NA	NA
Bra-miRN366	p<0.005	FALSE	down_early	down_late
Bra-miRN378	p<0.05	FALSE	up_late	peak
Bra-miR157a	p<0.005	FALSE	down_late	down_late
Bra-miR5724	p<0.005	FALSE	up_late	up_late
Bra-miR9560a	p<0.005	FALSE	up_early	up

Table 3.3: **List of miRNAs with cultivar specific abundance patterns.** The padj column is the statistical significance from differential expression analysis of the miRNAs with genotype x time interaction, highlighting miRNAs with genotype-specific abundance patterns (p<0.05). The curve registration column is a summary of whether the abundance profiles of these miRNAs are aligned in both cultivars when taking into account time/abundance level (curve registered = TRUE); or not aligned (curve registered = FALSE). The two columns "s14 pattern" and "ro pattern" are the abundance patterns corresponding with the cluster which the miRNAs are assigned to with Mfuzz clustering (see Fig. 3.3 and Table 3.1 for visualisation of the clusters and their abundance pattern description). Some miRNAs were not assigned by Mfuzz to any specific clusters in Sarisha-14 or R-o-18, thus have "NA" in those columns.

Although I predicted several target genes for these miRNAs (see Fig. 3.15), they don't display cultivar-specific gene expression unlike the miRNAs which have divergent abundance profiles. This could be due to redundant regulatory mechanisms besides the miRNAs which support target gene expression, but it is likely due to our limitation in target gene prediction without experimental validation.

```
Cultivar-divergent miRNAs
(a) Bra-miR398a - DREB2B (score = 4)     A01p054070.1_BraROA
    Target  5′ AAGGAG-GACCUGAGAACACU 3′
              ::: : ::::::::::::
    miRNA   3′ GUCCCCACUGGACUCUUGUGU 5′

(b) Bra-miR397 - LAC2 (score = 3.5)      A05p018330.1_BraROA
    Target  5′ UGAUCAAUGCUGCACUCGAUGA 3′
              ::::::.:::::::::.::::
    miRNA   3′ UGUAGUUGCGACGUGAGUUACU 5′

(ec Bra-miR408 - LAC3 (score = 4.0)      A05p017350.1_BraROA
    Target  5′ AACCAGUGAAGAGGCUGUGCA 3′
              : :::: :::::::: :::::
    miRNA   3′ UCGGUCCCUUCUCCGUCACGU 5′
```

Figure 3.15: **Predicted target sites for cultivar-divergent miRNAs** (Bra-miR398a, Bra-miR397). Each panel shows the miRNA–target duplex discussed in the case studies, with target (5′→3′) above and miRNA (3′→5′) below. Colons (:) indicate matched bases, dots (.) mismatches, and dashes (–) gaps. Lower TargetFinder scores indicate stronger predicted interactions. These miRNAs show distinct abundance trajectories between Sarisha-14 and R-o-18. Target predictions for Bra-miR169 and Bra-miR399 failed to identify potential target genes relevant to flowering, nor exhibited a reciprocal expression pattern to the miRNAs, so were not included in this figure.

3.3.5.1 *Bra-miR398a*

Bra-miR398a shows strong cultivar-specific abundance differences, with Sarisha-14 expressing more Bra-miR398a, which is rapidly upregulated almost 10-fold before the floral transition, in contrast to lower and declining abundance in R-o-18 (Figure 3.16). Its predicted target, *A01p054070.1_BraROA*, a putative orthologue of *DREB2B*, accumulates with similar trajectories in both cultivars, which suggests that the differences in Bra-miR398a abundance are not reflected by repression of this target gene. This points to either redundant regulatory mechanisms that also repress this gene, or the possibility that Bra-miR398a regulates a different set of genes in the context of early flowering.

3.3.5.2 *Bra-miR169*

Several members of the Bra-miR169 family display divergent abundance dynamics between cultivars (Figure 3.17, upper panel). In Sarisha-14, early downregulation of Bra-miR169 members is observed, while in R-o-18 these miRNAs show either late downregulation (Bra-miR169a) or a peak in abundance just prior to the floral transition (Bra-miR169e, Bra-miR169q). This suggests that Bra-miR169a plays a repressive role in flowering in both cultivars, but other members such as Bra-miR169e, which is the most highly expressed miR169 member in both cultivars, have divergent regulatory roles. Studies have shown that miR169 regulates Nuclear Factor Y subunit A (NF-YA) transcription factors in Arabidopsis to control flowering (M. Y. Xu et al. 2013; Jones-Rhoades and Bartel 2004), but target prediction failed to identify any NF-YA orthologues as target genes of this miRNA.

3.3.5.3 *Bra-miR399*

Bra-miR399a and Bra-miR399e also show cultivar-specific patterns, with higher abundance levels in Sarisha-14 compared to R-o-18 (Figure 3.18). While these miRNAs are downregulated early in Sarisha-14, they peak in abundance in R-o-18, reflecting a clear temporal shift in regulation. The elevated abundance of Bra-miR399 members in Sarisha-14 suggests a potential role in regulating flowering pathways, but current target predictions do not identify any floral genes or display reciprocal abundance dynamics, so their precise function is still uncertain.

3.3.5.4 *Bra-miR397 and Bra-miR408*

Bra-miR397 is strongly expressed in Sarisha-14 and inversely expressed with its predicted target *A05p018330.1_BraROA*, a putative orthologue of *LACCASE2 (LAC2)* (Figure 3.19). However, the target also declines in abundance in R-o-18, where Bra-miR397 levels are low, indicating that additional regulators are likely at play. Bra-miR408, another highly expressed miRNA in Sarisha-14, is also predicted to target a *LACCASE3* orthologue. Target gene expression is undetectable, raising the possibility of cleavage and degradation, but this requires experimental validation.

These cultivar-specific miRNAs we've highlighted in our case studies display how Sarisha-14 and R-o-18 may diverge in the timing and abundance of miRNA abundance during the floral transition. In several cases, such as Bra-miR169 and Bra-miR399, its higher abundance and different abundance dynamics in Sarisha-14 could feasibly contribute to its early flowering phenotype. However, the absence of consistent relationships between miRNA abundance and predicted target gene expression, highlights a gap in functional validation of these miRNAs. While these miRNAs remain promising candidates for explaining cultivar-specific developmental timing, further investigation is required to clarify their targets and determine their roles in regulating flowering.

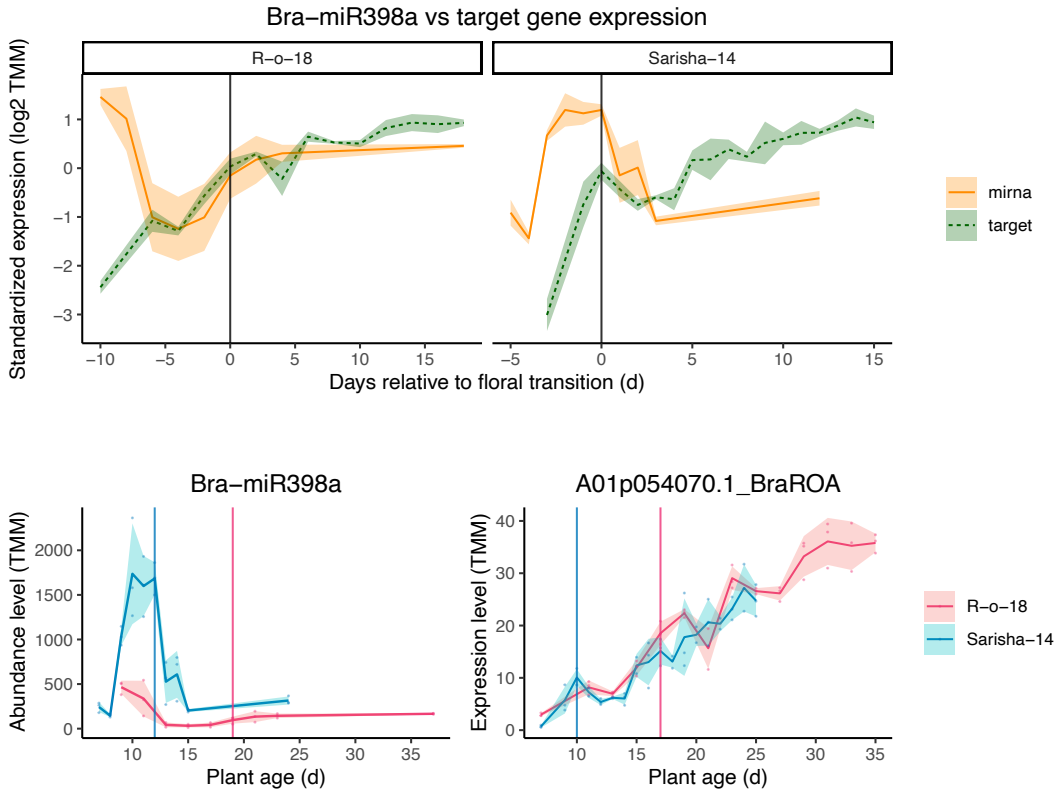
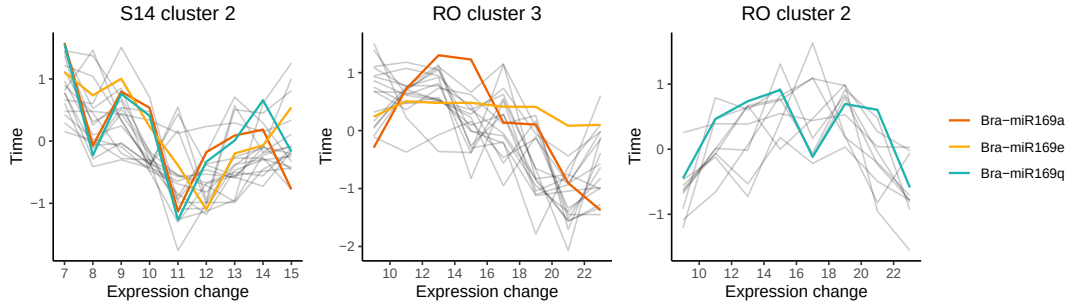
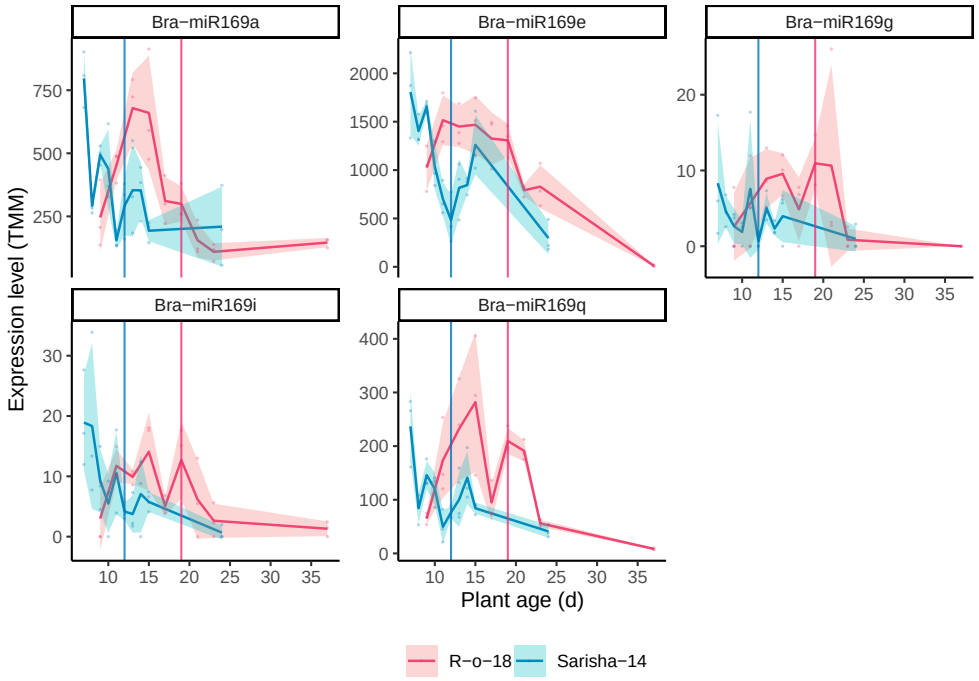


Figure 3.16: **Bra-miR398a is predicted to target an orthologue of *DREB2B*, which is a dehydration and heat inducible transcription factor. However, diverging abundance patterns and levels between Sarisha-14 and R-o-18 do not align with target gene expression patterns.**
(Top) Overlay of normalized abundance profiles for Bra-miR398a and *A01p054070.1_BraROA*, aligned to the floral transition (day 0).
(Bottom left) abundance pattern of Bra-miR398a, which is rapidly upregulated in Sarisha-14 prior to the floral transition (marked with vertical lines for each genotype), while in contrast, R-o-18 shows lower abundance and downregulation before the floral transition.
(Bottom right) abundance pattern of target gene *A01p054070.1_BraROA*, which increases throughout development, with similar abundance patterns in both genotypes. This shows that the differential expression levels of Bra-miR398a are not reflected in differential regulation of this target gene.

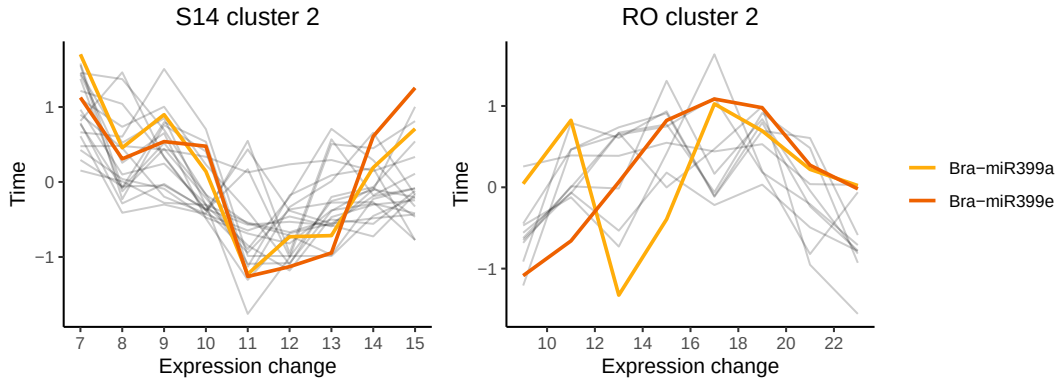


(a) Cluster comparisons

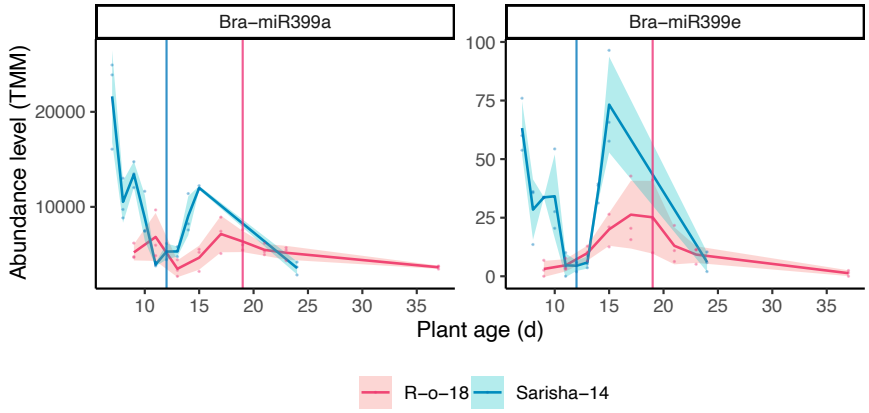


(b) Bra-miR169 abundance

Figure 3.17: **Bra-miR169 members display cultivar specific abundance patterns, with early downregulation in Sarisha-14, but late downregulation or peak in abundance before the floral transition in R-o-18.**
a) Comparisons between clusters from Sarisha-14 and R-o-18, with Bra-miR169a, Bra-miR169e and Bra-miR169q abundance patterns highlighted in either genotype. These miRNAs are the only members which were strongly associated with its respective cluster using Mfuzz. MiRNAs in S14 cluster 2 display early downregulation pattern, while miRNAs in RO cluster 3 undergo late downregulation, and RO cluster 2 shows peak abundance.
b) abundance of Bra-miR169 members across development in R-o-18 and Sarisha-14, highlighting cultivar-specific differences in abundance pattern. Vertical bars mark the floral transition timepoint for each cultivar.

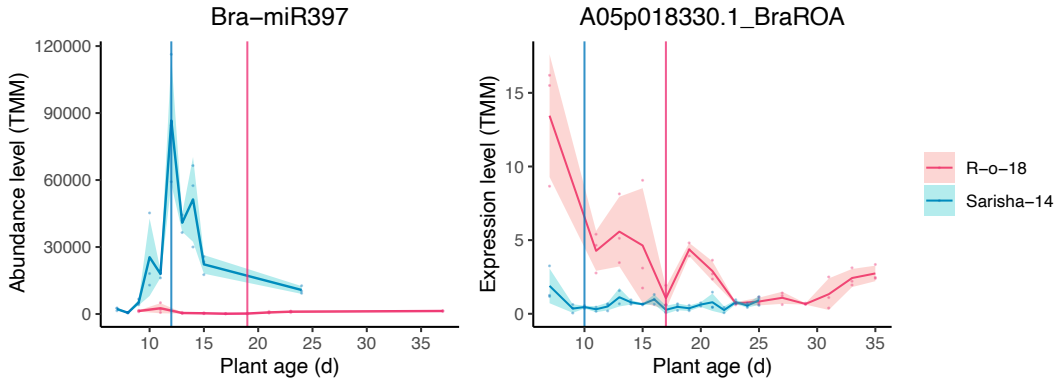


(a) Cluster comparisons

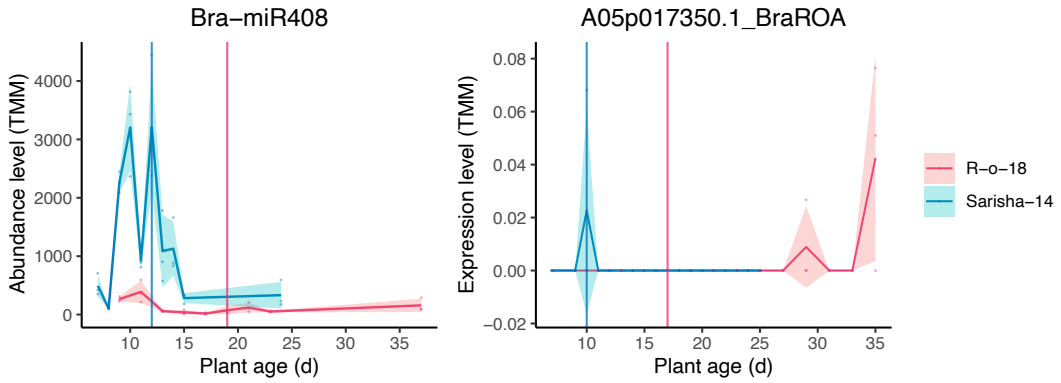


(b) Bra-miR399a and Bra-miR399e abundance

Figure 3.18: **Bra-miR399a and Bra-miR399e display cultivar-specific abundance patterns, and is more highly expressed in Sarisha-14 than R-o-18.**
a) Bra-miR399a and Bra-miR399e are highlighted in their respective clusters from each genotype, showing early downregulation in S14 cluster 2, and peak in abundance in RO cluster 2.
b) abundance of Bra-miR399a and Bra-miR399e throughout development in R-o-18 and Sarisha-14, highlighting differences in abundance patterns and abundance between cultivars. Vertical bars mark the floral transition timepoint for each cultivar.



(a) Bra-miR397



(b) Bra-miR408

Figure 3.19: **Other miRNAs which are highly expressed in Sarisha-14 but not R-o-18 suggest a role in early flowering, but target gene predictions are not consistent with miRNA abundance patterns.**
a) Bra-miR397 abundance and its predicted target gene *A05p018330.1_BraROA* abundance. High Bra-miR397 abundance in Sarisha-14 is consistent with low *A05p018330.1_BraROA* abundance, but this target gene is also downregulated despite low Bra-miR397 abundance in R-o-18.
b) Bra-miR408 is predicted to target *A05p017350.1_BraROA*, an orthologue of *LACCASE3 (LAC3)*. Undetectable target gene expression could be due to high Bra-miR408 abundance and cleavage, although this needs to be validated experimentally.

3.4 DISCUSSION

In this study, I identified miRNAs with potential roles in regulating the floral transition in rapid-cycling *Brassica rapa*. By integrating abundance profiles of miRNAs and their predicted targets across two genotypes with different flowering times, I find evidence for promoter and repressor miRNAs whose temporal abundance profiles align with the floral transition.

3.4.1 *MiRNAs that share regulatory roles in R-o-18 and Sarisha-14*

Several miRNAs displayed conserved abundance dynamics in both Sarisha-14 and R-o-18, which implies similar regulatory roles in *B. rapa* flowering. Some of these miRNAs have been previously characterised as flowering regulators, like miR157, miR319, miR396 and miR159, and their miRNA-target gene expression dynamics in *B. rapa* are consistent with the current literature. However, other miRNAs such as Bra-miR390 and Bra-miR394 have inconsistent abundance profiles compared to its known roles, possibly implying opposing regulatory roles from current literature. Furthermore, for some miRNAs like Bra-miR159c and Bra-miR390, target gene prediction or abundance profiles of the target genes were not consistent with the current literature, revealing a gap for target gene validation for these miRNAs.

Bra-miR157a, the only miR157 copy in B. rapa, shares similar target genes as Bra-miR156a, displaying conservation of miR156/miR157-SPL regulatory pathway

Bra-miR157a, which has only one nucleotide difference from Bra-miR156a, was much more abundantly expressed and rapidly downregulated in both cultivars prior to the floral transition (Fig. 3.6). Its downregulation and consequent upregulation of their shared target *SPL9*, among several other *SPL* orthologues, strongly supports a floral repressor role similar to Bra-miR156. Similarly, the faster decline of Bra-miR157a in Sarisha-14 may contribute to its earlier floral transition.

Bra-miR319 - TCP regulatory module is conserved in B. rapa

miR319 is a well known regulator of TCP transcription factors (Fang et al. 2021), and here its dynamics aligned with this role. Bra-miR319c was highly expressed in both cultivars, with abundance levels reaching nearly 50,000 CPM and 75,000 CPM in Sarisha-14 and R-o-18 respectively during the floral transition (Fig. 3.13). The reciprocal downregulation of its target, a TCP4 orthologue, further supports this repressive interaction. Interestingly, the timing of target repression differed slightly between cultivars: in R-o-18, TCP4 appeared already suppressed well before the floral transition, while in Sarisha-14 repression peaked closer to the transition itself. The variation in

TCP4 abundance, may also reflect circadian regulation, as TCP genes have been linked to clock function (Giraud et al. 2010).

Bra-miR396 - GRF conserved regulatory module may underlie Sarisha-14's rapid growth and subsequently earlier floral transition

Bra-miR396, which targets multiple Growth-Regulating Factor (GRF) genes, was also downregulated approaching the floral transition, resulting in upregulation of *GRF4* (Fig. 3.8) and other *GRF*s. As GRFs are key regulators of cell proliferation and floral organ development (Liebsch and Palatnik 2020), the Bra-miR396-*GRF* module may be more relevant for Sarisha-14's rapid growth rate rather than its timing of floral transition. Bra-miR396 upregulation at BBCH51, coupled with reduced GRF4 abundance, highlights the dynamic role of this regulatory pair during later stages of development.

Bra-miR159c targets GAMYB transcription factors, but only TIE4 and an orthologue of metallo-hydrolase/oxidoreductase-encoding gene display abundance profiles consistent with Bra-miR159c repression

Bra-miR159c was more highly expressed in R-o-18 than Sarisha-14, but in both cultivars it was strongly upregulated during the floral transition, consistent with its role as a floral promoter (Fig. 3.12). Its predicted targets, *TIE4* and a metallo-hydrolase/oxidoreductase-encoding gene, were downregulated across development in both cultivars, supporting the possibility of direct repression. While most studies on miR159 link it to leaf development and promoting flowering time through interactions with the miR156 pathway (J.-W. Wang, Czech, and Weigel 2009), the specific role of the metallo-hydrolase orthologue is unknown, raising the possibility that miR159c may have additional functions beyond canonical flowering regulators.

Bra-miR394a's role in positively regulating the floral transition may be independent of LCR repression

Bra-miR394a also displayed dynamics consistent with repression of its predicted target, *LCR*, a gene involved in stress responses. Bra-miR394a increased approaching the floral transition in both genotypes, though this increase began earlier in R-o-18 than Sarisha-14, and the corresponding target downregulation reflected this timing difference (Fig. 3.14). However, studies in Arabidopsis indicate that mir394a/b double mutants are early flowering, and that this phenotype is independent of LCR regulation (Bernardi et al. 2022). This discrepancy raises questions about whether Bra-miR394a has a conserved repressive function in flowering, or whether its role in *B. rapa* differs from Arabidopsis, potentially linking flowering time with stress adaptation.

Although previous studies have found a miR390-TAS3-ARF4 regulatory module which promotes flowering, Bra-miR390a downregulation suggests a repressive role instead

In Sarisha-14, Bra-miR390a was abundant during the vegetative stage and downregulated at the floral transition, whereas in R-o-18 the decline occurred earlier (Fig. 3.7). Both genotypes clustered in similar Mfuzz patterns and successfully curve-registered, suggesting functional similarity despite the timing shift. I did not detect a *TAS3* orthologue in *B. rapa*, which may explain why this well-characterised miR390-TAS3-ARF4 module in Arabidopsis was not found in *B. rapa* (Fahlgren, Montgomery, et al. 2006). Its predicted target, a putative orthologue of *ALFIN-LIKE 7*, encodes a chromatin-associated PHD protein that reads H3K27me3 marks (Molitor et al. 2014; Liang et al. 2018). Although this target has no established role in flowering, it raises the possibility that miR390a may integrate developmental timing with chromatin regulation.

Novel predicted miRNAs Bra-miRN347 and Bra-miRN367 target genes involved in stress response, possibly linking Sarisha-14's earlier flowering time with environmental adaptation

Finally, novel miRNAs identified by pmiREN, such as Bra-miRN347 and Bra-miRN367, introduce other regulatory pathways which intersect with flowering regulation. Bra-miRN347, predicted to target isopentenyltransferase, *IPT2* (a cytokinin biosynthesis gene), was downregulated in both cultivars (Fig. 3.9), which could suggest a role of cytokinin signalling and stress response in floral regulation. Interestingly, *IPT* is already a candidate gene for crop improvement efforts, by delaying senescence, increasing yield and improved stress response (Nguyen et al. 2021). On the other hand, Bra-miRN367 is predicted to target *VDAC1*, a mitochondrial channel protein. However, *VDAC1* in *B. rapa* was upregulated during the floral transition (Fig. 3.10), which is in contrast to Arabidopsis *VDAC1* which acts as a floral repressor by directly binding FT (J. Xu et al. 2021). Instead, its abundance may relate more to increased metabolic activity during vegetative growth instead of floral repression.

3.4.2 *Cultivar specific miRNAs and their potential regulatory role in Sarisha-14's early flowering*

The set of cultivar-specific miRNAs identified in this study presents exciting opportunities for investigating the early flowering phenotype of Sarisha-14. However, despite strong differential expression between cultivars, clear links to target regulation and flowering time remain unresolved for many candidates.

Bra-miR398a is upregulated 3-fold at the floral transition in Sarisha-14, but downregulated in R-o-18

Bra-miR398a was strongly upregulated in Sarisha-14 compared to R-o-18 (Fig. 3.16), consistent with previous reports connecting miR398 (and the related miR408) to dehy-

dration responses (Trindade et al. 2010). Its predicted target, a putative orthologue of *DREB2B*, a dehydration response gene, did not show abundance changes that corresponded with miRNA differences between genotypes. This suggests either that *DREB2B* is not the true regulatory target in this context or that additional buffering mechanisms obscure direct repression. Interestingly, *DREB2C* overabundance in Arabidopsis delays flowering (Song et al. 2022; Y. Wu et al. 2022), raising the possibility of a conserved regulatory connection between DREB-related genes and flowering. Nonetheless, the exact functional contribution of Bra-miR398a to flowering time regulation in *B. rapa* remains unclear.

Bra-miR169 downregulation in Sarisha-14 is not consistent with its known role in promoting flowering under stress

In Arabidopsis, miR169d regulates flowering under stress by targeting Nuclear Factor Y subunit A2 (NF-YA2), and higher abundance of miR169 has been linked to stress-induced early flowering (M. Y. Xu et al. 2013). In Sarisha-14, however, I observed downregulation of Bra-miR169 rather than upregulation (Fig. 3.17), which contradicts this model, and implies that Bra-miR169 represses flowering rather than promotes flowering (under stress). Moreover, no predicted targets of NF-YA2 orthologues were identified in our dataset, limiting the ability to directly connect these dynamics to known flowering-time pathways.

PHOSPHATE 2 (PHO2) is not a predicted target of Bra-miR399, which suggests alternate regulation of flowering by this miRNA in B. rapa

Bra-miR399 displayed strong cultivar-specific differences, being more highly expressed in Sarisha-14 during the vegetative phase (Fig. 3.18). Literature suggests that miR399 promotes early flowering under ambient temperature (23 °C) but not under cooler conditions (16 °C) in Arabidopsis (W. Kim et al. 2011). Given Sarisha-14's breeding history in warmer climates, elevated Bra-miR399 levels may contribute to its early flowering phenotype. However, its abundance declined approaching the floral transition, and none of the predicted targets displayed derepression consistent with miRNA-target dynamics. Thus, while Bra-miR399 is an attractive candidate for genotype-specific flowering regulation, its mode of action in *B. rapa* is yet to be resolved.

Bra-miR397 and Bra-miR408 are highly expressed in Sarisha-14 and both target cell wall biosynthesis genes

Bra-miR397 and Bra-miR408 are highly expressed in Sarisha-14 relative to R-o-18 (Fig. 3.19), and both are known to target laccase genes involved in cell wall development (Huang et al. 2020). In Sarisha-14, Bra-miR397 high abundance level was consistent with *LAC2* repression, but the same target was also downregulated in R-o-18 despite low Bra-miR397 abundance, suggesting that other regulatory mechanisms may

repress *LAC2* in R-o-18 (Fig. 3.19). Bra-miR408, which is associated with dehydration responses (Gao et al. 2022; Trindade et al. 2010), was also highly expressed in Sarisha-14, but no reciprocal abundance were detected in its predicted target gene expression. The *LAC3* orthologue transcript was undetectable, which could be due to extremely efficient mRNA cleavage/degradation by Bra-miR408 - however this is speculative and needs further target validation.

Overall, while these miRNAs with cultivar-specific abundance patterns are promising candidates which could contribute to Sarisha-14's early flowering phenotype, their regulatory mechanisms are not yet clear. Most of these miRNAs regulate target genes involved in stress responses or cell wall development, with unclear roles in flowering regulation. The lack of clear target gene responses emphasizes the need for experimental validation beyond gene expression profiling. These findings suggest that these miRNAs could contribute to flowering time diversity in *B. rapa*, either through direct regulation of flowering pathways or by integrating environmental and developmental signals in a genotype-dependent manner.

3.4.3 *Future investigation on the functions of candidate miRNAs in regulating flowering time in Sarisha-14*

This chapter compared the abundance dynamics of candidate miRNAs to assess whether they have the potential to influence flowering time in rapid-cycling *B. rapa*. Although I was able to visualize the correlation between miRNA abundance and their predicted target gene expression to elucidate the potential functions of these miRNAs, experimental validation is still needed to understand how these miRNAs contribute to earlier flowering time in Sarisha-14. Furthermore, miRNA-target predictions performed with TargetFinder is mainly based on sequence complementarity between the miRNA and target site, which could lead to potential false positives, but is very useful for exploratory analysis in this chapter. TargetFinder is also limited by the reference annotation of the R-o-18 genome, for example, based off current literature the miR390-*TAS3* regulatory module would be expected to be conserved in *B. rapa*, however, since there is no annotation for a *TAS3* orthologue in the R-o-18 reference genome, I didn't observe this conserved regulation.

Experimental validation of miRNA targets

RNA ligase-mediated 5' amplification of cDNA ends (RLM-RACE) (César Llave et al. 2011; Cesar Llave et al. 2002) is an established method to experimentally validate miRNA cleavage sites on its target genes, and would be useful for confirming the target transcripts of the miRNAs described in this chapter. A more high thoroughput method

would be degradome sequencing (also known as Parallel Analysis of RNA Ends, PARE), which allows target validation for all miRNAs within a single RNA sample.

Besides target site validation, functional studies are also necessary to assess the biological impact of these miRNAs and its downstream effects when their abundance or target site identification is altered. For example, generating transgenic R-o-18 lines with miRNA overabundance or target mimicry mutants, would provide insight on how manipulating miRNA abundance or inhibiting its repressive function could influence development. We could also take advantage of the R-o-18 TILLING (Targeting Induced Local Lesions In Genomes) population (Stephenson et al. 2010), by identifying mutant alleles of miRNA target genes or miRNA encoding loci which could lead to a change in flowering phenotype, thus giving us a better understanding of the biological functions of these miRNAs in *B. rapa* development.

Conservation of miRNA-target regulatory modules in other Brassicas

Another promising direction would be to examine whether the candidate miRNA regulatory modules identified in this chapter are conserved across other *Brassica* species. Comparative analysis of orthologous miRNA–target pairs in related cultivars, such as *B. oleracea* DH1012 rapid-cycling cultivar, or spring type oilseed rape (*B. napus*), could reveal whether these miRNA-target regulatory modules are conserved flowering-time regulators or lineage-specific adaptations. Sequence conservation would suggest a broader role in *Brassica* flowering regulation, while divergence may highlight unique regulation of flowering in Sarisha-14. This could also be paired with available timeseries profiling data for DH1012 (Woodhouse et al. 2021) and various *B. napus* crop types from Calderwood, Lloyd, et al. (2021), to assess the abundance patterns of the pri-miRNA precursors and their predicted target genes. Thus, integrating comparative genomics and abundance profiling across species will be important to understand the evolutionary and mechanistic basis of miRNA regulation of flowering in various Brassica crop types.

3.5 CONCLUSION

The comparative analysis of miRNA dynamics in Sarisha-14 and R-o-18 reveals both conserved and cultivar-specific regulatory modules. Shared floral promoter and repressor miRNAs, including Bra-miR159c, Bra-miR319c, Bra-miR390a, Bra-miR394a, Bra-miR157a, and Bra-miR396a, displayed broadly consistent abundance patterns across cultivars, which aligns with their roles as core miRNA regulators of flowering development. Their reciprocal abundance patterns against target genes such as *SPL*s (floral promoters), *GRF*s (growth regulators) , and *TCP* transcription factors suggest conserved regulatory modules in flowering and growth. Meanwhile, other miRNA-target

modules (Bra-miR390 - *AL7*, Bra-miRN347-*IPT*) suggest other flowering regulatory pathways which involve chromatin remodelling and cytokinin signalling.

In contrast, miRNAs such as Bra-miR398a, Bra-miR169, Bra-miR399, Bra-miR397, and Bra-miR408 displayed significant differences in abundance abundance or timing between Sarisha-14 and R-o-18, but their target gene dynamics did not reflect cultivar-specific abundance patterns reciprocal to the miRNAs. Several of these miRNAs are associated with stress response in Arabidopsis, which raises the possibility that Sarisha-14's rapid flowering is due to positive selection for stress-responsive miRNA networks, particularly since it was bred in warmer climates. These findings suggest that Sarisha-14's early flowering phenotype may come from a combination of conserved core miRNA-target regulatory modules that ensure proper developmental timing, and cultivar-specific rewiring of stress-related miRNA regulation that could fine tune the timing of the floral transition in specific environmental conditions.

4

INVESTIGATING THE AGEING PATHWAY IN RAPID-CYCLING *B. RAPA*

4.1 INTRODUCTION

This thesis chapter investigates the abundance of two miRNAs, miR156 and miR172, in the rapid-cycling *Brassica rapa* cultivars R-o-18 and Sarisha-14, which initiate flowering at different ages. MiR156 and miR172 act antagonistically as key regulators of the vegetative-to-reproductive transition as part of the ageing pathway. Since Sarisha-14 and R-o-18 don't require vernalization to progress to the floral transition, the ageing pathway could play a central role in regulating flowering time in these rapid-cycling cultivars.

4.1.1 *The role of miR156 and miR172 in regulating flowering time*

Teotia and Tang (2015) described miR156 and miR172 as developmental timers for the ageing pathway, with miR156 acting as a "count-down" timer to the vegetative-to-reproductive transition, in contrast to miR172 which acts as a "count-up" timer. During plant development, miR156 is highly expressed in the vegetative seedling stage, then its abundance decreases, while miR172 exhibits the opposite pattern, showing increasing abundance with plant age (G. Wu, Park, et al. 2009). This antagonistic abundance pattern is observed across diverse plant lineages, including the model species *Arabidopsis thaliana*, monocotyledonous crops like rice and maize, as well as perennial trees like poplar and *Acacia* (G. Wu, Park, et al. 2009; Lauter et al. 2005; Xie, C. Wu, and Xiong 2006; J.-W. Wang, Park, et al. 2011). Despite this highly conserved regulatory pathway, the mechanistic details of miR156 and miR172 abundance has not yet been characterized in rapid-cycling *B. rapa*, particularly how they can contribute to the different flowering times within the same species.

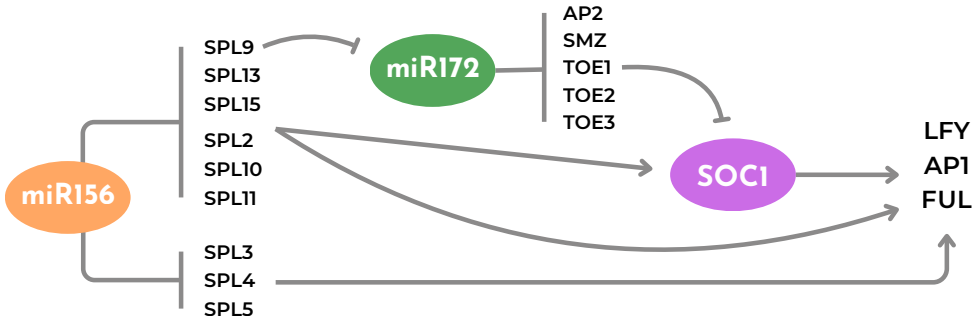


Figure 4.1: **An overview of miR156, miR172 and their target genes, and the regulatory network it forms to control flowering.** *LEAFY (LFY), APETALA1 (AP1)* and *FRUITFULL (FUL)* are a group of floral genes that have individual roles in promoting floral organ identity. During the vegetative stage, these genes are repressed by *APETALA2 (AP2)*-like gene family, comprising of *AP2, SCHLAFMÜTZE (SMZ)*, and *TARGET OF EAT (TOE1, TOE2, TOE3)*. At the same time, miR156 maintains the vegetative state by repressing members of the *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* gene family, which promote flowering. As the plant ages, miR156 abundance decreases, leading to the upregulation of its *SPL* targets, where *SPL9* directly activates miR172 abundance, *SPL3, SPL4 and SPL5* promote floral identity genes *LFY, AP1* and *FUL*, and developmental transition genes *SPL9, SPL13, SPL15, SPL2, SPL10* and *SPL11*, which promote expression of floral integrator gene *SUPPRESSOR OF CONSTANS 1 (SOC1)* and are necessary for the floral transition. The activated miR172 targets the *AP2*-like gene family, repressing them and lifting its repression of *SOC1* and floral identity genes *LFY, AP1* and *FUL*. The balance between miR156 and miR172 abundance dynamics is therefore the core of age-dependent flowering pathway. This diagram is adapted from M. Xu, T. Hu, Zhao, et al. (2016).

miR156 delays flowering by maintaining the juvenile stage

During the vegetative stage, miR156 maintains the juvenile vegetative stage by repressing transcription factors from the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) gene family (Fig. 4.1), which promote flowering (G. Wu, Park, et al. 2009). MiR156 abundance declines as the plant grows older, leading to SPL upregulation, thus promoting the juvenile-to-adult and vegetative-to-reproductive transitions (G. Wu and Poethig 2006; J.-W. Wang, Czech, and Weigel 2009). In the model plant *Arabidopsis thaliana*, the SPL gene family has 16 members (*SPL1-SPL16*), with 10 members being targets of miR156 (Xing et al. 2010; M. Xu, T. Hu, Zhao, et al. 2016). Among these SPLs, *SPL9, SPL13*, and *SPL15* are particularly important for the floral transition under long days, with mutants flowering 4-5 days later than wild type. Furthermore, miR156 repression of SPLs is also linked to miR172 abundance - loss-of-function *spl* mutants have reduced expression of precursor pri-miR172 and mature miR172 (M. Xu, T. Hu, Zhao, et al. 2016).

While miR156-mediated repression of *SPL* genes is well characterized in Arabidopsis, whether this regulatory module performs the same function of maintaining juvenility in Brassica is not yet established. Following the whole-genome triplication (WGT) event in *B. rapa*, many duplicated protein-coding genes were lost over time (fractionation), yet miRNA loci are preferentially retained and their sequences were

highly conserved (Sun et al. 2015). Despite this conservation of miRNA sequences, the same study shows that target genes often diverge in their miRNA binding sites, leading to expression divergence between gene copies. This raises the possibility that in *B. rapa*, duplicated *SPL* paralogues may respond differently to miR156-mediated repression, potentially changing the dynamics of the miR156–SPL module after WGT. Identifying conserved miR156 binding sites in *B. rapa SPL* transcripts, combined with timeseries expression profiling of these genes, can clarify whether the miR156-SPL regulatory module is conserved and functionally relevant to flowering time in *B. rapa*.

miR172 promotes flowering by targeting floral repressors

In contrast to miR156, which maintains juvenility and delays the transition to flowering, miR172 acts as a promoter of flowering. MiR172 targets members of the APETALA2 (AP2)-like transcription factor family, including *AP2, TARGET OF EAT1 (TOE1, TOE2, TOE3), SCHLAFMÜTZE (SMZ)*, and *SCHNARCHZAPFEN (SNZ)* (Fig. 4.1). These AP2-like proteins act as floral repressors, delaying both the floral transition and the acquisition of floral meristem identity (Aukerman and Sakai 2003; Schmid et al. 2003; Jung et al. 2007). By downregulating these floral repressors, miR172 facilitates the subsequent activation of downstream flowering pathways, most notably through the floral integrator FLOWERING LOCUS T (FT) (Mathieu et al. 2009; Jung et al. 2007).

The temporal abundance of miR172 is inverse of miR156's abundance: while miR156 declines with developmental age, miR172 abundance gradually increases. With miR172 upregulation, this leads to the repression of *AP2*-like floral repressors, thus promoting flowering. This antagonistic relationship between miR156 and miR172 is reinforced by evidence that *SPL9*, a miR156 target gene, transcriptionally activates miR172b abundance by directly binding upstream of the pri-miR172b locus. This suggests a sequential cascade that links the vegetative juvenile-to-adult transition with the promotion of flowering, and forms the basis of our knowledge of the ageing pathway (G. Wu, Park, et al. 2009).

miR156 and miR172 act together as a developmental switch: high miR156 maintains vegetative growth, while rising miR172 levels relieve repression of FT and promote the floral transition. Since this regulatory network is highly conserved across species, it could underlie the differences in flowering time between Sarisha-14 and R-o-18, but still remains to be determined.

Figure 4.2: *Illustration of temporal changes in miR156 and miR172 abundance and their target gene families during plant development.* MiR156 levels decline with age, permitting the accumulation of SPL transcription factors, whereas miR172 levels increase, leading to repression of AP2-like floral repressors. Together, these antagonistic abundance dynamics form a developmental switch that controls the timing of the floral transition.

4.1.2 *miR156 and miR172 abundance in rapid-cycling Brassica rapa*

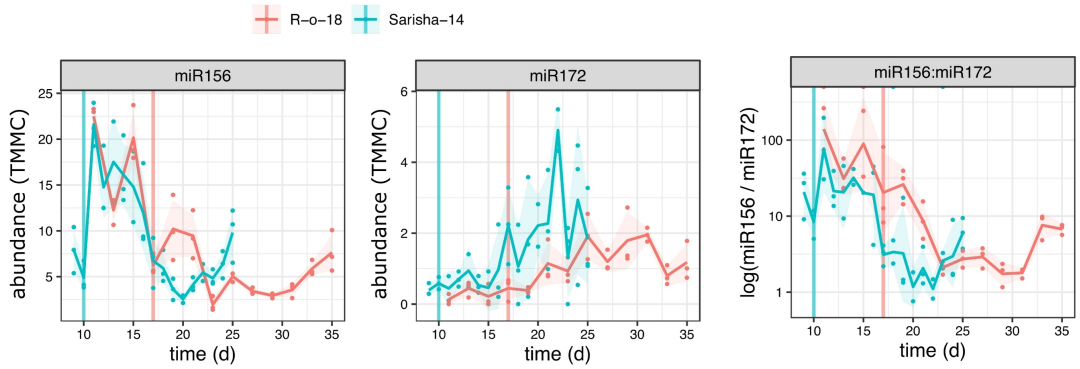


Figure 4.3: **Expression of pri-miR156 and pri-miR172 precursors in R-o-18 (red) and Sarisha-14 (blue) shoot apex during development.** This figure is adapted from Calderwood, Hepworth, et al. (2021). Calderwood et. al. reported that pri-miR156 expression decreases during development, accompanied by pri-miR172 upregulation. The log ratio of pri-miR156 to pri-miR172 expression are also similar in magnitude at the floral transition, which forms the basis of our hypothesis that miR156/miR172 expression ratio during shoot apex development determines the timing of the floral transition in *B. rapa*. Vertical lines mark the first day of the floral transition in the respective cultivars.

Calderwood, Hepworth, et al. (2021) compared gene expression between the two rapid-cycling *B. rapa* cultivars, R-o-18 and Sarisha-14, and showed that the different timing of the floral transition in Sarisha-14 is linked to shifts in temporal expression patterns of flowering-related genes. Among the flowering time genes, Calderwood, Hepworth,

et al. (2021) reported that the expression of pri-miR156 precursor decreased leading to the floral transition, while pri-miR172 expression increased (Fig. 4.3). Additionally, they found that the ratio of pri-miR156 to pri-miR172 expression level was similar at the floral transition timepoint, even though they occur 7 days apart (Fig. 4.3). This suggests that the ageing pathway is a core regulatory pathway for rapid-cycling *B. rapa*, and more specifically, that Sarisha-14's earlier floral transition could be achieved via an earlier shift in timing of miR156 and miR172 differential expression, thus achieving a specific ratio of miR156 to miR172 abundance that coincides with the floral transition.

While this provided our first insight of how these ageing-related miRNAs behave in rapid-cycling *B. rapa*, precursor expression does not always reflect the abundance of mature miRNAs, which are the actual regulators of target gene repression. Thus, the dynamics of mature miR156 and miR172 activity in these cultivars still needs to be confirmed. Moreover, Calderwood, Hepworth, et al. (2021) only examined the expression of an orthologue of *SCHLAFMÜTZE (SMZ)*, one of the AP2-like transcription factors targeted by miR172, but not any other targets of miR156 and miR172. Since flowering time regulation by the ageing pathway not only depends on miR156 and miR172's abundance, but also on the expression of their target genes that act within the flowering time pathway (Fig. 4.1), we can further elucidate the impact of these miRNAs by investigating their target genes.

In Arabidopsis, the balance between miR156 and miR172 provides a timing mechanism for the juvenile-to-adult transition and the initiation of flowering. In rapid-cycling *B. rapa*, the key questions remain: do R-o-18 and Sarisha-14 show similar temporal abundance dynamics of miR156 and miR172 during the floral transition? Does the ratio of miR156 to miR172 abundance reflect the timing of their floral transition? And most importantly, is the miR156–SPL / miR172–AP2 regulatory module conserved in *B. rapa*, or has genome triplication introduced divergence that modifies its role, possibly influencing Sarisha-14's earlier flowering time?

4.1.3 *Hypothesis*

In this chapter, the hypothesis I am testing is that there exists a specific ratio of miR156 and miR172 abundance that defines the timing of the floral transition in rapid-cycling *Brassica rapa*. This critical ratio could be the same in R-o-18 and Sarisha-14, which suggests that Sarisha-14 is able to accelerate initiating the floral transition due to achieving the specific miR156/miR172 ratio earlier.

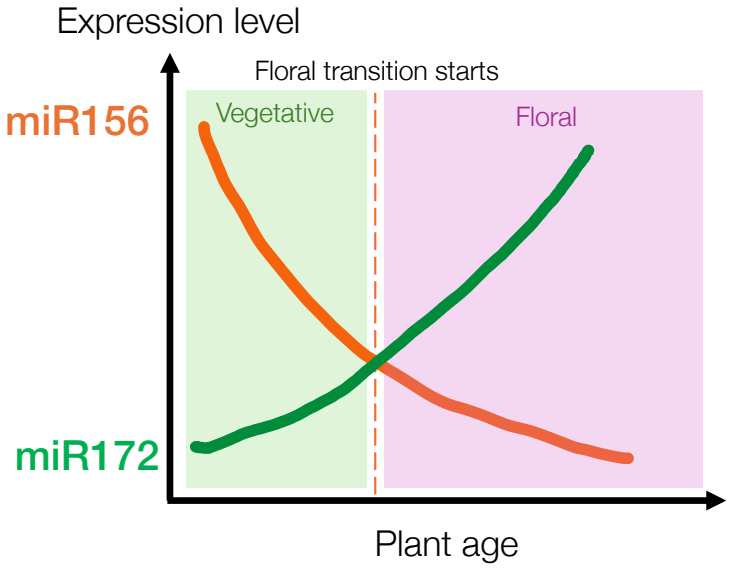


Figure 4.4: **A visualisation of the ratio hypothesis.** We hypothesise that miR156 downregulation (orange), accompanied by miR172 upregulation (green), converges to an abundance ratio which determines the timing of the floral transition (vertical dashed red line). The actual ratio may not actually be 1-to-1 as depicted in this diagram, which is merely for illustration purposes.

4.1.4 *Objectives*

The objectives of this chapter are to:

1. Characterize the timeseries abundance profiles of miR156 and miR172 in R-o-18 and Sarisha-14, and quantify their abundance ratio across developmental timepoints

2. Test whether predicted target genes of miR156 and miR172 show downregulation consistent with increased miRNA abundance, providing evidence for functional repression and conserved regulation by the ageing pathway.

4.2 METHODS

4.2.1 *Plant growth conditions, sampling, and small RNA sequencing*

Brassica rapa cultivars Sarisha-14 and R-o-18 plants were sown in 24 cell trays in cereals mix (40% medium grade peat, 40% sterilised soil, 20% horticultural grit, 1.3 kg/m^3 PG mix 14-16-18 + Te base fertiliser, 1 kg/m^3 Osmocote Mini 16-8-11 2 mg + Te 0.02% B, wetting agent, 3 kg/m^3 maglime and 300 g/m^3 Exemptor). Material was grown in a Conviron MTPS 144 controlled environment room with Valoya NS1 LED lighting (250 µmol m^{-2} s^{-1}) 18 °C day/15 °C night, 70% relative humidity with a 16 hour day (lights on from 00.30hrs until 16.30hrs).

Sampling of Sarisha-14 and R-o-18 leaf and apex was performed in the 10th hour of the day (ZT10). Sampling was performed on Sarisha-14 every day from 7 days after sowing until 15 days after sowing, with a final sampling timepoint at day 24 which was when the floral buds were visible (developmental stage BBCH51 (Meier et al. 2009))). R-o-18 was sampled every two days from day 9 until day 23, and also at BBCH51 on day 37. These timepoints were chosen to capture small RNA abundance prior to and during the floral transition stage, until the plants were considered past the floral transition. A total of 9 timepoints were collected from R-o-18, while 10 timepoints were collected for Sarisha-14. For each timepoint, 4 replicates were sampled, with 4 pooled apices/leaf tissue per replicate. Only 3 out of 4 replicates were processed for sequencing.

To identify the timepoint when the floral transition occured, 4 plants were inspected at the same time of day as the sampling, by dissection under a light microscope. The floral transition timepoint was marked when at least 3 out of 4 plants showed domed meristems with round floral primordia on the flanks of the apex (Kinoshita et al. 2020), compared to flat vegetative meristems flanked by triangular leaf primordia.

During tissue sampling, apex and leaf tissue were dissected and pooled on ice inside the growth chamber, then flash frozen in liquid nitrogen once all sampled were pooled for each replicate. Apex tissue were pooled in RNAse-free 1.7 mL microtubes, while leaf tissue was stored in aluminium packets. These samples were then stored in -70 °C storage, until total RNA extraction was performed at a later date.

Apex tissue samples were ground in the original microtube using a micropestle attachment on a drill while immersed in liquid nitrogen until a fine powder was formed. Leaf tissue was ground in a pestle and mortar while liquid nitrogen was added to keep the tissue frozen. Total RNA extraction was performed following the manufacturers standard protocol from Zymo Research's Direct-zol Microprep Total RNA extraction kit, supplemented with TRI Reagent for tissue lysis and DNAse treatment (https://zymoresearch.eu/products/direct-zol-rna-microprep-kits).

The total RNA samples were sent for sequencing with RealSeq Biosciences, California, United States of America. RNA libraries were constructed using Singular Genomics G4 F3 Sequencing Kit and sequencing was performed on the Singular Genomics G4 sequencer, at an average depth of 30 million single-ended reads.

4.2.2 *MiRNA identification and abundance quantification*

Raw small RNA sequencing reads (75 bp single-end) were quality-checked with FastQC (ver. 0.11.9, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapters trimmed using Cutadapt (ver. 2.10, Martin (2011)). After size filtering for reads between 15-30 nt, reads were aligned to the *Brassica rapa* R-o-18 reference genome (EnsemblPlants release 59) with Bowtie (ver. 1.2.2,Langmead et al. (2009)). Reads mapping perfectly (no mismatch) to ribosomal, transfer, small nuclear, or small nucleolar RNAs (sequences downloaded from Rfam release 14.10, Kalvari et al. (2021)) and to transposable elements (TrepDB release 19) were discarded. The remaining reads were aligned to a deduplicated set of mature *B. rapa* miRNA sequences from the Plant microRNA Encyclopedia (PmiREN, release 2.0, Guo, Kuang, and X. Yang (2025)) to identify known miRNAs, only keeping miRNAs perfectly mapping to canonical *B. rapa* miRNA sequences with no mismatch. Raw miRNA read counts for each sample were extracted using Samtools-idxstats (ver. 1.10, Danecek et al. (2021)). Read count normalization or transformation was performed depending on the requirements of the analysis, and are described below. For miRNA abundance plots, library size normalised counts (CPM) were scaled with Trimmed Mean of M factors (TMM) to allow more accurate sample-to-sample gene expression abundance comparisons (Smid et al. 2018).

4.2.3 *MiRNA target gene prediction*

MiRNA target gene prediction was performed locally with TargetFinder (https://github.com/carringtonlab/TargetFinder) (2024). *B. rapa* miRNA sequences were downloaded from PmiREN (release 2.0) (Guo, Kuang, and X. Yang 2025), and used as small RNA sequence input for TargetFinder. *B. rapa* R-o-18 reference coding DNA sequences (EnsemblPlants release 59) were used as the target database input. The prediction score cutoff was kept at the default value (-c 4).

Potential gene orthologues were referenced from EnsemblPlants: https://ftp.ebi.ac.uk/ensemblgenomes/pub/plants /release-59/tsv/ensembl-compara/homologies/brassica_rapa_ro18/

4.2.4 *Target gene mRNA expression quantification and normalisation*

For gene expression profiling from Sarisha-14 and R-o-18 timeseries, I used sequencing data generated from (2021). The Illumina sequencing reads are available from the NCBI Sequence Read Archive, project ID PRJNA593493. The raw sequencing reads were quality checked, adapter trimmmed and aligned against the *B. rapa* R-o-18 reference genome (EnsemblPlants release 59), then gene expression counts was calculated using stringtie. Read alignment and gene expression quantification was performed using an automated script on the High Performance Computing cluster, provided by Dr Hugh Woolfenden (script available on gitlab: https://git.nbi.ac.uk/morris-group/bravo-scripts/-/tree/master/alignment_scripts). For gene expression abundance plots, library size normalised counts (CPM) were scaled with Trimmed Mean of M factors (TMM) to allow more accurate sample-to-sample gene expression abundance comparisons.

4.3 RESULTS

4.3.1 *Timeseries abundance profiling of Bra-miR156 and Bra-miR172*

4.3.1.1 *Identification of B. rapa miR156 and miR172 sequences*

To characterize the abundance of Bra-miR156 and Bra-miR172 amongst other miRNAs during *B. rapa* development, small RNA-sequencing was performed on Sarisha-14 and R-o-18 shoot apex and leaf tissue from the vegetative stages, floral transition and floral stages. The first day of the floral transition was marked when 3 out of 4 sampled apices exhibited signs of the apex doming and undergoing the floral transition (Kinoshita et al. 2020). For Sarisha-14, this happened 12 days after sowing, while in R-o-18, the first day of the floral transition was on day 19, a week later (Fig. 4.8).

From the Plant MiRNA Encyclopedia (PmiREN) database (Guo, Kuang, and X. Yang 2025), there are 20 annotated members of the Bra-miR156 family (Bra-miR156a-t) and 10 members of the Bra-miR172 family (Bra-miR172a-j). Because highly conserved plant miRNA families like miR156 and miR172 often have multiple miRNA encoding loci (Sun et al. 2015), most mature miRNA sequences within the same family tend to be identical although they are expressed from different loci. For example, the mature miRNA sequence encoded by Bra-miR156a is shared with 13 other Bra-miR156 members (Fig. 4.5), while Bra-miR172a mature miRNA sequence is identical with three other members Bra-miR172b, Bra-miR172e and Bra-miR172g (Fig. 4.6).

To avoid ambiguous read assignment during miRNA identification with `bowtie`, I first performed sequence deduplication of miRNA members with identical mature sequences, collapsing them to a single member which appeared first alphabetically. For example, all 13 Bra-miR156 members which share the same mature miRNA sequence are collapsed under Bra-miR156a. This approach ensured accurate mapping of small RNA sequencing reads for miRNA identification, while minimizing the confounding effects of `bowtie` multi-mapping among redundant mature miRNA sequences.

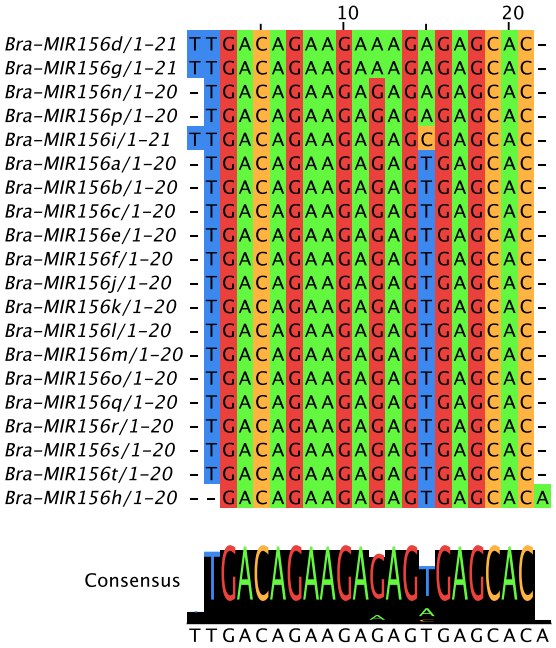


Figure 4.5: **Multiple sequence alignment of all Bra-miR156 members.** 20 reference *B. rapa* miR156 mature miRNA sequences were downloaded from PmiREN, and multiple sequence alignment was performed using ClustalOmega. These sequences are displayed in 5′-3′ alignment. Bra-miR156 members with similar sequences were aligned together, showing that 13 out of 20 members had the same sequence as Bra-miR156a. Bra-miR156h and Bra-miR156i are the only members with unique sequences, while Bra-miR156d and Bra-miR156g shared the same 21nt sequence.

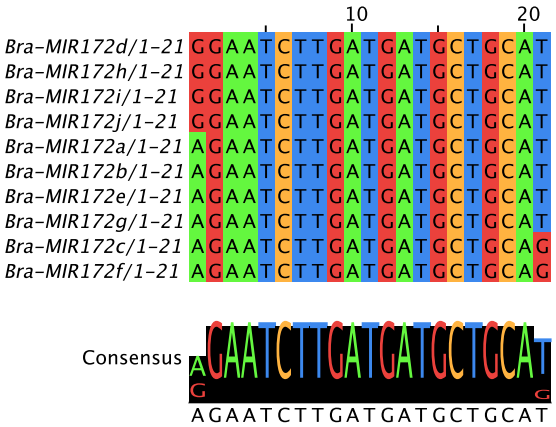


Figure 4.6: **Multiple sequence alignment of all Bra-miR172 members.** 10 reference *B. rapa* miR172 mature miRNA sequences were downloaded from PmiREN, and multiple sequence alignment was performed using ClustalOmega. These sequences are displayed in 5′-3′ alignment. There are three groups of Bra-miR172 members sharing the same mature miRNA sequence. The consensus diagram shows highly conserved nucleotides from position 2-20, with different Bra-miR172 members having different nucleotides in either the first or last position.

4.3.1.2 *abundance profiles of individual Bra-miR156 and Bra-miR172 members*

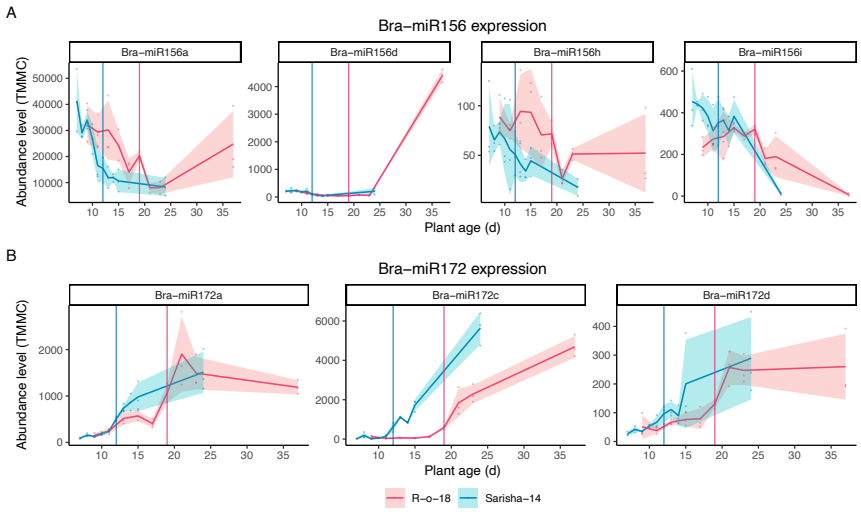


Figure 4.7: **Timeseries abundance pattern of individual miR156 and miR172 copies during shoot apex development.** Each plot illustrates the average abundance of individual miRNA members across development in R-o-18 (red) and Sarisha-14 (blue). Each point in the plot corresponds to the abundance level in a replicate sample, while the shaded ribbon is the standard deviation of abundance. The first day of the floral transition for either cultivar is denoted with vertical lines (Sarisha-14 on day 12; R-o-18 on day 19). abundance is shown in counts scaled by Trimmed Mean of M factors (TMMC)
A) All Bra-miR156 members (except Bra-miR156d) are highly expressed in the vegetative stage, followed by downregulation as the age progresses. Bra-miR156a is the most highly expressed miR156 member, and its abundance rapids declines in Sarisha-14 while in R-o-18 the decline is more gradual.
B) All Bra-miR172 members are lowly expressed during the vegetative stage, and its rise in abundance coincides with the floral transition timepoint. Of all the members, Bra-miR172c is the most highly expressed by the final sampling timepoint (developmental stage BBCH51).

In both Sarisha-14 and R-o-18, Bra-miR156a is the most highly expressed member of the miR156 family, with an average abundance of over 32000 TMMC in R-o-18, and over 41000 TMMC in Sarisha-14 at the first vegetative timepoint (Fig. 4.7). During the vegetative stage, all the Bra-miR156 members (except Bra-miR156d) are expressed at high levels, but their abundance declines as development progresses. The rate of abundance decline differs between cultivars: in Sarisha-14, Bra-miR156a abundance drops sharply before the floral transition at day 12, while in R-o-18 the decrease is more gradual until the floral transition at day 19. This difference suggests that the earlier repression of Bra-miR156 in Sarisha-14 may facilitate its earlier entry into the reproductive phase.

In contrast, all three Bra-miR172 members show the opposite trend. They are expressed at very low levels during the vegetative stage, then increase around the floral transition (Fig. 4.7). The timing of Bra-miR172 upregulation also illustrates that Bra-miR172 members are upregulated earlier in Sarisha-14, and their increase coincides with the earlier floral transition. Bra-miR172c is the most abundant member by the

final sampling point (BBCH51), which suggests that it plays a more dominant role in promoting flowering. The timing of miR172 upregulation coincides with the floral transition, reinforcing the idea that rising miR172 levels act as a developmental signal that the plant is ready to flower.

Curiously, in R-o-18, Bra-miR156a and Bra-miR156d show a peak in abundance in the final timepoint at day 37, coinciding with developmental stage BBCH51 (when floral buds are finally visible) (Fig. 4.7). This isn't reflected in the other members of Bra-miR156 or Bra-miR172 in R-o-18, nor in Sarisha-14 Bra-miR156 members, which rules out potential sequencing effects where deeper sequencing depth could result in higher raw read counts. Why this happened is yet to be determined.

Thus, the abundance patterns of Bra-miR156 and Bra-miR172 members in both cultivars follows the expected antagonistic pattern where Bra-miR156 decline is accompanied by Bra-miR172 increase. What distinguishes Sarisha-14 from R-o-18 is the sharper decline of Bra-miR156a, which may reduce repression of its *SPL* targets earlier and permit earlier floral transition. Similarly, the earlier rise of Bra-miR172, particularly Bra-miR172c, coincides with the earlier initiation of the floral transition in Sarisha-14. These observations suggest that the balance between declining miR156 activity and increasing miR172 abundance is conserved in *B. rapa*, but that the relative timing of these abundance changes differs between cultivars and may underlie their differences in flowering time. Further analysis of *SPL* and *AP2* target genes expression will help clarify how these two miRNA families jointly regulate the floral transition in rapid-cycling *B. rapa*.

4.3.1.3 *Total miR156 and miR172 abundance abundances are vastly different*

The abundance profiles of the sum of Bra-miR156 members and Bra-miR172 members exhibit a clear inverse pattern, in line with previous findings in Arabidopsis (G. Wu, Park, et al. 2009) (Fig. 4.8). However, we also observe a large difference in the total abundance levels between Bra-miR156 and Bra-miR172 in *B. rapa*. Figure 4.8A shows that Bra-miR156 is expressed at much higher levels than Bra-miR172 in both cultivars, and it remains more highly expressed than Bra-miR172 even at the onset of the floral transition, after Bra-miR156 already declined to almost half its initial vegetative abundance level. On the other hand, Bra-miR172 showed a modest but consistent increase , increasing almost three-fold by the floral transition (Fig. 4.8).

This suggests that the floral transition in *B. rapa* is associated with an increase in Bra-miR172 abundance, even while Bra-miR156 remains relatively abundant. In both cultivars, Bra-miR172 increases around the floral transition despite Bra-miR156 not being fully repressed, implying that the initiation of the floral transition may depend on Bra-miR172 increasing to a sufficient abundance level rather than on Bra-miR156 decreasing to very low levels.

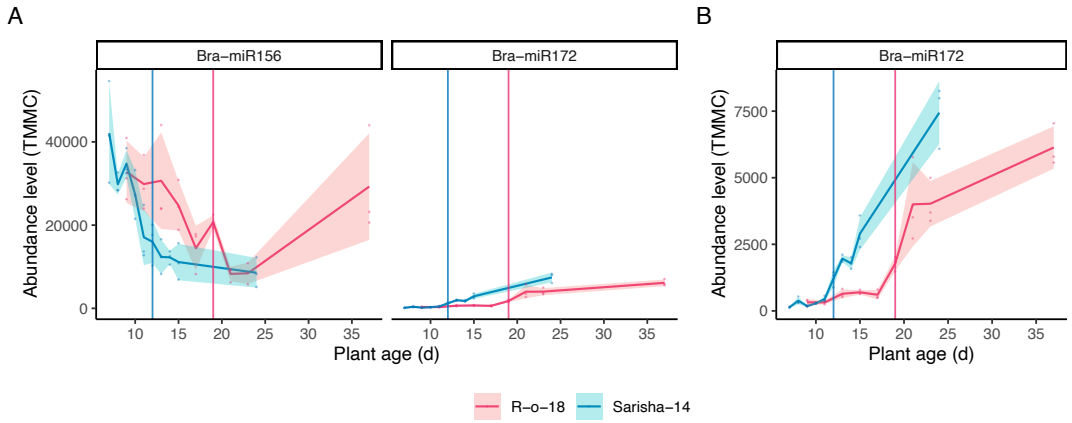


Figure 4.8: **Total abundance of Bra-miR156 decline is accompanied by Bra-miR172 upregulation, displaying antagonistic abundance dynamics.** The sum of abundance of all Bra-miR156 members and Bra-miR172 members was calculated for each replicate per timepoint. The centre line in each plot shows the average abundance of the sum of Bra-miR156 and Bra-miR172 members across development in R-o-18 (red) and Sarisha-14 (blue). Each point in the plot corresponds to the total miRNA abundance level in a replicate sample, while the shaded ribbon is the standard deviation of abundance. The first day of the floral transition for either cultivar is denoted with vertical lines (Sarisha-14 on day 12; R-o-18 on day 19). abundance is shown in counts scaled by Trimmed Mean of M factors (TMMC)
A) Total abundance for Bra-miR156 and Bra-miR172, with the same y-axis scale to demonstrate the differences in abundance level. It is clear that although Bra-miR156 abundance decreases to around half its initial vegetative abundance level at the floral transition timepoint, it is still more highly expressed than Bra-miR172 overall.
B) A zoomed-in plot for total Bra-miR172 abundance on a smaller scale, which shows that Bra-miR172 rises almost 3-fold, coinciding with the floral transition in both cultivars.

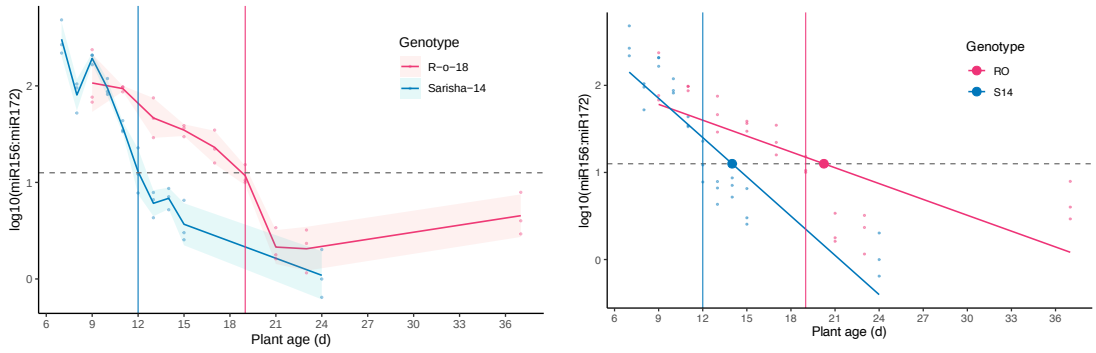
4.3.1.4 *Ratio of miR156/miR172 abundance coincides with the initiation of the floral transition*

In Figure 4.8, Bra-mir156 is more highly expressed than Bra-mir172 by over 10-fold during the floral transition in both cultivars. However, when observing the relative abundance of the two miRNAs, a more informative dynamic is revealed where the ratio of Bra-miR156 to Bra-miR172 decreases steadily across development and reaches a consistent value at the floral transition. Using replicate measurements at the morphologically defined floral transition, the mean $\log_{10}$(Bra-miR156/Bra-miR172) ratio was approximately 1.1 in both cultivars (Fig. 4.9A).

Morphological staging identified the floral transition at day 12 in Sarisha-14 and day 19 in R-o-18 1.2. To determine whether the miRNA ratio trajectories quantitatively recapitulate this developmental shift, genotype-specific linear models were fitted to the $\log_{10}$(Bra-miR156/Bra-miR172) ratio across time (Fig. 4.9B). With the model we estimated the time at which each genotype crossed a common ratio threshold of 1.1, which predicted the age at the floral transition was 14.1 ± 0.5 days for Sarisha-14 and 20.4 ± 1.5 days for R-o-18 (mean ± SE). These model-based estimates indicate a 6.3

± 1.5-day shift between genotypes, closely mirroring the 7-day difference observed morphologically and corresponding to an effect size of approximately four standard deviations.

Thus, while the absolute abundance levels of Bra-miR156 remain higher than Bra-miR172 during the floral transition, it is the relative ratio between them that better reflects the timing of the vegetative-to-reproductive transition. In line with our hypothesis, the earlier transition of Sarisha-14 compared to R-o-18 may therefore be explained by Sarisha-14 reaching this optimum miR156/miR172 ratio threshold sooner. The agreement between morphological staging and linear model estimations strengthens the hypothesis that the balance between these miRNAs functions as a developmental switch for initiating the floral transition. Further analysis of the downstream targets of these miRNAs is also needed to establish whether the miR156-SPL-miR172-AP2 regulatory network is also conserved in rapid-cycling *B. rapa* and how it regulates its flowering time.



(a) Log10 miR156/miR172

(b) Linear model

Figure 4.9: **Log10 transformed ratio of Bra-miR156/Bra-miR172 expression of 1.1 coincides with the floral transition in both cultivars.**
(A) The ratio of miR156 to miR172 expression reduces as Bra-miR156 expression decreases, and Bra-miR172 increases, and when the floral transition starts in either cultivar, the average ratio is a log10 value of 1.1. Each point represents the ratio of total miR156 divided by total miR172 expression from a replicate apex sample. The main line is the average ratio for all apex replicates in a single timepoint, separated by genotype (R-o-18 in red, Sarisha-14 in blue). The standard deviation is represented by the shaded region. The grey dashed line indicates the estimated miR156-to-miR172 ratio which coincides with the floral transition in both R-o-18 and Sarisha-14.
(B) Linear models confirm genotype-specific declines. RO: slope=-0.061*** (R^2=0.56), predicted transition at 25d; S14: slope=-0.150*** (R^2=0.79), predicted transition at 14d. Points are individual replicates; lines are fitted slopes.

4.3.2 *Timeseries expression of predicted targets of miR156 and miR172*

4.3.2.1 *Target gene prediction with TargetFinder*

To establish whether the miR156-*SPL* and miR172-*AP2* regulatory modules are conserved in *B. rapa*, target prediction was performed for the individual miR156 and miR172 sequences identified in *B. rapa*, using TargetFinder (*carringtonlab/TargetFinder* 2024), which computationally predicts small RNA binding sites on candidate target transcripts. A total of 30 target genes were predicted for the Bra-miR156 family, while 36 target genes were predicted for the Bra-miR172 family.

Out of the 30 genes predicted to be targeted by Bra-miR156 members, 10 of these genes come from the *SQUAMOSA BINDING PROMOTER LIKE (SPL)* gene family which encodes *SPL* transcription factors, confirming conservation of the miR156-SPL regulatory interactions (Fig. 4.10). Figure 4.11 depicts alignment of the predicted target binding site in the *SPL* genes against the mature Bra-miR156 sequences. The majority of predicted target sites showed high sequence conservation, especially in *SPL* genes such as *SPL9*, *SPL2*, and *SPL15.A07*. In contrast, we observe one nucleotide mismatch in *SPL13* copies and in *SPL15.A04*, and other Bra-miR156 members such as *SPL5.A03* and *SPL10.A08* with more sequence divergence in the first 6 nucleotides of the target site, potentially reducing binding affinity by Bra-miR156 members. These results suggest that while most *SPL* target genes retain strong conservation of the Bra-miR156 binding site, certain *SPL* paralogues may have diverged following genome triplication, possibly leading to divergent regulation.

```
(a) Bra-miR156a - SPL15 (score = 1)     A07p029860.1_BraROA
    Target  5' GUGCUCUCUCUCUUCUGUCA 3'
              :::::: ::::::::::::::
    miRNA   3' CACGAGUGAGAGAAGACAGU 5'

(b) Bra-miR156a - SPL9 (score = 1)       A04p037800.1_BraROA
    Target  5' GUGCUCUCUCUCUUCUGUCA 3'
              :::::: ::::::::::::::
    miRNA   3' CACGAGUGAGAGAAGACAGU 5'

(c) Bra-miR156a - SPL2 (score = 1)       A06p056880.1_BraROA
    Target  5' GUGCUCUCUCUCUUCUGUCA 3'
              :::::: ::::::::::::::
    miRNA   3' CACGAGUGAGAGAAGACAGU 5'

(d) Bra-miR156a - SPL15.A04 (score = 1.5) A04p003490.1_BraROA
    Target  5' GUGUUCUCUCUCUUCUGUCA 3'
              :::.:: ::::::::::::::
    miRNA   3' CACGAGUGAGAGAAGACAGU 5'

(e) Bra-miR156a - SPL13 (score = 2)      A03p017260.1_BraROA
    Target  5' GCGCUCUCUCUCUUCUGUCA 3'
              : :::: ::::::::::::::
    miRNA   3' CACGAGUGAGAGAAGACAGU 5'
```

Figure 4.10: **Representative TargetFinder predictions of Bra-miR156a target sites across textitSPL paralogues.** Panels show representative duplexes for highly conserved *SPL9*, *SPL2*, *SPL15.A07* (score=1), the single nucleotide mismatch in *SPL15.A04* (score=1.5), and mismatch in *SPL13* paralogues (score=2). Target (5′→3′) above, miRNA (3′→5′) below. Colons (:) indicate matches, dots (.) mismatches.

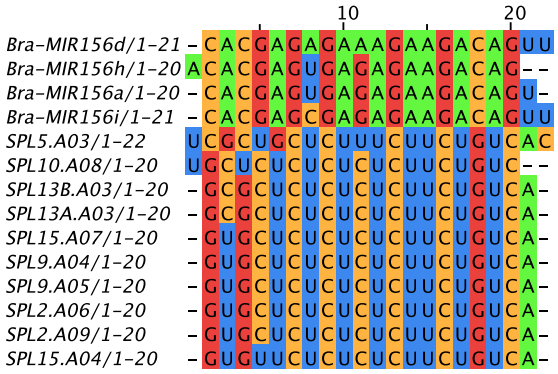


Figure 4.11: **Sequence comparison of Bra-miR156 mature sequence aligned against *SPLs'* target binding sites reveals high conservation in target binding sequences, except for a few *SPLs*.**
The target binding site sequences for *SPL15.A04*, both copies of *SPL9* and *SPL2* are highly conserved, with one nucleotide difference at position 4 for *SPL15.A04*. In contrast, *SPL5.A03, SPL10.A08* have sequence divergence in the first 6 nucleotides, which could potentially impact miRNA binding efficiency. Target binding site predictions of Bra-miR156 members from TargetFinder from each *SPL* target gene was aligned with the mature Bra-miR156 sequences using ClustalOmege. The miRNA sequences are displayed in reverse 3'-5' alignment, while the target binding sites are displayed in forward 5'-3' alignment, displaying the complementary binding between miRNA and target.

For the Bra-miR172 members, target predictions identified 7 *AP2*-like genes (Fig. 4.13), which include *AP2, TARGET OF EAT1 (TOE1, TOE2, TOE3)* and *SCHLAFMÜTZE (SMZ)* (Fig. 4.12). The binding sites were generally conserved, with only 1-2 nucleotide differences in the first 2–3 nucleotides, except for *TOE3.A07*, which exhibited more sequence divergence across nucleotides 1–6 of the predicted site. This would also likely change the binding affinity of Bra-miR172 to *TOE3.A07*, which could have an impact on its temporal expression pattern. Similar to the *SPL* target genes, most of the predicted *AP2*-like target genes have high sequence conservation at the miR172 target binding sites, but nucleotide differences in individual paralogues could lead to divergent Bra-miR172 mediated regulation.

```
(a) Bra-miR172a - TOE2 (score = 1)        A10p023550.1_BraROA
    Target  5′ AUGCAGCAUCAUCAGGAUUCU 3′
              ::::::::::::::.::::::
    miRNA   3′ UACGUCGUAGUAGUUCUAAGA 5′

(b) Bra-miR172a - AP2 (score = 2)         A03p071180.1_BraROA
    Target  5′ CUGCAGCAUCAUCAGGAUUCU 3′
              ::::::::::::::::.::::::
    miRNA   3′ UACGUCGUAGUAGUUCUAAGA 5′

(c) Bra-miR172a - SMZ (score = 3)         A09p060690.1_BraROA
    Target  5′ UUGCAGCAUCAUCAGGAUUCC 3′
              ::::::::::::::.:::::
    miRNA   3′ UACGUCGUAGUAGUUCUAAGA 5′

(d) Bra-miR172a - TOE3 (score = 2.5)      A07p022340.1_BraROA
    Target  5′ AUGGCGGCAUCAUCAGGAUUCU 3′
              ::: :.:::::::::.::::::
    miRNA   3′ UAC-GUCGUAGUAGUUCUAAGA 5′

(e) Bra-miR172c - TOE2 (score = 2)        A10p023550.1_BraROA
    Target  5′ AUGCAGCAUCAUCAGGAUUCU 3′
              ::::::::::::::.::::::
    miRNA   3′ GACGUCGUAGUAGUUCUAAGA 5′
```

Figure 4.12: **TargetFinder rediction of Bra-miR172 target sites across *AP2*-like paralogues.** Panels show representative duplexes for highly conserved *TOE2* (score=1), *AP2* (score=2), *SMZ* (score=3), more divergent *TOE3.A07* (score=2.5), and Bra-miR172c:TOE2 (score=2) showing 1-2 nucleotide differences in positions 1–3. Target (5′→3′) above, miRNA (3′→5′) below. Colons (:) indicate matches, dots (.) mismatches.

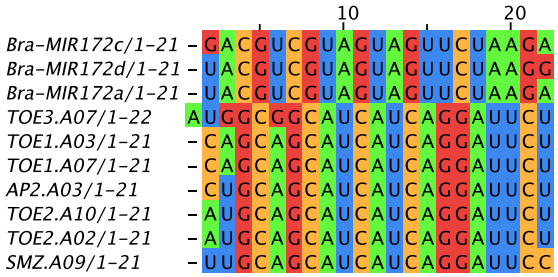


Figure 4.13: **Sequence comparison of Bra-miR172 mature sequence aligned against *AP2*-like genes' target binding sites reveals divergence in some target binding sequences, like *TOE3.A07*.**
Most target genes' binding sites have different nucleotides in the first 2-3 nucleotides of the sequence, while the rest of the target site are highly conserved. *TOE3.A07* shows the most sequence divergence from nucleotide position 1-6, potentially impacting miRNA binding to this target gene. Target binding site predictions of Bra-miR172 members from TargetFinder from each Bra-miR172 target gene was aligned with the mature Bra-miR172 sequences using ClustalOmege. The miRNA sequences are displayed in reverse 3'-5' alignment, while the target binding sites are displayed in forward 5'-3' alignment, displaying the complementary binding between miRNA and target.

These target gene predictions highlight both sequence conservation and divergence in the Bra-miR156-*SPL* and Bra-miR172-*AP2* regulatory networks in *B. rapa*, suggesting that overall these modules are conserved, but individual gene copies may have divergent regulation.

4.3.2.2 *Diverging expression profiles of B. rapa SPL transcripts*

To characterize the temporal expression profiles of the target genes of Bra-miR156 and Bra-miR172, I used time series gene expression data of Sarisha-14 and R-o-18 published by Calderwood, Hepworth, et al. (2021). These time series are similar to my miRNA timeseries, but of expressed mRNA transcripts instead. Note that in this gene expression time series, the floral transition timepoint was earlier, with Sarisha-14 reaching the floral transition at day 10, and R-o-18 at day 17 (Calderwood, Hepworth, et al. 2021). Due to the earlier timing of the floral transition in this time series, comparisons of target gene expression with miRNA abundance was anchored to the same developmental timepoint, at the first day of the floral transition.

Out of the 10 *SPL* orthologues predicted to be targets of Bra-miR156, 9 of them were expressed in R-o-18 and Sarisha-14, while *SPL5.A03* was not detected (see Appendix Fig. A.1). It is likely that *SPL5.A03*'s high sequence divergence (Fig. 4.11) led to it being non-functional, therefore it was not expressed in either cultivar.

At the seedling stage (day 7), all *SPL* transcripts were expressed at very low levels, consistent with strong miR156-mediated repression early in development (Fig. 4.14A). As plants approached the floral transition, most *SPL* orthologues increased in expression. This upregulation was generally more rapid in Sarisha-14 than in R-o-18, aligning with Sarisha-14's earlier floral transition. The expression dynamics of *SPL9.A04*,

SPL9.A05, and *SPL10.A08* coincided with the floral transition (day 12 in Sarisha-14; day 19 in R-o-18), suggesting these paralogues may play a role in promoting the transition once their transcript levels cross a critical threshold.

However, not all *SPL* copies exhibited cultivar-specific upregulation patterns, such as the *SPL2* and *SPL13* copies, which had similar rates of upregulation but not a shift in expression in relation to the floral transition stage in either cultivar. For example, both copies of *SPL13* increased in expression until day 10, at which it hits a plateau and then downregulation near BBCH51 (Fig. 4.14). In line with our target binding site sequence comparisons in Figure 4.11, *SPL13* copies had 1 nucleotide mismatch at position 2, which could allow it to escape from Bra-miR156 repression.

Among the predicted *SPL* targets, the expression profile of *SPL15.A04* stood out as uniquely cultivar-specific. In (Fig. 4.14A), *SPL15.A04* expression was consistently higher in Sarisha-14 across development, while in R-o-18 its expression remained relatively low. We also observe a similar higher expression abundance for its other copy *SPL15.A07*, although the difference is not as noticeable. This suggests that *SPL15.A04* may be subject to stronger genotype-specific regulation than other *SPL* targets, possibly contributing to different flowering times.

Overlaying Bra-miR156 abundance with individual *SPL* profiles revealed further insights (Fig. 4.14B). While many *SPLs* showed the expected reciprocal expression profile with Bra-miR156 upregulation, some *SPL*s did not exhibit this pattern. For example, *SPL13* transcripts were upregulated despite no clear decline in Bra-miR156a, suggesting that *SPL13* activation may occur independently of Bra-miR156 repression (Fig. 4.14B). In contrast, *SPL15.A04* in Sarisha-14 displayed a strong reciprocal relationship with Bra-miR156a, but not so much in R-o-18, which suggests stronger Sarisha-14 specific regulatory control.

Overall, these results indicate that while many *SPL* orthologues show conserved upregulation patterns, a subset—including *SPL15.A04* and several *SPL9/SPL10* copies—exhibit cultivar-specific dynamics that may underlie differences in the timing of the floral transition between R-o-18 and Sarisha-14.

Figure 4.14 (next page)
A) Each row displays the mean expression of an *SPL* orthologue across development in R-o-18 (red) and Sarisha-14 (blue). Most *SPL* orthologues display a rise in expression prior to the floral transition, with Sarisha-14 showing a rapid increase compared to R-o-18's gradual increase. Only the *SPL13* copies show temporally similar expression profiles, with rapid upregulation at around day 10 in both cultivars. No genotype specific interactions were identified from differential expression analysis in any *SPL* orthologue, suggesting that their expression dynamics are not genotype-specific. However, *SPL15.A04* is expressed more in Sarisha-14 than R-o-18 consistently throughout development, which implies divergent expression regulation.
B) Overlay of standardised expression of Bra-miR156 against each *SPL* target gene, aligned to the floral transition developmental stage (day 0). Each row corresponds to the target gene in panel A, with each column displaying the genotype-specific Bra-miR156 vs *SPL* expression profiles. The Bra-miR156 member with the strongest binding score predicted by TargetFinder is plotted against each target gene. Bra-miR156a-SPL13 expression profiles don't show reciprocal expression, which could suggest that *SPL13* activation is independent of Bra-miR156a downregulation. On another note, *SPL15.A04* displays reciprocal expression to Bra-miR156a in Sarisha-14 leading to the floral transition, but less so in R-o-18, which suggests genotype-specific regulation of *SPL15.A04*.

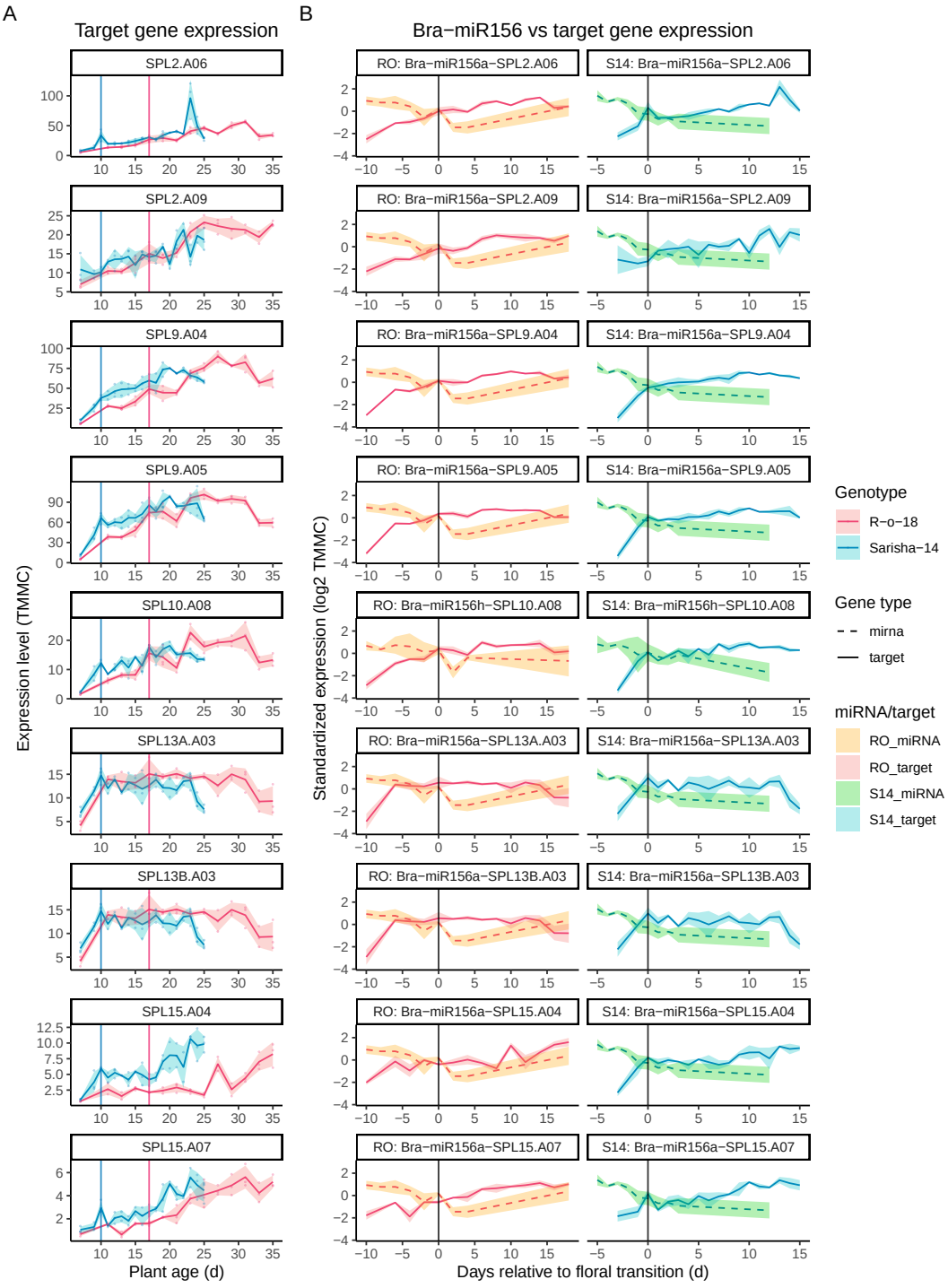


Figure 4.14: **Expression of Bra-miR156 predicted targets in R-o-18 and Sarisha-14 shoot apex.** The expression of target *SPL* orthologues is shown alongside Bra-miR156 vs. individual *SPL* expression profiles to assess aligned reciprocal expression dynamics, or the lack of it.

4.3.2.3 *Most B. rapa AP2-like genes expression profiles are consistent with miR172-mediated repression*

All 7 *AP2*-like target genes that were predicted by TargetFinder to be Bra-miR172 targets were expressed in R-o-18 and Sarisha-14 shoot apex. At the earliest vegetative timepoint (day 7), when Bra-miR172 abundance was at its lowest, only two predicted target genes, *SMZ.A09* (which was already characterized by Calderwood, Hepworth, et al. (2021)) and *TOE2.A10*, were highly expressed (Fig. 4.15A). Their transcript abundance gradually declined over time, with more rapid downregulation in Sarisha-14 coinciding with earlier floral transition. For example, *SMZ.A09* expression dropped below 5 TMMC at the floral transition, while *TOE2.A10* fell below 20 TMMC, suggesting that these transcripts must be downregulated to a critical threshold before the floral transition can begin. This expression pattern is consistent with their known roles as floral repressors, maintaining the vegetative stage while it is highly expressed until subsequent repression by the increasing Bra-miR172 abundance. Interestingly, *TOE2.A10* expression was much lower in Sarisha-14 in the seedling stage, which could imply that lower *TOE2.A10* expression had a weaker repressive effect on flowering in Sarisha-14.

AP2.A03 was expressed at low levels at the seedling stage but increased in abundance during vegetative development, reaching peak expression just prior to the floral transition (Fig. 4.15A). After its peak during the floral transition, its expression declines as Bra-miR172 abundance increases further in development. Bra-miR172c is the best-scoring miRNA for binding to *AP2.A03*, and comparison between Bra-miR172c and *AP2.A03* expression in Figure 4.15B shows *AP2.A03* upregulation follows Bra-miR172c upregulation in the vegetative stage, but only after the floral transition does *AP2.A03* expression decrease. This pattern is also reflected in certain *TOE* members, particularly *TOE1.A03, TOE1.A07* and *TOE2.A02*, which have variable high expression in the vegetative stage, and only decreases after Bra-miR172 reaches a certain abundance level at the floral transition. These patterns suggest that some *AP2-like* targets require sustained or elevated miR172 abundance for repression to take effect, in contrast to the earlier and sharper decline seen for *SMZ.A09* and *TOE2.A10*.

Finally, *TOE3.A07* displayed an expression pattern distinct from the other predicted targets. Instead of being downregulated when Bra-miR172a abundance increased, it showed a steady increase in expression across apex development in both cultivars (Fig. 4.15B). Given that *TOE3.A07* had the least conserved Bra-miR172 binding site sequence compared to the other genes (Fig. 4.13), this upregulation suggests that its transcript may be resistant to miR172-mediated repression, potentially due to sequence divergence at the binding site.

In conclusion, the expression dynamics of Bra-miR156 targets (*SPLs*) and Bra-miR172 targets (*AP2*-like genes) suggest that the ageing pathway in *B. rapa* regulates flowering via a combination of threshold-dependent repression and binding site sensitivity to

miRNA regulation. While most targets followed the expected reciprocal expression pattern consistent with miRNA mediated repression, some target genes like *SPL13* and *TOE3.A07* highlight how sequence divergence at their binding sites can enable these genes to partially evade miRNA-mediated repression. Furthermore, the consistent timing of target downregulation at the floral transition in both cultivars suggests that the developmental switch is not driven by the mere presence of miRNA abundance, but rather by the balance between miRNA and target expression, coupled with cultivar-specific differences in how rapidly these thresholds are reached.

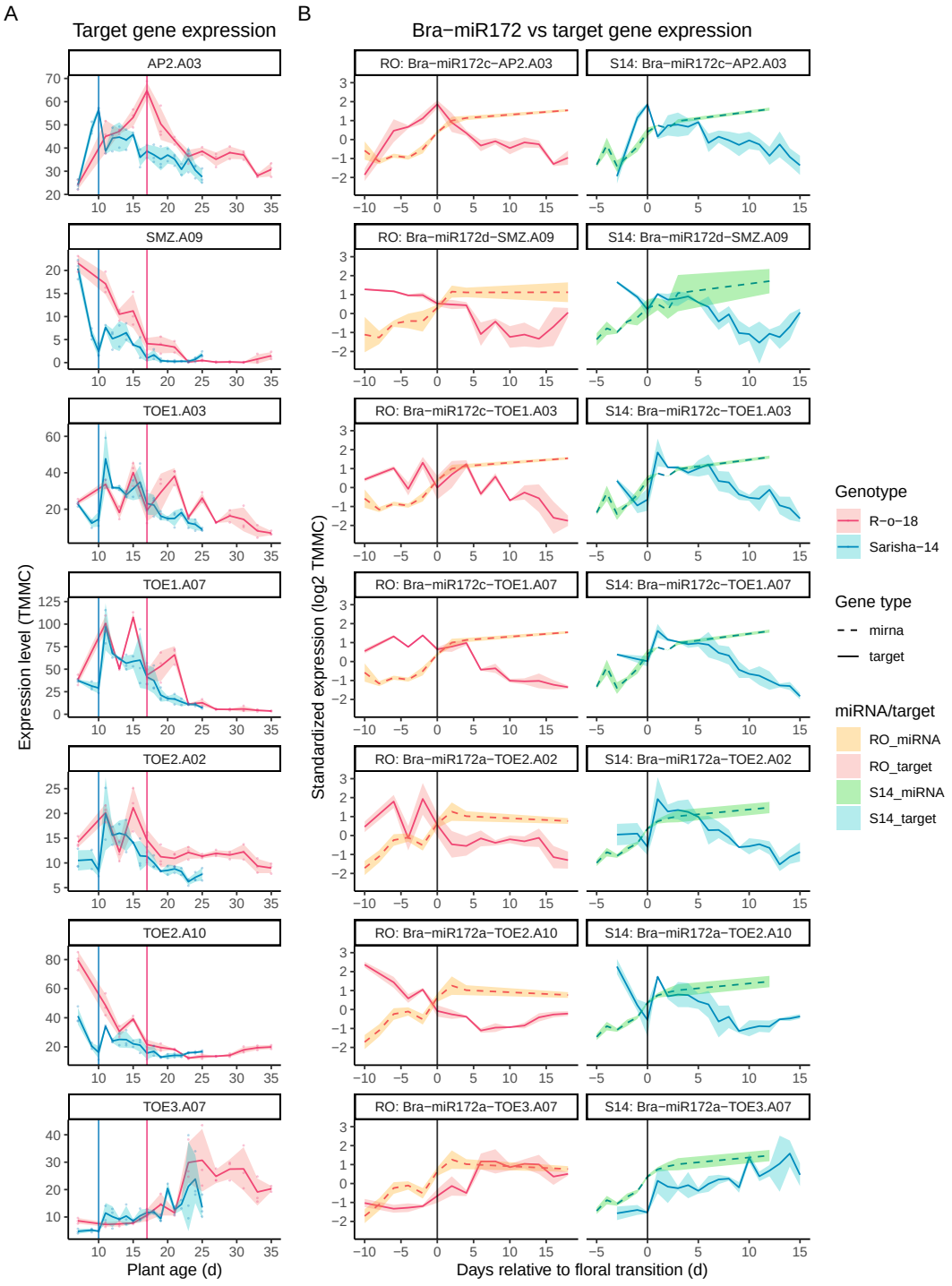


Figure 4.15: **Expression of Bra-miR172 predicted targets in R-o-18 and Sarisha-14 shoot apex.** The expression of target *AP2*-like orthologues is shown alongside Bra-miR172 vs. individual target gene expression profiles to assess aligned reciprocal expression dynamics.

Figure 4.15 (continued)

A) Each row displays the mean expression of an *AP2*-like orthologue across development in R-o-18 (red) and Sarisha-14 (blue). Most *AP2*-like orthologues decrease in expression after the floral transition, consistent with Bra-miR172-mediated repression.

SMZ.A09 and *TOE2.A10* are highly expressed early and sharply downregulated as the floral transition begins, suggesting a threshold-like requirement for their repression before flowering can initiate. Other paralogues such as *AP2.A03*, *TOE1.A03*, *TOE1.A07*, and *TOE2.A02* peak just before the floral transition and decline more gradually, consistent with delayed repression by miR172.
B) Overlay of standardised expression of Bra-miR172 against each *AP2*-like target gene, aligned to the floral transition developmental stage (day 0). Each row corresponds to the target gene in panel A, with each column displaying the genotype-specific Bra-miR172 vs target gene expression profiles. The Bra-miR172 member with the strongest binding score predicted by TargetFinder is plotted against each target gene. Reciprocal expression dynamics are evident for *SMZ.A09* and *TOE2.A10*, whereas other paralogues such as *AP2.A03* and certain *TOE* members show repression only after Bra-miR172 abundance reaches higher levels after the floral transition. In contrast, *TOE3.A07* expression increased steadily despite rising Bra-miR172 abundance, consistent with its divergent target site sequence (Fig. 4.13) and suggesting partial resistance to miR172-mediated repression.

4.4 DISCUSSION

The molecular mechanisms which regulates the timing of the floral transition in rapid-cycling *Brassica* cultivars, which are able to flower without requiring vernalization, is yet to be uncovered. In this chapter, I demonstrated that Bra-miR156's rapid decline alongside Bra-miR172's upregulation coincides with earlier initiation of the floral transition in Sarisha-14. I also identified a potential optimum miR156/miR172 abundance ratio which coincides with the same developmental transition point in both cultivars. Although we see evidence that the miR156-*SPL* and miR172-*AP2*-like regulatory modules are conserved and transcriptomic evidence confirms miRNA-mediated repression, further analysis is required to establish the downstream molecular mechanisms that allows Sarisha-14 and R-o-18 to flower without vernalization.

4.4.1 *Conservation and divergence of miR156–SPL regulation*

The reciprocal dynamics we observed between Bra-miR156 and most *SPL* orthologues are consistent with findings in *A. thaliana*, where miR156 represses *SPL*s early in development, delaying flowering (G. Wu, Park, et al. 2009; M. Xu, T. Hu, Zhao, et al. 2016). In *B. rapa*, *SPL9* and *SPL10* copies were more rapidly upregulated at the floral transition in Sarisha-14 than in R-o-18, suggesting that accelerated *SPL* derepression via Bra-miR156 decline contributes to Sarisha-14's earlier flowering time.

However, not all *SPL*s followed this reciprocal expression pattern from the vegetative stage. In R-o-18, *SPL13* copies were upregulated at the same rate and day as Sarisha-14, reaching a peak in expression at around day 11, although Bra-miR156 abundance in R-o-18 was still relatively high in the vegetative stage. This implies that *SPL13* upregulation in R-o-18 may be independent of Bra-miR156 decline, possibly due to mismatches in the predicted binding site. The single nucleotide mismatch in the Bra-miR156 binding site of *SPL13* may explain this partial escape, highlighting how small sequence divergences can modulate miRNA–target interactions. Among all SPLs, *SPL15.A04* was especially cultivar-specific: its expression was consistently higher in Sarisha-14 and closely tracked Bra-miR156 decline, whereas in R-o-18 the relationship was weaker. This suggests that Sarisha-14 may rely on tighter *SPL15* regulation to achieve earlier flowering.

Our identification of *SPL9*, which is strongly upregulated during the floral transition, *SPL13*, which is upregulated independent of Bra-miR156 decline, and *SPL15*, which has cultivar-specific expression patterns, emphasise their importance in regulating the floral transition, and parallel to their Arabidopsis orthologues, where *AthSPL9, AthSPL13* and *AthSPL15* are dominant promoters of the floral transition, with the triple *spl9/spl13/spl15* mutant taking the longest time to flower compared to other *spl* mutants (M. Xu, T. Hu, Zhao, et al. 2016).

4.4.2 *Differential regulation of AP2/TOE targets by miR172*

The expression profiles of Bra-miR172 targets also confirm the conservation of the miR172–AP2 module, and are consistent with Arabidopsis studies. Jung et al. (2007) showed that in Arabidopsis, *TOE1, TOE2, SMZ,* and *SNZ* are miR172-targeted floral repressors which decline in expression during development, except *TOE3*, which continues to be upregulated until flowering. We observed a remarkably similar pattern in *B. rapa*, where *SMZ.A09* and *TOE2.A10* were highly expressed in the vegetative stage and declined before the floral transition, while other targets such as *AP2.A03, TOE1.A03*, and *TOE1.A07* remained high during vegetative development and were only repressed after Bra-miR172 levels rose substantially, suggesting they require sustained Bra-miR172 abundance for repression. Meanwhile, *TOE3.A07* increased throughout development in both cultivars and was resistant to Bra-miR172 repression, mirroring the same expression pattern in Arabidopsis (Jung et al. 2007). We also observed subtle differences in *TOE2.A10* expression level, which is notably lower in Sarisha-14 than R-o-18 at the seedling stage, which could reduce its floral repressor effect in the shoot apex and make it more permissive for the floral transition to take place.

From these observations, I would hypothesize that *SMZ.A09* and *TOE2.A10* are the most sensitive to Bra-miR172 repression, due to their rapid response to Bra-miR172's modest upregulation before the floral transition. Repression of these two genes leads to the initiation of the floral transition, and following our hypothesis, this occurs earlier in Sarisha-14 since Bra-miR172 is upregulated earlier. Meanwhile, the other *AP2*-like genes, *AP2.A03, TOE1.A03*, and *TOE1.A07* were less sensitive to Bra-miR172 during the vegetative stage, maintaining its high expression, and only when Bra-miR172 reached sufficently high levels after the floral transition, were these floral repressors downregulated. This suggests that these less sensitive genes are more important for progression to flowering, but not necessarily the switch from vegetative to reproductive growth, since they are not repressed prior to the floral transition.

4.4.3 *Ratio-based model of flowering time control*

Overall, these findings support our hypothesis that the timing of the floral transition is defined not just by the antagonistic abundance dynamics of Bra-miR156 or Bra-miR172, but that there is a ratio between these two miRNAs which coincide with the initiation of the floral transition.

In both cultivars, *SPL* floral promoters increases once Bra-miR156 levels fall sufficiently, while *AP2*-like floral repressors decline once Bra-miR172 rises past a crucial threshold. The earlier flowering of Sarisha-14 can thus be explained by its accelerated decline in Bra-miR156 and more rapid accumulation of Bra-miR172, enabling it to reach the same ratio threshold earlier than R-o-18 (Fig. 4.9). This provides quantitative

evidence for the idea that timing of developmental transitions can be remodulated via the abundance dynamics of miR156 and miR172, and has exciting potential for accelerated plant breeding and developing rapid-cycling cultivars.

However, it remains unclear whether the observed miR156/miR172 ratio is causative of the floral transition or merely predictive of an underlying developmental shift. Although the ratio coincides precisely with the first day of the floral transition in both cultivars, correlation does not establish mechanistic sufficiency. The ratio may represent a quantitative developmental threshold that actively determines the balance between SPL accumulation and AP2-like repression, thereby driving meristem reprogramming. Alternatively, it may function as a biomarker that reflects upstream developmental or physiological changes that independently commit the shoot apex to flowering. Distinguishing between these possibilities requires temporal resolution beyond coincidence, particularly to determine whether shifts in the ratio precede morphological commitment at the shoot apex or occur as a consequence of it.

Furthermore, it is possible that the floral transition is governed not by absolute miRNA abundance but by the relative stoichiometry between miR156 and miR172, establishing a threshold-based regulatory checkpoint. In such a model, minor quantitative differences in abundance dynamics could disproportionately influence developmental timing. Conversely, the ratio may simply track the progressive decline of juvenility and accumulation of flowering competence without acting as the initiating trigger. Clarifying whether the ratio represents a regulatory switch or a developmental readout is therefore central to interpreting its biological significance.

4.4.4 *Future directions*

While our results provide strong support for the ratio hypothesis, further work is needed to rigorously test this model:

Degradome/proteome analysis to trace miR156/mir172-mediated repression

Although we predicted the target genes of Bra-miR156/Bra-miR172 and confirmed the conservation of the miR156-SPL and miR172-AP2 regulatory modules in *B. rapa* by visualising their reciprocal expression dynamics, this does not prove direct regulation. MiRNAs repress the expression of their targets either by DICER-mediated cleavage at the binding site, or by preventing protein translation of the transcripts when bound to the target binding site. Our current transcriptome datasets from Calderwood, Hepworth, et al. (2021) only capture the resulting expression dynamics of target genes, but not specifically the degradation of the target genes, or its functional repression by inhibiting mRNA translation. It is possible that the transcripts profiled were the

actually remaining transcripts that were yet to be cleaved by miRNA-recruited DICER complexes.

Thus, degradome profiling of R-o-18 and Sarisha-14 shoot apex would identify target genes which were cleaved via miR156/miR172, and experimentally validate the repression of its target transcripts. In addition, proteomics profiling would reveal the downstream effects of miRNA post-transcriptional repression for target genes that were not cleaved by DICER.

Perturbation of ageing pathway regulatory network

Experimentally testing the ratio hypothesis

To elucidate how the Bra-miR156 and Bra-miR172 regulatory network can be exploited for crop breeding, we could perturb it by altering the expression of either miRNAs, using targeted manipulation of their abundance levels. For instance, over expressing Bra-miR172 with a constitutive promoter in R-o-18 could test whether precocious up-regulation of Bra-miR172 would lead to accelerated flowering, thus achieving a similar phenotype as Sarisha-14. On the other hand, we could also test the oppposite effect by overexpressing Bra-miR156 in Sarisha-14, which we would expect to delay its flowering time. Further transcriptomic profiling could also reveal how these perturbations change the regulatory network, and provide a basis for modelling this gene regulatory network.

To directly test whether the miR156/miR172 ratio is causative rather than correlative, we can use targeted manipulation of the ratio, by generating inducible expression systems such as the dexamethasone- glucocorticoid system (Craft et al. 2005) Bra-miR172. This allows temporal control of miRNA abundance prior to the floral transition, where inducing the expression of Bra-miR172 at different levels of endogenous Bra-miR156 would achieve different miR156/miR172 ratios. By artificially shifting the ratio earlier or later in development, we could then determine whether altering the miRNA balance can accelerate or delay the floral transition, and in turn flowering time. Another method of testing the ratio is using target mimicry, where a non-cleavable RNA is introduced into the plant that can sequester its target miRNAs, thus preventing the miRNA from binding to its target genes (Franco-Zorrilla et al. 2007). Introducing a Bra-miR156 target mimic would effectively reduce the functional activity of Bra-miR156 without altering its transcription or genomic dosage, allowing selective modulation of miRNA activity while maintaining endogenous regulatory architecture. This can further establish whether a quantitative threshold for the miR156/miR172 exists, and whether flowering time responds in a threshold-dependent (a sharp transition at a critical ratio) or in a gradual manner (a continuous, dose-dependent response to ratio perturbation).

We could also alter the miRNA target binding sites of specific *SPL* and *AP2*-like target genes, in order to study whether nucleotide mismatches in the binding sites lead

to decreased susceptibility to miRNA repression, or if miRNA abundance is a factor for target gene abundance. For example, mutating the Bra-miR156 binding site in *SPL9*, one of the Bra-miR156 which shows aligned expression patterns in both cultivars, with a nucleotide mismatch similar to *SPL13*'s binding site, might provide insight on how *SPL13* is able to evade repression by Bra-miR156.

If the miR156/miR172 ratio defines the timing of the floral transition, it could also serve as a proxy for the floral transition in other Brassica cultivars, and potentially other species as well. Characterising the ratio of miR156/miR172 and modelling its abundance in different contexts such as vernalisation-requiring vs non-vernalisation requiring cultivars, or different crop types like *B. oleracea* vegetative cabbage vs. floral-arrested cauliflower, could provide insight on whether the ratio persists in these cultivars, and that there is a key ratio which needs to be achieved before the floral transition occurs. We could utilise pri-miRNA expression levels to also serve as a proxy for miRNA abundance, going back to how Calderwood, Hepworth, et al. (2021) first presented the ratio hypothesis based on pri-miR156 and pri-miR172 expression profiles.

A wheat study also investigated the ageing pathway in different crop types - in spring cultivars, flowering coincides with a decline in miR156 and an increase in miR172 that represses AP2-like genes, whereas in winter cultivars the same repression requires vernalization prior to age-dependent flowering (Debernardi, Woods, et al. 2022). Although it doesn't mention the ratio of miR156/miR172 abundance, it emphasizes the balance of these two miRNAs, and the temporal shift of flowering time due to the vernalisation requirement in winter wheat.

Thus, these approaches would shift the ratio hypothesis from correlative observation to mechanistic testing. By systematically perturbing miRNA abundance, activity, and target sensitivity, it would be possible to determine whether the miR156/miR172 ratio acts as a developmental threshold, a quantitative modulator of flowering time, or simply a ratio that correlates with flowering time. Establishing this distinction is critical before the ratio can be reliably used as a predictive or translational tool in Brassica breeding.

Epigenetic regulation of the ageing pathway

What controls miR156 abundance decline? Our miRNA abundance timeseries confirms that Bra-miR156 abundance decreases with plant age, and when it is sufficiently repressed, *B. rapa* proceeds to the floral transition. However, the regulation of the timing of miR156 decline itself is still being studied. A recent review by Poethig and Fouracre (2024) describes the different molecular mechanisms which regulate the actual timing of vegetative phase change, upstream of miR156 regulation. One of these mechanisms include chromatin remodelling of the *MIR156* encoding gene loci to regulate its abundance. During the vegetative stage, miR156 is highly expressed thanks to maintenance

of H2A.Z histone variants, and deposition of activating H3K4me3 marks by *BRAHMA (BRM)* and the *SWI2/SNF2-Related 1 Chromatin Remodeling (SWR1)*(Y. Xu et al. 2016; M. Xu, Leichty, et al. 2018). Antagonistic to this, miR156 abundance is downregulated transcriptionally by *Polycomb Repressive Complex 2 (PRC2)* components which promote the deposition of silencing H3K27me3 and H2Aub marks as plants age (M. Xu, T. Hu, Smith, et al. 2016; Picó et al. 2015; Y. Xu et al. 2016). These mechanisms reveal an additional layer of regulation at the chromatin level.

ChIP-seq (chromatin immunoprecipitation DNA sequencing) can identify DNA sequences associated with target chromatin binding factors. A study in *B. rapa* R-o-18 focused on the H3K27me3 epigenome reported that a histone methyltransferase *curlyleaf (clf)* mutant had reduced H3K27me3 at floral integrator genes (Payá-Milans et al. 2019). However, their study was limited to leaf and inflorescence tissues at the floral stage, and did not report on H3K27me3 deposition at Bra-miR156 encoding loci, so there is a gap in the literature to investigate the epigenetic landscape of *B. rapa* during vegetative development.

It is possible that epigenetic regulation of Bra-mir156 differs between R-o-18 and Sarisha-14. An earlier decline in Bra-miR156 in Sarisha-14 could be due to earlier repression via H3K27me3 deposition via PRC1 and PRC2 components, or reduced activating H2A.Z deposition in miR156 encoding loci. This could be tested by profiling chromatin states at Bra-mir156 loci using CUT&Tag for H3K4me3, H3K27me3, and H2A.Z, or by comparing expression of PRC1/PRC2 component genes between the two cultivars across development.

Physiological signals that regulate ageing

Nutrients are also key for plant development, and sugars have been shown to repress miR156 transcription, which in turn regulates flowering time. Seedlings grown in glucose media have decreased levels of miR156 by 50%, while *SPL9* reporter expression increased subsequently (S. Yu et al. 2013; L. Yang et al. 2013). The current theory is that as the seedling grows older and leaves get bigger, photosynthesis increases, thus sugar levels in the leaves and shoot apex increase, repressing miR156 transcription downstream. Mutants within the sugar singalling pathway such as *Arabidopsis cao/chlorina1 (ch1)* mutant and *trehalose-6-phosphate synthase1 (tps1))* mutants exhibited late flowering time phenotypes, suggesting that this pathway is linked to the juvenile-to-adult transition via miR156 repression (S. Yu et al. 2013; L. Yang et al. 2013). Sugar-mediated repression of miR156 was also observed in *Physcomitrella patens, Solanum lycopersicum,* and *Nicotiana benthamiana*, which strongly implies a conserved endogenous cue in angiosperms (L. Yang et al. 2013).

This is an exciting hypothesis to test in R-o-18 and Sarisha-14, and could be tested by measuring glucose levels in the leaf and shoot apex tissue prior to the floral transition, or applying sugar treatment to seedlings and measuring the impact on miR156

abundance via RT-qPCR.

4.4.5 *Agricultural implications*

Flowering time is a key agricultural trait in most flowering crops besides Brassica. But specifically in Brassica, flowering crop types such as *B. napus* oilseed rape and *B. rapa* mustard depend on optimum flowering time to achieve maximum yield, while vegetative crop types such as *B. oleracea* cabbage or broccoli requires continuous growth without triggering bolting.

The ratio hypothesis opens exciting opportunities to fine-tune flowering time in Brassica crops, where targeted manipulation of the ageing pathway through transgenic or genome-editing approaches can more precisely re-modulate flowering time. For flowering/oilseed crop types, inducible Bra-miR172 expression systems could be engineered to accelerate floral transition in response to external inducing signals, which would make more synschronized flowering time possible, and improve yield predictability. Conversely, for vegetative crops where premature bolting reduces yield and commercial value, stabilising Bra-miR156 expression, either through sustained overexpression or modification of upstream repressors could prolong the juvenile phase and extend vegetative biomass accumulation. Such approaches would allow temporal control rather than constitutive changes to development, potentially minimising pleiotropic effects on plant architecture or stress responses. However, careful evaluation of trade-offs between accelerated reproduction and biomass allocation, as well as regulatory and consumer acceptance of genetically modified crops, would be essential before practical implementation.

A more ambitious goal would be to develop rapid-cycling crops from winter-type crops, creating more climate resilient crops that don't depend on vernalization to initiate flowering. Winter oilseed rape and related cultivars typically require prolonged cold exposure to initiate flowering, which constrains geographic range, cropping flexibility, and resilience under increasingly unpredictable winters. If the miR156–miR172 balance contributes to establishing the competence to flower independently of cold exposure, then re-modulating the rate of Bra-miR156 decline or accelerating Bra-miR172 accumulation could, in principle, shift the developmental threshold at which flowering becomes permissible. Rather than bypassing vernalization entirely at the level of classical flowering repressors, manipulation of the ageing pathway could promote earlier acquisition of floral competence, potentially reducing cold dependency while preserving agronomically important winter growth traits.Thus, understanding and exploiting this miRNA balance could provide new avenues for crop improvement.

4.5 CONCLUSION

Together, these results draw a conserved miR156-SPL / miR172-AP2 ageing pathway in rapid-cycling *B. rapa*. Cultivar specific miR156 abundance is reflected in *SPL15.A04* which shows downregulation consistent with miR156 expression, while miR172 target genes *SMZ.A09* and *TOE2.A10* are downregulated at the same time miR172 is upregulated during the floral transition stage. Furthermore, we quantitatively confirmed that mature miR156/miR172 log 10 expression ratio of 1.1 coincides with the first day of the floral transition, which suggests that this ratio can be used as a proxy for the timing of the floral transition.

5

FUNCTIONAL ANNOTATION OF CANDIDATE SNPS LINKED TO EARLIER FLOWERING

5.1 INTRODUCTION

This thesis chapter investigates the genetic basis of flowering time divergence between two rapid-cycling *Brassica rapa* cultivars, R-o-18 and Sarisha-14. Under the same growth conditions, R-o-18 reaches developmental stage BBCH51 (floral buds visible, (Meier et al. 2009)) 37 days after sowing, while Sarisha-14 floral reaches the same stage earlier at day 24, which is almost twice as fast compared to R-o-18. Besides gene expression and miRNA expression profiling, having these two cultivars with contrasting phenotypes presents a unique opportunity to conduct bulked segregant analysis (BSA).

BSA allows us to identify genomic regions linked to this difference in flowering time, by detecting allele frequency differences between two 'bulks' of individuals with contrasting extreme phenotypes. The allele frequency at the same Single Nucleotide Polymorphism (SNP) position is calculated for each bulk (SNP-index). Candidate genomic loci which could underlie the variation in flowering time between the two bulks, would exhibit greater divergence in SNP-index between the bulks (Δ SNP-index).

In this chapter, 850 segregating individuals from an F2 cross of R-o-18 and Sarisha-14 were generated, and BSA was performed on the extremely early and extremely late flowering individuals (in total <5% of population). To complement the genomic loci identified from the BSA, we also utilised available time series RNA-seq datasets from Calderwood, Hepworth, et al. (2021) to visualise the expression patterns of candidate genes in the R-o-18 and Sarisha-14 cultivars.

For this study, I hypothesised that the difference in flowering time between R-o-18 and Sarisha-14 could be explained by allelic differences (SNPs) in Sarisha-14 within candidate flowering time genes, particularly floral repressors. This assumes that SNPs within floral repressors would lead to disruption of its regulation on flowering, presumably allowing Sarisha-14 to progress to the floral transition earlier.

This chapter presents the results of the BSA and time series transcriptome analyses, highlighting candidate flowering time genes which may contribute to earlier flowering

time in Sarisha-14. I describe the F1 genotyping and F2 phenotyping results, followed by a chromosome-by-chromosome analysis of potential candidate genes which contribute to an early flowering phenotype.

5.2 METHODS

5.2.1 *B. rapa crossing and F1 plant generation*

Brassica rapa cultivars Sarisha-14 and R-o-18 plants were sown in 24 cell trays in cereals mix (40% medium grade peat, 40% sterilised soil, 20% horticultural grit, 1.3 kg/m^3 PG mix 14-16-18 + Te base fertiliser, 1 kg/m^3 Osmocote Mini 16-8-11 2 mg + Te 0.02% B, wetting agent, 3 kg/m^3 maglime and 300 g/m^3 Exemptor). Material was grown in a Conviron MTPS 144 controlled environment room with Valoya NS1 LED lighting (250 µmol m^{-2} s^{-1}) 18 °C day/15 °C night, 70% relative humidity with a 16 hour day (lights on from 00.30hrs until 16.30hrs).

R-o-18 (female) and Sarisha-14 (male) plants grown were crossed after both cultivars were at developmental stage BBCH60 (first flower opens). Hand pollination was performed using freshly opened flowers. To perform crossing, R-o-18 flowers which were very close to opening were emasculated, leaving the unfertilised female pistil open for pollination from Sarisha-14 flowers. The pollinated flowers were then labelled and bagged using perforated bread bags to prevent further pollination. Mature seed pods were harvested when they turned brown, and threshed to obtain the F1 R-o-18 x Sarisha-14 crossed seeds.

5.2.2 *F1 genotyping*

16 F1 seeds were grown under CER conditions, and then selfed by Dr Monica Chhetry, followed by primer screening to identify successfully crossed progenies.

The F1 plants were genotyped for successful heterozygous crosses between R-o-18 x Sarisha-14, using RA2F04 SSR (Single Sequence Repeats) microsatellite markers (primer sequences in Table 5.1) (Lowe et al. 2004). Since all the crosses were performed onto R-o-18 (female), only plants producing an amplified band between 100-200bp were considered as successfully crossed. PCR was performed with AmpliTaq Gold™ DNA Polymerase (ThermoFisher Scientific) in a final reaction volume of 20 uL, with 2 uL of diluted genomic DNA (100 ng), 0.5 uM for each primer, 2 uL of 10X PCR buffer, 1 unit or 0.2 uL of AmpliTaq Gold™ DNA Polymerase (5u/µl), 0.26 uL of 10mM dNTPs and 13.14 uL distilled water. The PCR programme had an initial denaturation step at 95 °C for 10 minutes, followed by 35 cycles of denaturation at 95 °C for 30 seconds, annealing at 59.4 °C for 1 min 35 sec, and DNA extension at 72 °C for 1 min 35 sec,

then finished off with a final extension step at 72 °C for 10 minutes. DNA bands were stained with ethidium bromide and run on gel electrophoresis in 3% Agarose gel at 100V for 60 minutes, then imaged with a Syngene G:BOX gel imager.

Primer	Sequence	Tm (°C)
RA2F04 Forward	CCTACAAACACATAAATAAAGAGAGAG	56.3
RA2F04 Reverse	AACAACATAAAAGATTCATTTCG	55.3

Table 5.1: **Primer sequences of RA2F04 SSR markers used to genotype F1 R-o-18 x Sarisha-14 lines.**

5.2.3 *F2 growth conditions and phenotyping*

850 seeds from confirmed F1 R-o-18 x Sarisha-14 crossed lines were sown in 24-cell trays in cereals mix (40% Medium Grade Peat, 40% Sterilised Soil, 20% Horticultural Grit, 1.3 kg/m³ PG Mix 14-16-18 + Te Base Fertiliser, 1 kg/m³ Osmocote Mini 16-8-11 2 mg + Te 0.02% B, Wetting Agent, 3 kg/m³ Maglime, 300 g/m³ Exemptor). They were grown with regular watering for 8 weeks from May-June 2024, in a glasshouse under natural light supplemented with LED lighting for 16 hour days, and estimated temperatures at 21 °C day/18 °C night.

Each cell was numbered from 1-850. The plants were scored for developmental stage BBCH51, which is the first day where the floral buds are visible by eye. Individual plants which had reached BBCH51 were recorded by recording the date and the plant's number, and a marker was inserted into their pots to indicate plants that have reached BBCH51 stage. Around 21 days after sowing, leaf tissue from the plants were sampled into 96-microtube boxes, then kept frozen in -20 °C freezers.

To obtain the whole genome sequence from parent lines, 24 Sarisha-14 plants were also grown under glasshouse conditions, and leaf tissue was pooled and kept frozen in -20 °C. R-o-18 tissue was not sampled since this cultivar has a high quality published reference genome available on EnsemblPlants (https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-59/fasta/brassica_rapa_ro18/dna/Brassica_rapa_ro18.SCU_BraROA_2.3.dna.toplevel.fa.gz).

5.2.4 *DNA extraction and sequencing*

The leaf tissue collected from the individual F2 plants were bulked (see Table 5.2). DNA extraction was performed on the pooled leaf tissue and Sarisha-14 tissue. Pooled tissue was ground into a fine powder with a pestle and mortar while kept frozen with liquid nitrogen, and then transferred into a 50 mL Falcon tube using a cooled sterile spatula, and frozen in liquid nitrogen.

A sorbitol wash protocol was performed to remove secondary metabolites from the tissue and improve the purity of extracted DNA, following a published protocol (A. Jones and Schwessinger 2021; A. Jones, Torkel, et al. 2021). Sorbitol wash solution was prepared using the following recipe to create a 500 mL stock solution: 31.88 g of D-Sorbitol, 5 g of PVP 40000, 50 mL of 1M Tris-HCl (pH 8), 5 mL of 0.5M EDTA (pH 8), and Milli-Q Water to bring the solution to a total of 500 mL (A. Jones and Schwessinger 2021). Prior to DNA extraction, around 7.5 mg of DTT was added to 50 mL of the sorbitol wash solution, and around 10 mL of the solution was added to the 50 mL Falcon tube with the ground leaf tissue. The tube is shaken and vortexed to mix the wash solution and ground leaf tissue, then centrifuged at 5000 rcf for 5 minutes at room temperature to pellet the tissue. The supernatant is decanted carefully from the tube, keeping the tissue pellet, and the wash step was repeated, until the supernatant was no longer viscous and lighter green in colour.

DNA extraction was performed using the Illustra Nucleon Phytopure Genomic DNA Extraction kit (Cytiva Life Sciences, https://www.cytivalifesciences.com/en/gb), with an adapted protocol provided by Dr Roland Wouters from the JIC Molecular Biology platform. In this protocol, Reagent 1 refers to the extraction kit's lysis reagent named Reagent 1, while Reagent 2 contains potassium SDS to form complexes with polysaccharides in the lysed plant cells. First, 1 uL of RNAse was first added to 1 mL of Reagent 1 (lysis reagent), and then incubated at 37 °C for 30 minutes. After incubation, 600 uL of Reagent 1 was added to the washed leaf tissue, then shaken vigorously to mix the tissue and reagent, and then ~200 uL of the mix was pipetted into a DNAse-free 1.7 mL microtube, dividing the whole sample into three separate tubes. 200 uL of Reagent 2 (potassium SDS reagent) was added, and the microtubes were gently inverted to mix, and then incubated in a 65 °C water bath for 10 minutes, with regular manual agitation during the 65 °C water bath. After the water bath, 100-150 uL of potassium acetate was added and left to cool on ice for 20 minutes. 500 uL of cold chloroform was added to the sample, and then shaken on a rotator for 10 minutes at room temperature, forming a milky mixture. The tube was then centrifuged at 1300 g for 10 minutes. The clear upper DNA phase was pipetted into a new tube without disturbing the lower phase. DNA precipitation was performed with addition of an equal volume of cold isopropanol, and then kept in a -10 °C freezer for 1 hour to improve DNA precipitation. Afterwards, the DNA was pelleted by centrifuging at 4000 g for 5 minutes, then washed with cold 70% ethanol, and centrifuged again at 4000 g for 5 minutes. The supernatant was discarded, and leftover supernatant was carefully pipetted out. This is followed by quickly drying the tube upside down on tissue, dripping off the remaining ethanol, and the DNA was resuspended in TE buffer. DNA quantity and quality was measured with the Nanodrop and Tapestation.

Extracted DNA from the bulked pools and Sarisha-14 were sent to Novogene, Beijing, China, for Whole Genome Sequencing. PCR free library preparation was performed, then 150 bp paired-end reads at 50X coverage was generated on the Illumina

NovaSeq X Plus sequencing platform. On average, 18 Gb of raw data was generated per sample.

5.2.5 *SNP identification*

Raw sequencing data from Novogene was trimmed using cutadapt (ver. 2.1.0), then aligned against the R-o-18 reference genome (EnsemblPlants release 59, https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-59/fasta/brassica_rapa_ro18/dna/Brassica_rapa_ro18.SCU_BraROA_2.3.dna.toplevel.fa.gz) using bwa mem command (ver. 0.7.17, Heng Li (2013)). SNP variant calling was performed on the aligned Sarisha-14 sample using Freebayes (ver. 1.3.8 ,Garrison and Marth (2012)), with R-o-18 reference genome as the reference genome, to identify SNPs unique to Sarisha-14 as the alternate parent. SNP variant calling was also performed on the extreme early (Pool 1, Table 5.2) and extreme late pools (Pool 4, Table 5.2). Using a custom bash script and the bcftools (ver. 1.2.1, Danecek et al. (2021)) isec command, the early and late flowering SNP variant files (.vcf) are filtered for biallelic SNPs with QUAL value over 20.

5.2.6 *Further bioinformatic analysis*

Sarisha SNP distribution

The SNP variants identified in Sarisha-14 using Freebayes was used to visualise the distribution of Sarisha-14 specific SNPs across the reference genome. SNP density was visualised using ggplot2's barplot function (ver. 3.5.2, Wickham (2011)), with each bar representing a window of 1Mb, and the height corresponds to the square-root transformed SNP count in each window.

SNP-index and Δ SNP-index calculations

The SNP variant files from Sarisha-14 and early and late flowering pools were analysed in R (ver. 4.5.0) to calculate and plot Δ SNP-index between the early and late flowering pools. The SNP variant files from the early/late flowering pools were filtered for matching SNP positions against Sarisha-14, with at least 20 reads for the alternate allele read depth.

SNP-index of the early and late flowering pool was calculated separately, by dividing the number of alternate alleles (alternative allele observations, AO) by the total number of reads observed (total read depth, DP) at each SNP position. The SNP-index values were then averaged by 50kb bins across the genome for visualisation with ggplot2 (ver. 3.5.2).

The Δ SNP-index was calculated by subtracting the binned SNP-index values of the late flowering pool from the early flowering pool. The SNP-index and Δ SNP-index were plotted using ggplot2.

SnpEff annotation

SnpEff (ver. 5.2) was installed locally from https://snpeff.odsp.astrazeneca.com/versions/snpEff_latest_core.zip (Cingolani et al. 2012) using the wget command. The SnpEff reference library for R-o-18 was downloaded using the command " java -jar snpEff.jar download Brassica_rapa_ro18" for SnpEff annotations of the R-o-18 genome. The filtered early and late flowering SNP variant files were then annotated using the command "java -Xmx8g -jar snpEff.jar Brassica_rapa_ro18 input.vcf > output.vcf". Only annotated SNPs within regions of interest were selected for downstream analysis.

Candidate gene expression

time series gene expression data from Calderwood, Hepworth, et al. (2021) was accessed from JIC's internal BRAVO project directory. This dataset is also available in the NCBI Sequence Read Archive (SRA), with the project ID PRJNA593493 (Calderwood, Hepworth, et al. 2021). This dataset contains RNA-seq data from shoot apex and leaf tissue sampled at regular timepoints from Sarisha-14 and R-o-18 throughout development. For Sarisha-14, 18 timepoints were sampled (every day from day 7 until day 25), and 14 timepoints for R-o-18 (every other day from day 9 until day 35).

The raw RNA-seq paired end reads were trimmed with Trimmomatic (ver. 0.39 (Bolger, Lohse, and Usadel 2014)), aligned to the R-o-18 reference genome (EnsemblPlants release 59) with HISAT2 (ver. 2.1.0 (D. Kim et al. 2019)), and readcounts were quantified using Stringtie (ver. 2.1.1 (Shumate et al. 2022)), which was all run automatically with a pipeline written by Dr Hugh Woolfenden (script available on gitlab: https://git.nbi.ac.uk/morris-group/bravo-scripts/-/tree/master/alignment_scripts). For gene expression abundance plots, library size normalised counts (CPM) were scaled with Trimmed Mean of M factors (TMM) to allow more accurate sample-to-sample gene expression abundance comparisons, and plotted with ggplot2 (ver. 3.5.2, Wickham (2011)).

5.3 RESULTS

5.3.1 *Phenotyping results from R-o-18 x Sarisha-14 F2 generation*

R-o-18 (later-flowering female parent) was crossed with Sarisha-14 (earlier-flowering male parent) to create an F1 population. 16 F1 seeds were grown in CER conditions and then selfed to generate the F2 population. F1 lines were genotyped by PCR using

SSR markers RA2F04 to identify Sarisha-14 crossing into the R-o-18 parent. Of the 16 lines, R-o-18 x Sarisha-14 lines 3, 4, 8, 11, 13, 14, 15 and 16 were successfully crossed, displaying a band between 100-200bp similar to the Sarisha-14 parent (Fig. 5.1). 850 seeds from selfed line 11, 13 and 14 were sown to generate the F2 population since they had the most seeds collected (333, 328 and 301 seeds respectively).

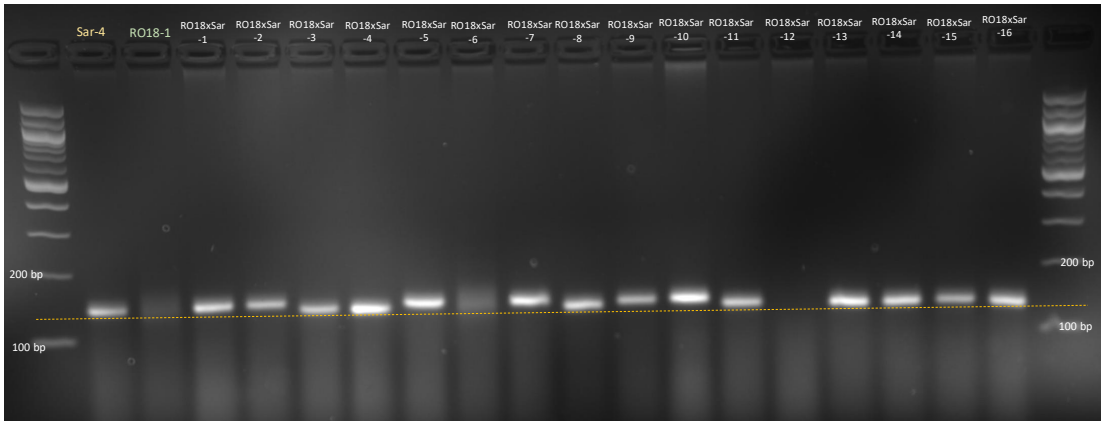


Figure 5.1: **Genotyping results for F1 population.** The yellow dotted line indicates the expected band weight for successful Sarisha-14 crosses.

The BSA mapping population of 850 F2 individuals were scored for days taken to reach developmental stage BBCH51, which is when the floral buds emerge and are visible to the naked eye (Meier et al. 2009). Figure 5.2 shows the distribution of the F2 population's phenotypes. Of the 850 plants, 816 had reached BBCH51, while the rest either failed to germinate, or failed to reach BBCH51 even after most plants have open flowers (past developmental stage BBCH60, first flower open (Meier et al. 2009)).

32 individuals (4% of the population) had an "extreme" early flowering phenotype with the average age at BBCH51 of 24 days after sowing (DAS), and were pooled together as pool 1, while pool 4 consisted of individuals with an "extreme" late flowering phenotype, with the average individual reaching BBCH51 at 33 DAS (21 individuals, 3% of population) (Fig. 5.2, Table 5.2). Although not directly comparable, Sarisha-14 grown in CER conditions reached BBCH51 at 24 DAS, while R-o-18 grown in the CER reached BBCH51 at 37 DAS. This strongly suggests that the F2 population flowering time segregated due to successful crossing between Sarisha-14 and R-o-18, and most of the population flowered earlier than R-o-18 due to inheriting alleles from Sarisha-14 associated with earlier flowering.

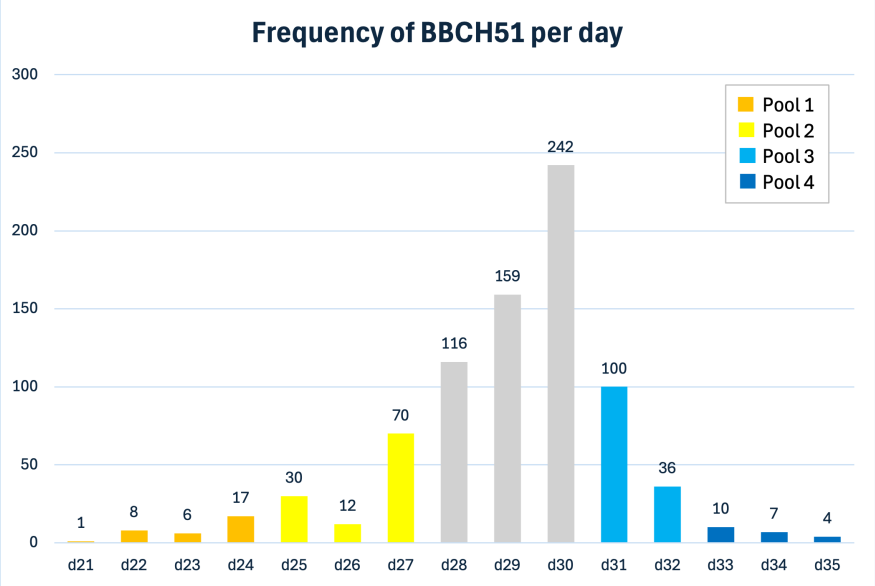


Figure 5.2: **Distribution of days to BBCH51 in the F2 crossing population of R-o-18 (female parent) and Sarisha-14 (male parent).** Pool 1 are bulked individual plants with "extremely early" BBCH51 phenotype, while Pool 4 are bulked individuals with "extremely late" BBCH51 phenotype.

Pool	Pool 1	Pool 2	Pool 3	Pool 4
Phenotype	Extreme early	Intermediate early	Intermediate late	Extreme late
Plant age at BBCH51 (DAS)	21-24	25	32	33-35
Number of individuals	22	30	36	18
% of population	2.7%	3.7%	4.4%	2.2%

Table 5.2: **Description of each pool sent for sequencing for bulk segregant analysis.**

5.3.2 *SNP comparisons between early flowering and late flowering pool*

Leaf tissue from the F2 population were sampled and bulked into each pool, as well as pooled leaf tissue from Sarisha-14 plants for DNA extraction and subsequent whole genome Illumina DNA sequencing at 50X coverage. Genome sequences from the R-o-18 parent was not necessary as it could be represented using the *B. rapa* R-o-18 reference genome (EnsemblPlants release 59).

The DNA sequencing data for Sarisha-14 and the 4 pooled samples was quality checked with FastQ and then aligned against the R-o-18 reference genome (EnsemblPlants release 59). SNP calling was performed using Freebayes (Garrison and Marth 2012). For the rest of this analysis, I will focus on only the extremely early flowering and extremely late flowering pools, which represents a total of ~5% of the population (Table 5.2). This is to increase the likelihood of detecting SNPs in loci which contribute to the contrasting phenotypes between the pools.

Freebayes called a total of 2305889 SNP variants in the Sarisha-14 genome, but after selecting for SNPs which were: 1) biallelic (detected SNPs had only one alternate allele compared to the reference), 2) SNP quality threshold (QUAL) $\geqslant$ 20, and 3) at least 20 reads detected (DP $\geqslant$ 20), only 935784 of Sarisha-14 SNPs passed (40 % of total SNPs passed). The distribution of these SNPs were found across the whole genome, with some 500kb regions with higher SNP density than average, particularly chr A01, A03, A06, A07 and A09 (Fig. 5.3A). This confirms divergence in Sarisha-14 from the R-o-18 genome, which may contribute to the difference in flowering time phenotype.

Freebayes SNP calling for the extremely early flowering pool identified a total of 1726742 SNPs, of which 413621 (24 %) SNPs passed the quality filtering and shared the same SNP positions with Sarisha-14. Meanwhile, the late flowering pool had a total of 1430282 SNPs, but only 401064 (28 %) passed quality filtering and intersected with Sarisha-14 SNP positions.

The SNP-index is a measure of the contribution of the reference (R-o-18) and alternate (Sarisha-14) parent allele in each SNP position. This is calculated by dividing the total observations of alternate alleles in each pool against the sum of all allele observations (alternate allele + reference allele from the R-o-18 reference genome). A SNP-index of 0.50 indicates equal contribution from both parents, and in Fig. 5.3B, SNP-index of 1.0 indicates all sequenced alleles came from the Sarisha-14 parent, while SNP-index of 0 means all observed alleles are from the R-o-18 parent.

Across the whole genome, there is a clear divergence in SNP-index between the early and late flowering pool, bar a few chromosomes such as chr A04, A06, A07 where the SNPs have near 50/50 parental allele contributions (Fig. 5.3B). The divergence in the pools suggests that there are parent-specific SNPs which contribute to early and late flowering time. We can calculate the difference in SNP-index by subtracting the SNP-index of the late pool from the early pool, which gives the Δ SNP-index (Fig. 5.3C). A peak in Δ SNP-index above a set threshold of $\pm$ 0.45 highlights regions with a significant difference in SNP-index between the pools, which should have candidate genes that contribute to the difference in flowering time. Despite the divergence of SNP-index across almost the whole genome between the two pools, only three chromosomes stand out with Δ SNP-index above the threshold, which are chromosomes A01, A02 and A03. Thus, the following analysis is focused on these three chromosomes only.

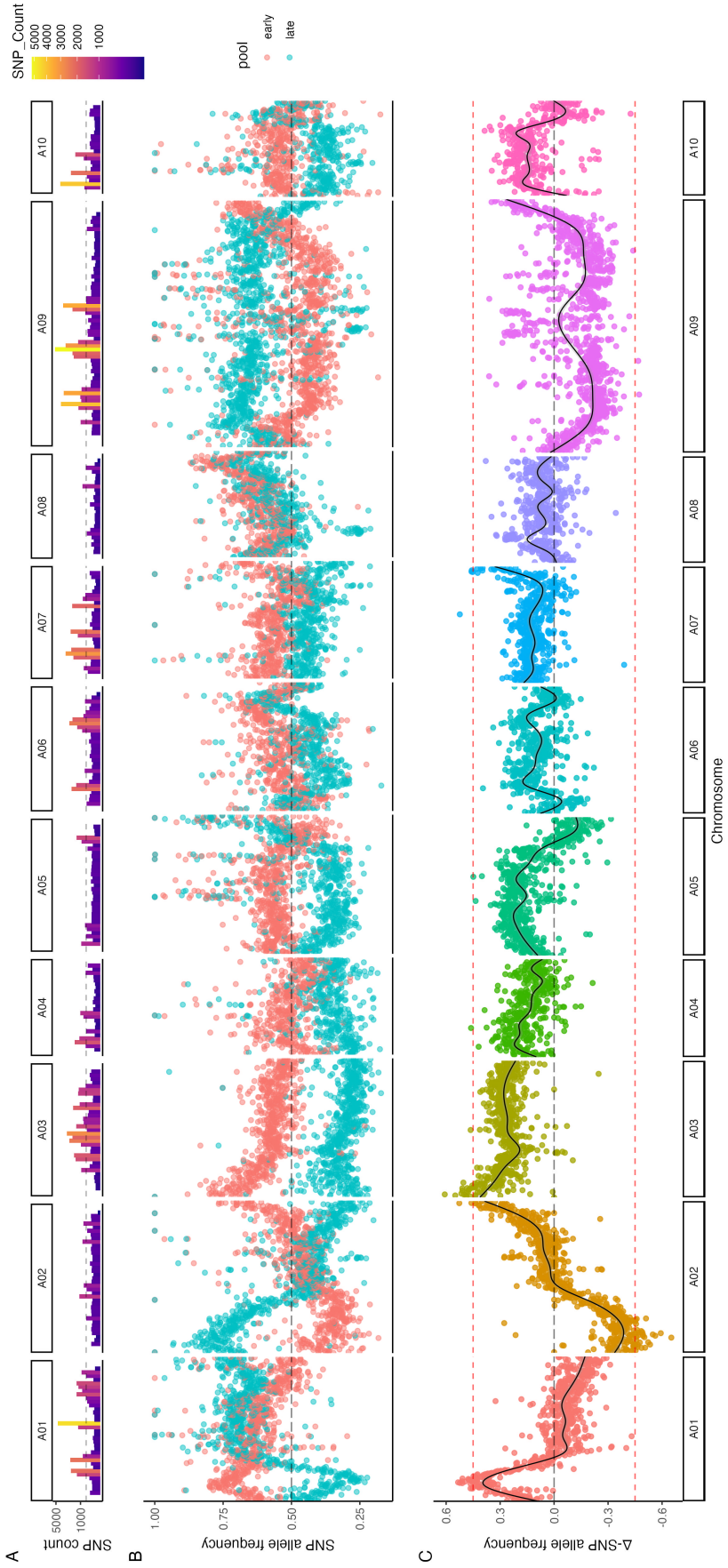


Figure 5.3: **Distribution of SNPs associated with flowering time (days to BBCH51).** (Figure legend continues in the next page)

Figure 5.3 (continued)
(A) Distribution of Sarisha-14 SNPs across the R-o-18 reference genome. Each column represents the number of SNPs per 500 kb, and is square-root transformed for visualization purposes. The gray dashed line denotes the mean SNP count across the genome (512 SNPs per 500kb). Sarisha-14 SNPs are present across the whole genome, with no SNP-deserts (500 kb regions with 0 SNPs).
(B) SNP-index across the genome. Each dot represents the average SNP-index across a 500 kb window for the early (red) and late (blue) flowering pools. The black dashed line indicates SNP-index value of 0.5, where SNPs have equal contributions from both parents. Chromosome A01, A02 and A03 show the most divergence between the early and late flowering pool, which suggests genomic loci in these chromosomes may be associated with earlier flowering.
(C) Δ SNP-index across the genome. Each dot represents the difference in SNP-index between early and late flowering pools. The black dashed line indicates no difference between the pools. Smoothed trend lines were fitted using generalized additive models (GAMs) with cubic splines to capture broad-scale allele frequency patterns along the chromosome (shown with fitted black line). The red dashed lines indicate a Δ SNP-index threshold of 4.50 and -4.50, which is estimated to be around 90% of the variance of the Δ SNP-index. Genomic regions in chromosome A01, A02 and A03 have multiple SNPs beyond these thresholds, which suggests potential genomic loci linked to earlier flowering.

Chr	Region Start	Region End	Region length	No. genes in region	Total flowering genes	No. genes High imp. SNPs	No. genes Mod. imp. SNPs	Flowering genes w. SNPs
A01	3Mb	6Mb	3Mb	1709	23	18	329	1
A02	1.2Mb	8.2Mb	7Mb	1714	24	8	246	2
A03	0.2Mb	3.5Mb	3.3Mb	1948	28	25	257	4

Table 5.3: **Overview of genes with SNPs in regions of interest.** This table summarizes the total number of genes with high-impact SNPs and flowering genes found in chrA01, chrA02 and chrA03. This is within the first 10 Mb of the chromosomes, which contains a region of interest where the Δ SNP-index is higher than the set threshold of ± 0.45.

5.3.3 *Candidate flowering loci in chromosome A01*

There is a strong positive Δ SNP-index peak within the 3-6Mb region of chromosome A01 (Fig. 5.4). To identify potential genes within this region, I expanded my search to flanking 3Mb upstream and downstream of this peak. In the first 10 Mb of the

chromosome, there are 1709 genes and 6499 SNPs identified within this region (Table 5.3). To identify SNPs with a functional impact on its genomic loci, SnpEff (Cingolani et al. 2012) was used to annotate the predicted functional impact of the SNPs in this region. SnpEff annotated 60 SNPs as "high impact", where these SNPs caused downstream impacts on protein coding, either by gain of stop sites or splicing variants. These high impact SNPs occured within 18 genes in this region, out of a total of 1709 genes. There are also 23 known flowering genes within this region of chrA02, but only 1 gene, *A01p012440.1_BraROA*, an orthologue of Arabidopsis *SUSPENSOR2 (SUS2)*, contained 2 "moderate impact" SNPs, which were predicted to cause a missense variant of the gene (Table 5.4). Figure 5.5 shows the locations of the candidate loci with high impact SNPs, where *SUS2* is found within the high Δ SNP-index region. Out of these 19 genes, 10 of them showed detectable expression (Table 5.4, Fig. 5.6).

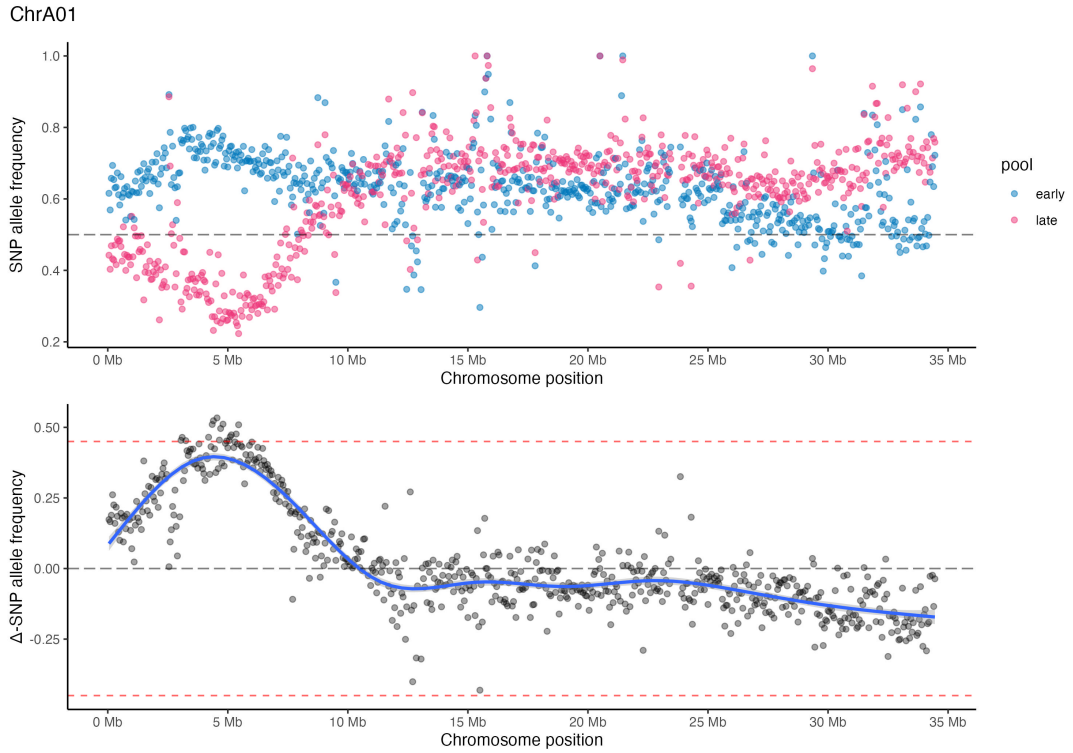


Figure 5.4: **SNP-index and Δ SNP-index plots across chromosome A01.**
SNP-index plot (top) shows changes in SNP allele frequency in both early and late flowering pools, and Δ SNP-index plot (bottom) highlights genomic regions with significant SNP-index differences between the pools. Detailed descriptions of these plots can be found in 5.3.2B and C. In chromosome A01, a distinct Δ SNP-index peak is observed around 3-6 Mb, which suggests there is a potential loci inherited from Sarisha-14 which contributes to earlier flowering in this region.

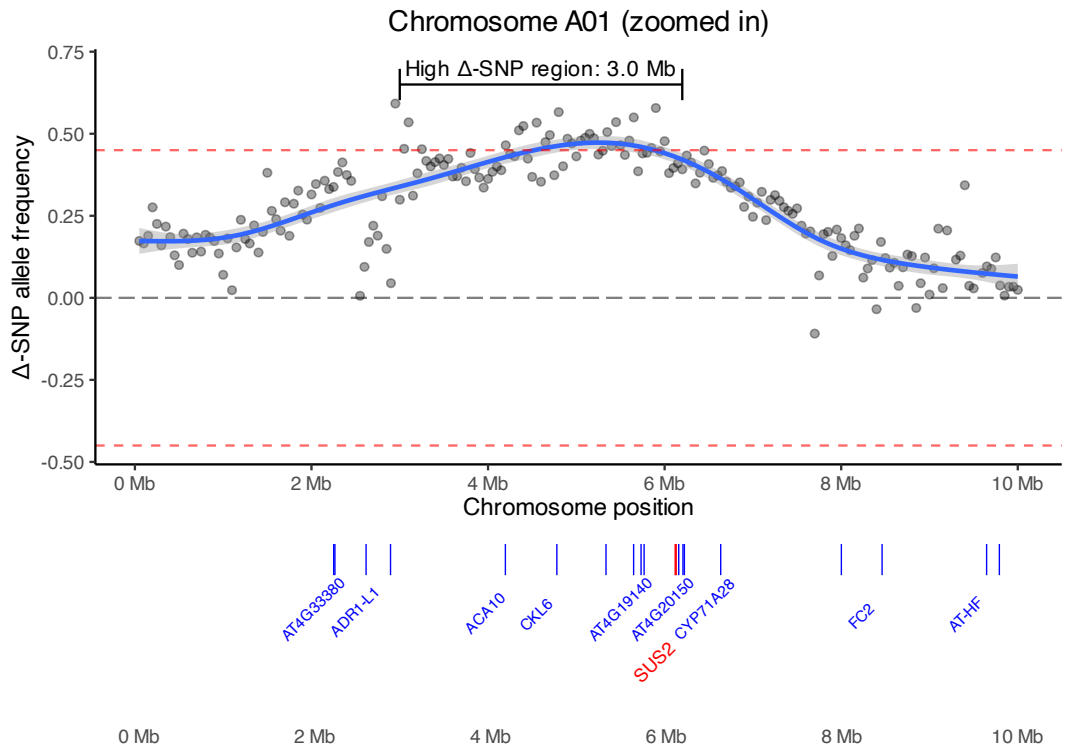


Figure 5.5: **A closer look at the first 10Mb of chrA01.** This figure shows the Δ SNP-index plot at the first 10Mb of chromosome A01, with the positions of candidate genes with high impact SNPs annotated with blue lines, and flowering gene *SUS2* in red. The location of *SUS2* within the high Δ SNP-index indicates that it could be a potential contributor to early flowering time.

Table 5.4: **List of candidate genes within chrA01 region of interest with high impact SNPs.** Only genes with detectable expression from either R-o-18 or Sarisha-14 are listed (Calderwood, Hepworth, et al. 2021).

BRAPA ID	POS	EFFECT	ATH ID	ATH NAME	NUM SNPS	ATH DESC
A01p012440.1 BraROA	6105582	missense variant	AT1G80070	SUS2	2	Pre-mRNA-processing-splicing factor
A01g501090.1 BraROA	4167202	stop gained	AT4G29900	ACA10	9	autoinhibited Ca(2+)-ATPase 10
A01p005190.1 BraROA	2212522	stop gained	AT4G33380	AT4G33380	2	dimethylallyl%2C adenosine tRNA methylthiotransferase
A01p005220.1 BraROA	2226678	stop gained	AT4G33300	ADR1-L1	1	ADR1-like 1
A01p009840.1 BraROA	4753123	splice donor variant & intron variant	AT4G28540	CKL6	3	casein kinase I-like 6

BRAPA ID	POS	EFFECT	ATH ID	ATH NAME	NUM SNPS	ATH DESC
A01p011790.1 BraROA	5711743	splice donor variant & intron variant	AT4G19140	AT4G19140	1	exopolysaccharide production negative regulator
A01p012610.1 BraROA	6188492	stop gained	AT4G20150	AT4G20150	2	excitatory amino acid transporter
A01p012650.1 BraROA	6204756	stop gained	AT4G20235	CYP71A28	2	cytochrome P450%2C family 71%2C subfamily A%2C polypeptide 28
A01p017080.1 BraROA	8455412	stop gained	AT4G24740	FC2	1	LAMMER-type protein kinase AFC2
A01p019660.1 BraROA	9645197	stop gained	AT4G26900	AT-HF	4	HIS HF

Using R-o-18 and Sarisha-14 time series gene expression data from Calderwood, Hepworth, et al. (2021), we can observe the expression pattern of these candidate genes in Sarisha-14 and R-o-18 shoot apex tissue during development.

Of the 10 genes with high impact SNPs that are expressed in Sarisha-14 and R-o-18, 5 of these genes show divergent expression patterns between the cultivars (Fig. 5.6). This could suggest that besides impact on protein function, the SNPs may also influence the regulation of gene expression, or the stability of mRNA transcript during development. For example, *CKL6*, *AT-HF* and *CYP71A28* are expressed higher in Sarisha-14 compared to R-o-18 consistently throughout development, while the reverse is observed for the orthologue of *AT4G20150* (Fig. 5.6). However, this remains speculative since the SNPs within these genes are predicted to disrupt protein translation due to a gain in STOP codon (Table 5.4).

SUS2, the only flowering orthologue identified with a missense variant SNP in this region, is also expressed in both the shoot apex and leaf tissue (Fig. 5.6). This gene orthologue is expressed at similar levels in the shoot apex of both cultivars, but in the leaf, *SUS2* expression is markedly higher in R-o-18 than Sarisha-14, where *SUS2* expression peaks at around 150 TMM compared to 100 TMM in Sarisha-14.

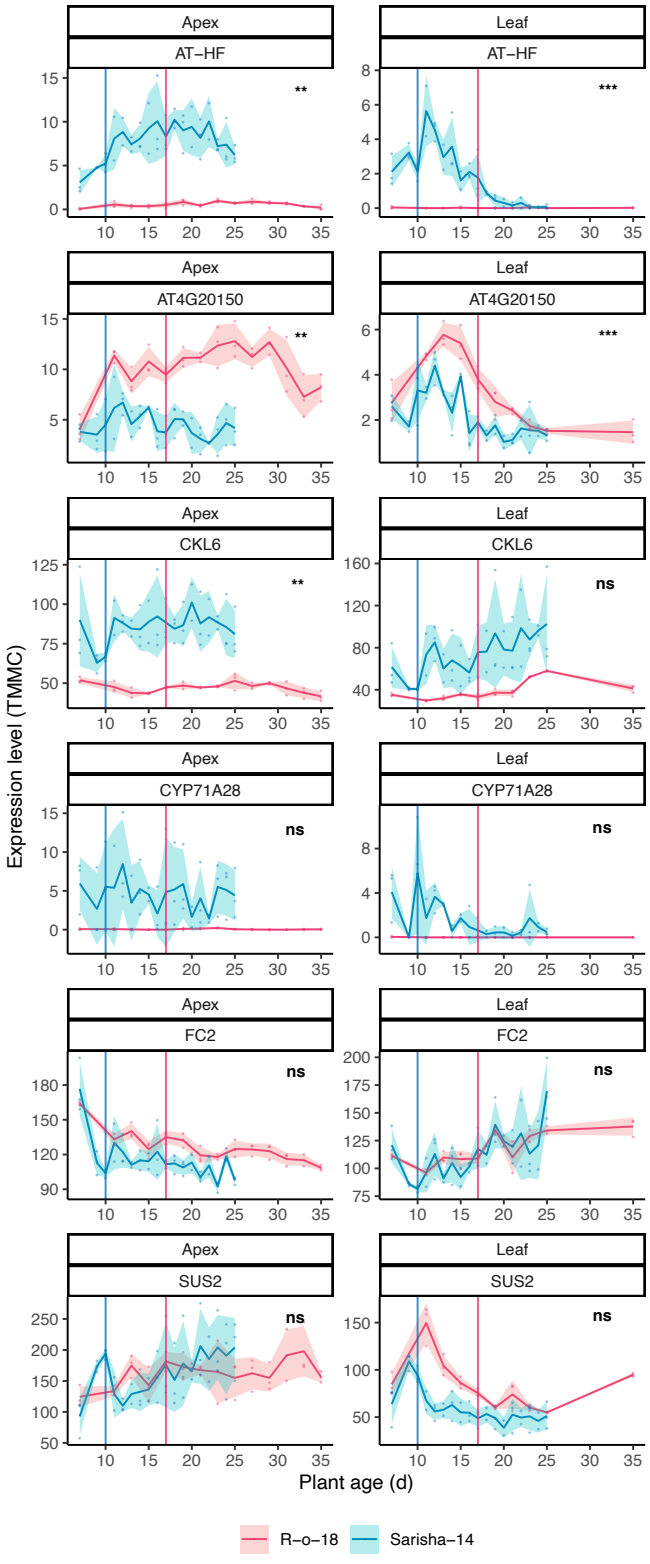


Figure 5.6: **Expression of selected candidate genes influencing flowering time in chrA01.** Time series expression of candidate genes throughout development in R-o-18 (red) and Sarisha-14 (blue). Vertical lines indicate first day of floral transition in R-o-18 (red) and Sarisha-14 (blue). P-values in the top-right corner show whether temporal expression patterns differ between genotypes (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, ns = no significant difference; results from likelihood ratio test for genotype × time spline interaction).

5.3.4 *Candidate flowering loci in chromosome A02*

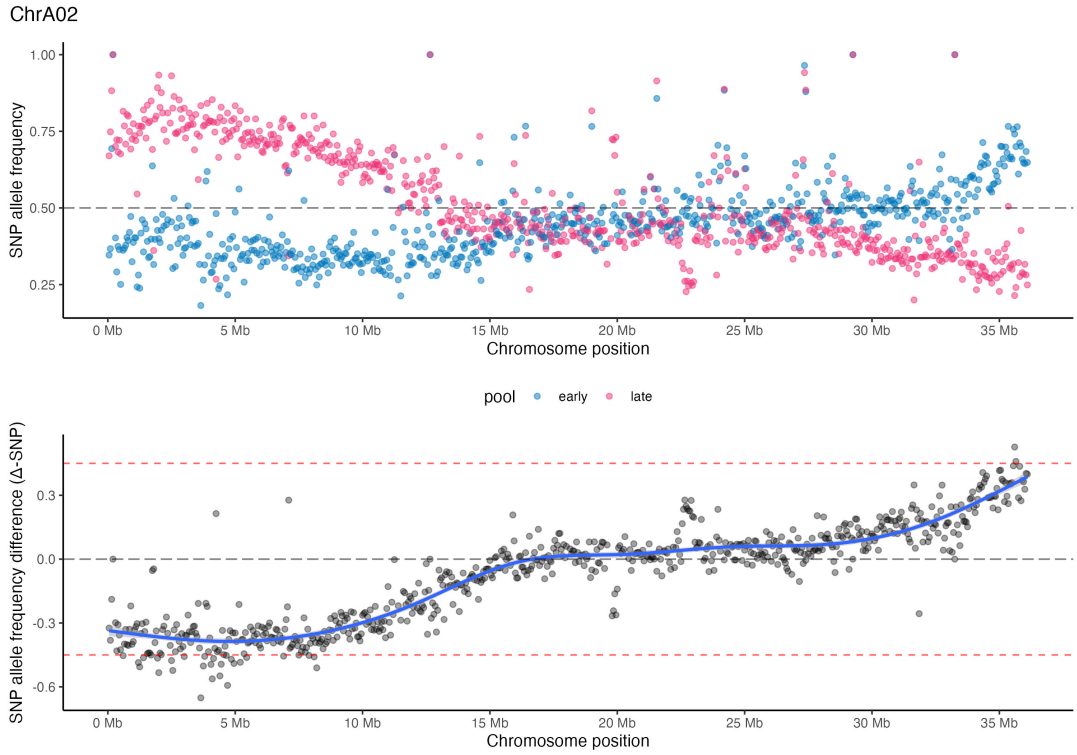


Figure 5.7: **A broad negative Δ SNP-index peak near the beginning of chromosome A02 indicates genomic loci inherited from R-o-18 which contribute to earlier flowering.** SNP-index (top) and Δ SNP-index (bottom) plots across chromosome A02. Detailed descriptions of these plots can be found in 5.3.2B and C.

In chromosome A02, there is a broad negative Δ SNP-index peak from 1.2Mb to 8.2Mb in the chromosome, which is in contrast to chromosome A01's strong positive Δ SNP-index peak (Fig. 5.7, 5.8). This suggests that the earlier flowering phenotype is inherited from alleles from the R-o-18 (reference) parent, instead of Sarisha-14. There are a total of 1714 genes within the first 10Mb of this chromosome, of which only 8 genes contained high impact SNPs, as well as 2 flowering gene orthologues with SNPs (Table 5.3, Fig. 5.8). However, only 6 of these high impact genes were expressed (Table 5.5, Fig. 5.9), as well as the 2 flowering genes, *A02p004560.1_BraROA (EMF1)* and *A02p005360.1_BraROA (PBS3)*.

These genes display similar expression patterns in the same tissues of both Sarisha-14 and R-o-18, except for *A02p005360.1_BraROA (PBS3)* and *A02p019100.1_BraROA*. *A02p019100.1_BraROA* expression in R-o-18 is lower and not as variable as Sarisha-14, and *PBS3* also has lower expression in R-o-18 than Sarisha-14. These two genes could be floral repressors that are non-functional in R-o-18, thus these alleles inherited from R-o-18 could allow the F2 individuals to flower earlier.

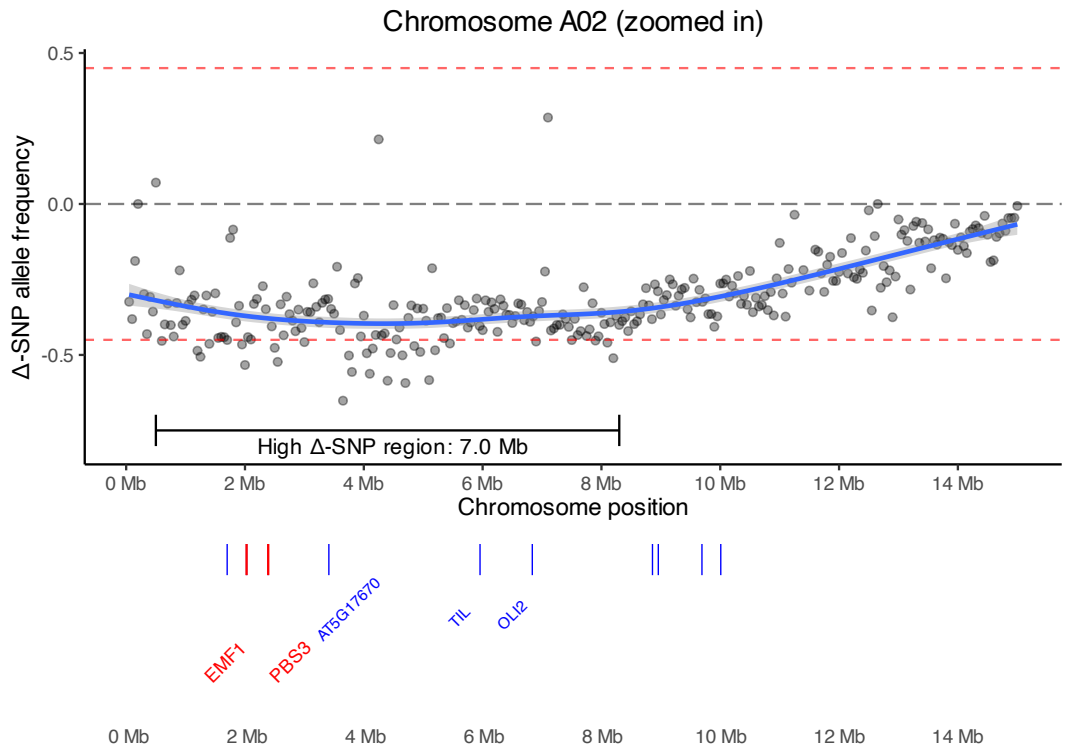


Figure 5.8: **A closer look at the first 10Mb of chrA02.** This figure shows the Δ SNP-index plot at the first 10Mb of chromosome A02, with the positions of candidate genes with high impact SNPs annotated with blue lines, and flowering genes *EMF1* and *PBS3* in red. The location of *EMF1*, *PBS3*, and the orthologue of *AT5G17670* within the high Δ SNP-index suggests that R-o-18 inherited alleles within these genes could influence early flowering time.

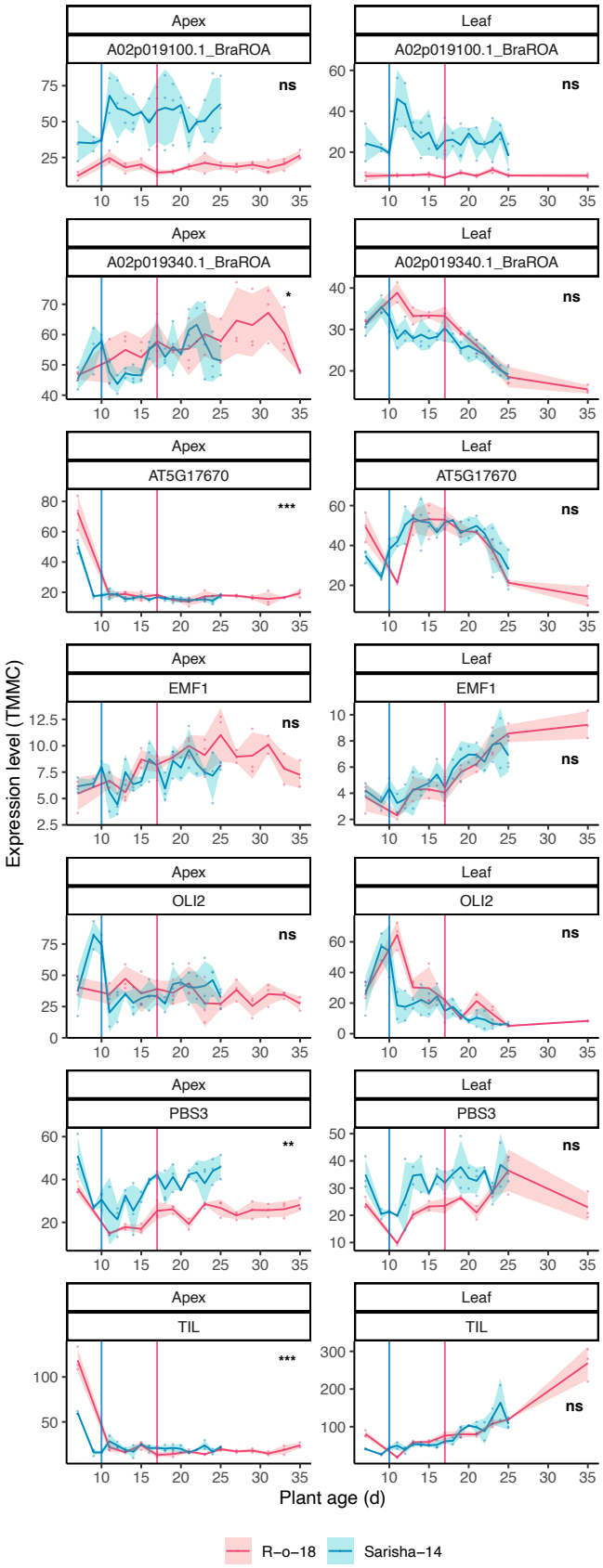


Figure 5.9: **Expression of candidate genes influencing flowering time in chrA02.** Time series expression of candidate genes throughout development in R-o-18 (red) and Sarisha-14 (blue). Vertical lines indicate first day of floral transition in R-o-18 (red) and Sarisha-14 (blue). P-values in the top-right corner show whether temporal expression patterns differ between genotypes (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, ns = no significant difference; results from likelihood ratio test for genotype × time spline interaction).

Table 5.5: **List of candidate genes within chrA02 region of interest with high impact SNPs.** A few of these genes have no predicted orthologues in Arabidopsis, but their gene IDs and SNP information are still included in this table. Only genes with detectable expression from either R-o-18 or Sarisha-14 are listed (Calderwood, Hepworth, et al. 2021).

BRAPA ID	POS	EFFECT	ATH ID	ATH NAME	NUM SNPS	ATH DESC
A02p004560.1_BraROA	1983249	missense variant	AT5G11530	EMF1	1	embryonic flower 1 (EMF1)
A02p005360.1_BraROA	2348821	missense variant	AT5G13320	PBS3	2	Auxin-responsive GH3 family protein
A02p008130.1_BraROA	3373136	stop gained	AT5G17670	AT5G17670	1	alpha/beta-Hydrolases superfamily protein
A02p013520.1_BraROA	5926157	stop gained	AT5G58070	TIL	3	temperature-induced lipocalin
A02p015300.1_BraROA	6808892	splice donor variant & intron variant	AT5G55920	OLI2	7	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
A02p019100.1_BraROA	8838742	splice donor variant & intron variant	NA	NA	1	NA
A02p019340.1_BraROA	8937088	splice donor variant & intron variant	NA	NA	1	NA
A02p021120.1_BraROA	9992079	stop gained	NA	NA	1	NA

5.3.5 *Candidate flowering loci in chromosome A03*

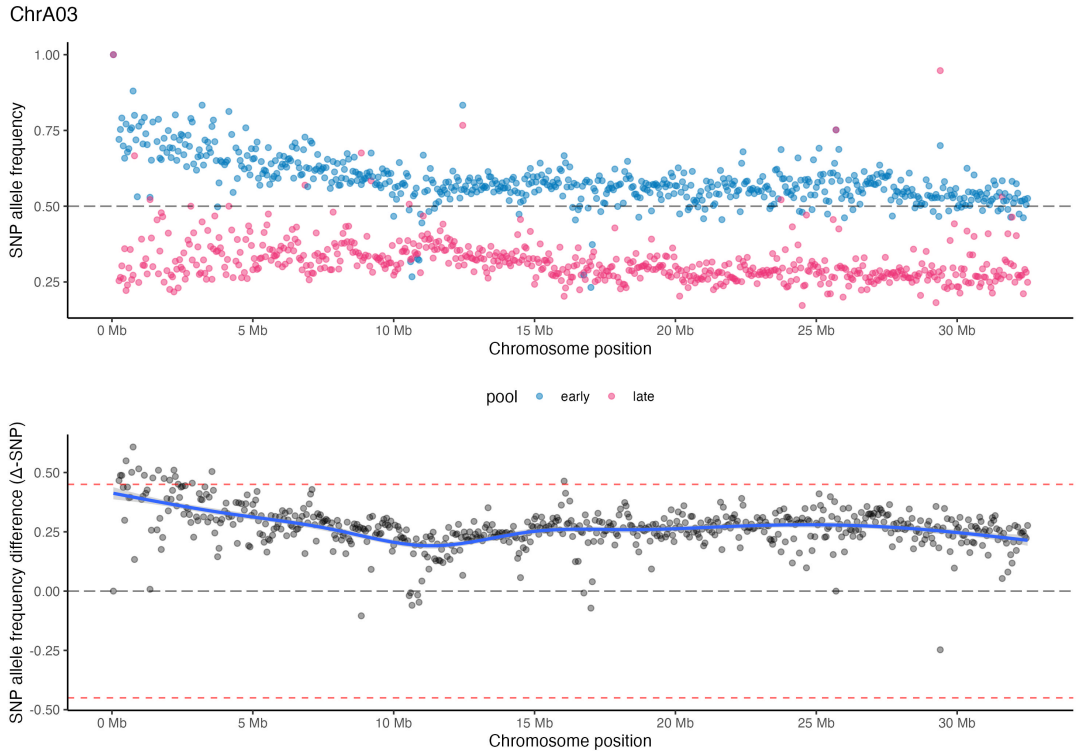


Figure 5.10: **A broader Δ SNP-index peak at the beginning of chromosome A03 indicates another potential loci controlling flowering time closer to the start of this chromosome.** SNP-index (top) and Δ SNP-index (bottom) plots across chromosome A03. Detailed descriptions of these plots can be found in 5.3.2B and C.

Similar to chrA02, there is also a broad peak at the beginning of chrA03, however, it is a positive Δ SNP-index peak, indicating that SNPs inherited from Sarisha-14 within this region influences earlier flowering time. More SNP groups (visualised as gray dots in the Δ SNP-index plot (bottom) of Fig.5.10) are found above the +0.45 threshold at the beginning of chrA03, which suggests potential loci associated with earlier flowering are found here. A total of 25 "high impact" candidate genes are found in the first 10Mb of this region, in addition to 4 flowering time orthologues (*TFL1, EMF1, RD20, NF-YB1*) (Table 5.3, Table 5.6).

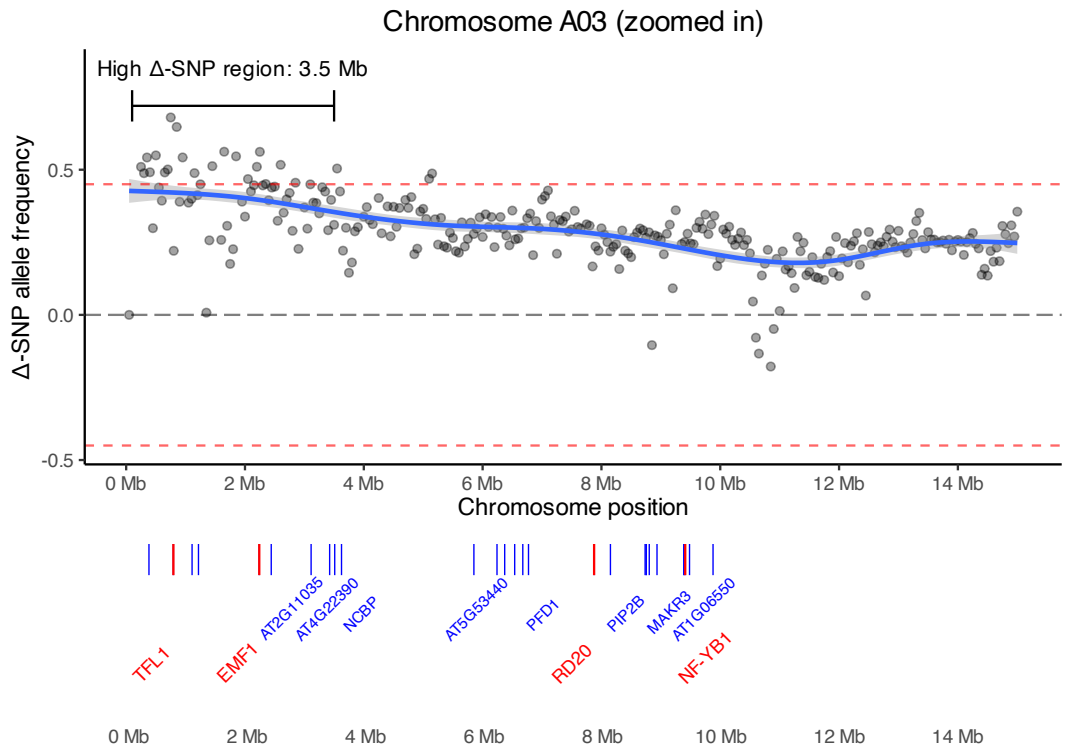


Figure 5.11: **A closer look at the first 10Mb of chrA03.** This figure shows the Δ SNP-index plot at the first 10Mb of chromosome A01, with the positions of candidate genes with high impact SNPs annotated with blue lines, and flowering gene *SUS2* in red. The location of *SUS2* within the high Δ SNP-index indicates that it could be a potential contributor to early flowering time.

Several of these candidate genes had lower expression in Sarisha-14 than R-o-18, such as *A03p022300.1_BraROA*, *A03p021860.1_BraROA (AT2G38330)*, *A03p022380.1_BraROA (NF-YB1)*, and an orthologue of *TFL1*, *A03p001500.1_BraROA*.

TFL1.A03 stands out because it is the only flowering gene annotated with a high impact SNP, leading to a loss of a start codon (Table 5.6). The genomic coordinates of *TFL1.A03* shows that it is located near the start of chrA03, where there is a high frequency of SNPs associated with earlier flowering time (Fig. 5.11). This points to *TFL1.A03* to being a strong candidate for contributing to earlier flowering time in Sarisha-14, which I expand further in the discussion.

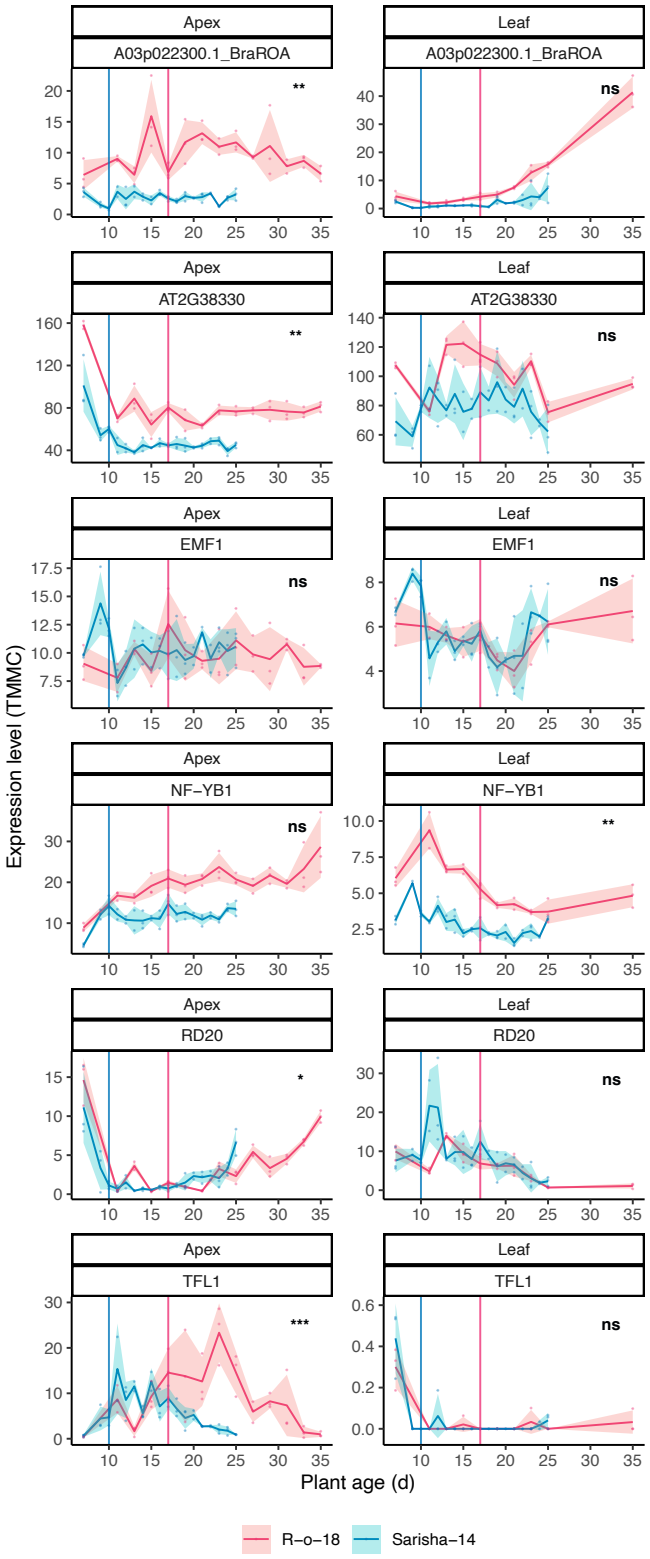


Figure 5.12: **Expression of potential genes influencing flowering time in chrA03.** time series expression of candidate genes throughout development in R-o-18 (red) and Sarisha-14 (blue). Vertical lines indicate first day of floral transition in R-o-18 (red) and Sarisha-14 (blue). P-values in the top-right corner show whether temporal expression patterns differ between genotypes (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, ns = no significant difference; results from likelihood ratio test for genotype × time spline interaction).

Table 5.6: **List of genes which contained high-impact SNPs and flowering genes with high or moderate-impact SNPs within chrA03 region of interest.** Only genes with detectable expression from either R-o-18 or Sarisha-14 are listed (Calderwood, Hepworth, et al. 2021).

BRAPA ID	POS	EFFECT	ATH ID	ATH NAME	NUM SNPS	ATH DESC
A03p001500.1_BraROA	747008	start lost	AT5G03840	TFL1	1	PEBP (phosphatidyl-ethanolamine binding protein) family protein
A03p005240.1_BraROA	2197714	missense variant	AT5G11530	EMF1	3	embryonic flower 1 (EMF1)
A03p019260.1_BraROA	7852873	missense variant	AT2G33380	RD20	1	Caleosin-related family protein
A03p022380.1_BraROA	9391686	missense variant	AT2G38880	NF-YB1	19	nuclear factor Y%2C subunit B1
A03g502800.1_BraROA	8734436	stop lost & splice region variant	AT2G37170	PIP2B	1	plasma membrane intrinsic protein 2
A03g502800.1_BraROA	8734436	stop lost & splice region variant	AT2G37180	RD28	1	Aquaporin-like superfamily protein
A03p002580.1_BraROA	1170856	splice donor variant & intron variant	NA	NA	1	NA
A03p007420.1_BraROA	3073240	splice acceptor variant & intron variant	AT4G22390	AT4G22390	3	F-box associated ubiquitination effector family protein
A03p008970.1_BraROA	3586793	stop gained	AT5G18110	NCBP	1	Putative cap-binding protein
A03p015710.1_BraROA	6212878	stop gained	AT5G53440	AT5G53440	1	LOW protein: zinc finger CCCH domain protein
A03p016380.1_BraROA	6511492	stop gained	NA	NA	2	NA
A03p016650.1_BraROA	6648382	stop gained	AT2G07340	PFD1	1	PREFOLDIN 1
A03p019830.1_BraROA	8128461	start lost	AT2G34750	AT2G34750	2	RNA polymerase I specific transcription initiation factor RRN3 protein
A03p021270.1_BraROA	8784314	stop lost & splice region variant	AT2G37380	MAKR3	10	membrane-associated kinase regulator

BRAPA ID	POS	EFFECT	ATH ID	ATH NAME	NUM SNPS	ATH DESC
A03p021860.1_BraROA	8915159	stop lost & splice region variant	AT2G38330	AT2G38330	1	MATE efflux family protein
A03p022270.1_BraROA	9367059	stop gained	AT1G06550	AT1G06550	1	ATP-dependent caseinolytic (Clp) protease/crotonase family protein
A03p022300.1_BraROA	9370240	stop gained	AT2G38800	AT2G38800	1	NA
A03p022530.1_BraROA	9464280	splice donor variant & intron variant	AT2G28290	SYD	6	P-loop containing nucleoside triphosphate hydrolases superfamily protein

5.3.6 *Expression profiles of all TFL1 copies between Sarisha-14 and R-o-18*

B. rapa has three homologues of *TFL1*, one in chrA02, chrA03 and chrA10 respectively. All three copies were expressed in the shoot apex tissue of both cultivars, and were almost undetectable in the leaf tissue (Fig. 5.13). The chrA03 copy which was annotated with a high impact SNP has the highest expression level in both Sarisha-14 and R-o-18 (Fig. 5.13) compared to the other copies. Based off just the expression abundance, this could suggest that *TFL1.A03* plays a stronger role in repressing flowering compared to the other copies.

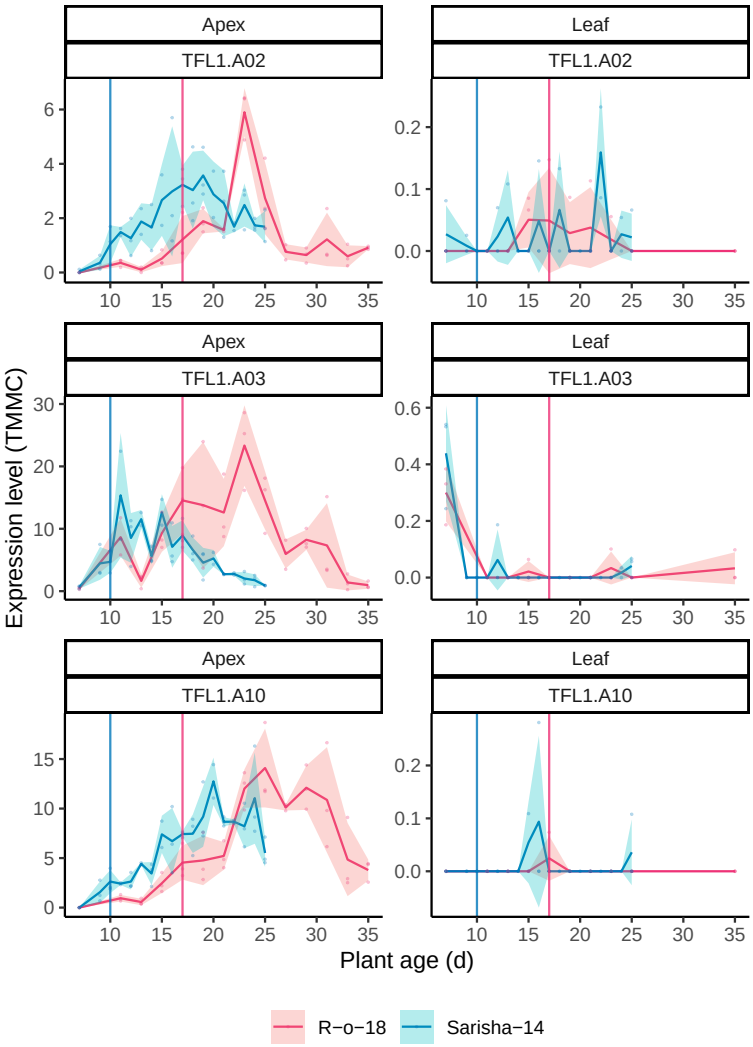


Figure 5.13: ***TFL1.A03* is the only copy with a high-impact SNP within a region of high SNP divergence.** Comparison of all *TFL1* copies between R-o-18 and Sarisha-14. time series expression of all *TFL1* copies in R-o-18 and Sarisha-14. Data from Calderwood, Hepworth, et al. (2021).

In conclusion, this bulk segregant analysis has identified three genomic regions of interest in chrA01, chrA02 and chrA03 displaying allele divergence between the early and late flowering pooled F2 R-o-18 x Sarisha-14 progenies. From these three genomic regions, I identified 7 orthologues of known flowering time genes, alongside 50 other candidate genes containing high impact SNPs.

5.4 DISCUSSION

5.4.1 *SNPs in TFL1.A03 could influence Sarisha-14's earlier flowering phenotype*

TFL1 is an evolutionarily conserved floral repressor in many plant lineages such as woody perennials, grasses and ornamental species (Wickland and Hanzawa 2015). It is a member of the phosphatidylethanolamine binding protein (PEBP) family, where it forms a repressive complex with FD transcription factor to repress floral meristem identity genes *APETALA1 (AP1), SEPETALLA1 (SP1), SP3, FRUITFUL (FUL)* and *LEAFY (LFY)* (Hanano and Goto 2011). Transgenic Arabidopsis overexpressing *TFL1* have extended vegetative and reproductive stages (Ratcliffe et al. 1998), pointing to *TFL1*'s repressive role in delaying flowering time.

This BSA found that 1 SNP in the *TFL1.A03* orthologue was predicted to cause a loss in a start codon of this gene. This would suggest that in the early flowering pool, the *TFL1.A03* gene was non-functional, leading to reduced floral repression, contributing to their early flowering phenotype.

5.4.2 *SNPs in other flowering gene orthologues identified in the BSA could contribute to early flowering.*

EMF1

EMF1 is a key floral repressor in *A. thaliana*, maintaining the vegetative stage and preventing precocious flowering. It functions as a component of the Polycomb Group (PcG) repressive complex, targeting the floral homeotic gene *AGAMOUS (AG)* (Picó et al. 2015), and also *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE (SPL)* genes and *MIR172* genes (Picó et al. 2015), which are both floral promoters. Loss-of-function *emf1* mutants flower extremely early, producing abnormal inflorescence meristems soon after germination (Sung et al. 1992). This highlights *EMF1*'s essential role in regulating flowering time, and could tie into the regulation of flowering time in rapid-cycling *B. rapa*.

Two orthologues of *EMBRYONIC FLOWER 1 (EMF1)*, *A02p004560.1_BraROA* and *A03p005240.1_BraROA*, were identified in the BSA. Both copies of *EMF1* displayed a rapid upregulation prior to the floral transition in Sarisha-14, and R-o-18 also went through the floral transition at similar expression levels (3 TMMC for *EMF1_A02*, and 4 TMMC for *EMF1_A03*).

SUS2

SUS2 encodes a core splicesosome subunit PRP8, which is necessary for the correct splicing of *COOLAIR* transcripts, which represses *FLC* transcription (Schwartz, Yeung, and Meinke 1994). Hypomorphic mutations of *SUS2* delayed flowering in Arabidopsis,

due to unrepressed *FLC* transcripts by alternatively splices *COOLAIR* (Schwartz, Yeung, and Meinke 1994).

In rapid-cycling *B. rapa*, *FLC* copies are non-functional, thus it lacks a vernalization requirement. It is possible however, that *BraFLC* remains repressed via SUS2-dependent repression by *COOLAIR*. Transcriptome profiling shows that *SUS2* is expressed at 10 TMMC in the vegetative stage (day 7) in both cultivars, and the floral transition coincides with a *SUS2* expression level of around 25 TMMC (Fig. 5.6). The gene expression profiles suggest that *SUS2* expression may contribute to both rapid-cycling cultivar's repression of *FLC*, leading to its lack of vernalization requirement, but its role in regulating flowering time is yet to be known.

RD20

Another gene with similar expression patterns between Sarisha-14 and R-o-18 is *A03p019260.1_BraROA*, an orthologue of *Responsive to Dehydration/Dessication 20* (*RD20*, AT2G33380). This gene encodes a stress-responsive caleosin protein induced by drought, salt and abscisic acid (ABA) (Takahashi et al. 2000). Interestingly, *RD20* overexpression mutants also present early flowering phenotypes, and it is suggested that oxylipins produced by *RD20* as a stress response interacts with the gibberellin-signalling pathway to repress flowering in stress conditions (Blée et al. 2014).

The temporal expression pattern of *RD20* in Sarisha-14 and R-o-18 is also interesting, as it first expressed highly at around 12-15 TMMC in the vegetative stage, and then decreases rapidly at day 10 in both cultivars, followed by a slow increase back to initial levels by the BBCH51 stage. It could be that *RD20* represses flowering during the vegetative stage, but as the floral shoot apex begins to form inflorescence meristems, *RD20* plays another unknown role perhaps related to stress response. Its expression pattern in rapid-cycling *B. rapa* suggests that downregulation of this gene is also crucial for the timing of the floral transition.

NF-YB1

A03p022380.1_BraROA is considered a putative orthologue of Arabidopsis *NUCLEAR FACTOR Y beta subunit 1 (NF-YB1)*, which is also known as *HEME ACTIVATOR PROTEIN 3a (HAP3a)*. This gene is part of the Nuclear Factor Y (NF-Y) gene family which regulates many aspects of plant development, including flowering time (Siriwardana et al. 2016; Ben-Naim et al. 2006; Wenkel et al. 2006). HAP3a is suggested to act antagonistically with CONSTANS (CO), a floral promoter which controls flowering in response to changes in photoperiod, by sequestering CO into inactive complexes, which leads to FT reduction (Wenkel et al. 2006). Overexpressing HAP3a in Arabidopsis causes a delayed flowering phenotype and reduced FT expression, but no reduction in CO expression (Wenkel et al. 2006).

In agreement with Arabidopsis studies, it seems that *A03p022380.1_BraROA*'s higher expression in R-o-18 may contribute to its later flowering. It is also possible that SNPs in this gene may lead to reduced repressive function in Sarisha-14, which may explain how it is still able to progress to the floral transition even when *HAP3a* orthologue is expressed at a similar level as R-o-18 on the same timepoint.

Another HAP3 subunit, HAP3b, is shown to positively regulate flowering time in Arabidopsis - overexpression of this gene promotes early flowering, while *hap3b* null mutant displayed delayed flowering phenotype (Cai et al. 2007). It is also possible that in Sarisha-14, *A03p022380.1_BraROA* is more similar functionally to *HAP3b* than *HAP3a*, which could promote earlier flowering in Sarisha-14. Sequence homology and protein characterisation for *A03p022380.1_BraROA* could provide more insight on its regulatory mechanism on floral transition in rapid-cycling *B. rapa*.

5.4.3 *Candidate genes with early flowering alleles inherited from R-o-18 in chromosome A02*

This bulk segregant analysis identified a region of interest in the beginning of chromosome A02 where R-o-18 alleles are associated with earlier flowering phenotype (Fig.5.7). Two candidate genes,
A02p005360.1_BraROA and *A02p019100.1_BraROA*, are found within this region, and both genes are expressed lower in R-o-18.

A02p005360.1_BraROA is another predicted orthologue of a known floral repressor, *AvrPphB susceptible3 (PBS3)*. *PBS3* is known for its role in salicylic acid biosynthesis and plant immunity, but it also functions as a negative regulator of flowering time. Loss-of-function *pbs3-2* Arabidopsis mutants flowered earlier than wild-type Arabidopsis, alongside increased floral promoter *FT* expression and reduction of floral repressor *FLOWERING LOCUS C (FLC)*. This flowering phenotype is independent of salicyclic acid biosynthesis, as salicylic acid mutants such as *sid2-1* and *pad4-1* do not show an early flowering phenotype (G.-F. Wang et al. 2011). A missense variant SNP in this gene inherited from R-o-18 could render this protein non-functional, therefore contributing to an earlier flowering phenotype.

A02p019100.1_BraROA is also very interesting as it is highly expressed in both cultivars, with around 80 TMMC peak expression in R-o-18, and around 200 TMMC expression in Sarisha-14 at its peak. *A02p019100.1_BraROA* function is not yet annotated, and sequence homology predictions with BLAST suggest similarities with secEsec61-gamma protein transport protein (AT4G24920 and AT5G50460). Current literature is lacking on its function in plants, nor in regulating flowering time/development, but it is described as an essential eukaryotic protein translocation complex (Pohlschröder et al. 1997).

Due to its location within an early-flowering associated region, and their different expression levels in both cultivars, these two candidate loci may act as floral repressors,

where SNPs within these genes altered its expression and impacted its protein function, thus contributing to an early flowering phenotype.

5.4.4 *Potential promoters of flowering in Sarisha-14*

A03p002580.1_BraROA, a potential orthologue of an Arabidopsis epigenetic regulator

The *B. rapa* gene *A03p002580.1_BraROA* had no predicted *A. thaliana* orthologue from EnsemblPlants, so its DNA sequence was manually BLASTed against the *A. thaliana* (TAIR10) protein database to identify potential protein orthologues. BLASTX alignment revealed an 86.7% alignment with the reverse strand of an Arabidopsis gene *Zuotin-Related Factor 1a (ZRF1)*, AT3G11450, with a predicted E-value of 2.1e-130. This protein is a DNA-binding protein which acts as a chromatin regulator, by promoting the deposition of silencing histone modifications H2Aub1 and H3K27me3 (Feng et al. 2016). Loss-of-function *zrf1a/b* Arabidopsis mutants had many developmental defects, such as seed germination delay, poor fertility and abnormal flower development (due to misregulation of floral homeotic genes), which emphasises the crucial role of this protein in reproductive development (Feng et al. 2016).

A03p002580.1_BraROA is more highly expressed in Sarisha-14 than R-o-18 after it undergoes the floral transition. In Sarisha-14 this gene's expression level ranges around 40-60 TMMC from day 11 until day 25, while in R-o-18 the range is between 30-45 TMMC from day 11 until day 37. It is possible that the increased expression of *ZRF1* potential orthologue in Sarisha-14 allowed for earlier upregulation of floral homeotic genes, thus driving earlier floral development.

Casein Kinase-Like 6 (CKL6) may contribute to accelerated growth in Sarisha-14

A01p009840.1_BraROA is predicted to be a *B. rapa* orthologue of Arabidopsis *Casein Kinase-Like 6 (CKL6)*, which is a member of the *CKL* group of proteins. A study by Ben-Nissan et al. (2008) showed no characterized function in regulating flowering time, but instead *CKL6* is involved in anisotropic cell growth and shape formation via microtubule binding and organization.

This gene is more highly expressed in Sarisha-14 by double the expression level in R-o-18 5.6, and although it may not directly regulate floral development, it may contribute to Sarisha-14's accelerated growth rate in general by promoting cell growth.

A putative histidine biosynthesis pathway is upregulated in earlier flowering Sarisha-14

A01p019660.1_BraROA is the putative orthologue of an Arabidopsis gene *Arabidopsis hisHF (AT-HF)* or alternatively known as *HISN4*. In Arabidopsis, this gene encodes a glutamine amidotransferase and cyclase which is involved in Arabidopsis histidine biosynthesis pathway. Histidine biosynthesis is crucial for microorganism survival, and was first characterized in *Eschericia coli* and *Salmonella typhimurium* (Winkler and

Ramos-Montañez 2009). Disruption of histidine biosynthesis in Arabidopsis leads to embryo lethality, aborted seeds, and inability to grow true leaves unless supplemented with histidine in growth media (Muralla et al. 2007). Muralla et al. (2007)'s systematic screen of histidine biosynthsis mutants in Arabidopsis only reported seed abortion and seedling phenotypes in heterozygous mutant lines, but no phenotypes for flowering time or floral development.

Although the histidine biosynthesis pathway has not been characterised as a regulator of flowering time, it is clear that it is crucial for early plant development. It is curious that it has almost zero expression in R-o-18, while it is expressed at around 10TMMC for the majority of Sarisha-14's shoot apex development. Furthermore, this particular gene has 8 recorded SNPs (Table 5.4), which suggests functional divergence of *AT-HF* orthologue between Sarisha-14 and R-o-18. This may be tied to earlier flowering time, although the molecular mechanism which drives this is yet to be understood.

Other possible B. rapa genes that repress flowering

There are several *B. rapa* genes that are not yet characterised in flowering time in either *B. rapa* or its Arabidopsis orthologues. These include *A01p012610.1_BraROA (AT4G20150)*, *A01p017080.1_BraROA (AT4G24740)* , *A03p021860.1_BraROA (AT2G38330)* and *A03p022300.1_BraROA (AT2G38800)*. These genes have distinct expression profiles between R-o-18 and Sarisha-14, and are all expressed lower in Sarisha-14 - which suggests that lower expression of these genes could contribute to earlier flowering time.

5.4.5 *Future lines of investigation*

In this chapter, we identified several candidate loci associated with earlier flowering time in Sarisha-14 and R-o-18. To establish a mechanistic link between sequence variation in the candidate loci with earlier flowering phenotype in *B. rapa*, we can take a multi-level approach combining genetic, proteomic and epigenetic analyses of these candidate loci.

One of the key candidate genes that stood out was *TFL1*, which is a well-known floral repressor in plants. The A03 copy of *TFL1* has divergent expression profiles between Sarisha-14 and R-o-18 in contrast with its other copies, where it is expressed lower in Sarisha-14. This suggests that SNPs in this candidate gene could contribute to reduced *TFL1* transcription, and possibly altered protein structure which prevents it from repressing flowering in Sarisha-14. More experimental work can be done to test this hypothesis, firstly by protein structure comparison, time series proteomics and genetics studies such as mutations in R-o-18's *TFL1* sequence to mimic Sarisha-14's sequence and observing its effect in R-o-18 flowering time. The binding efficiency of

TFL1 to its target genes would also shed more insight on its downstream regulatory effects in Sarisha-14 flowering time.

Several other potential promoters and repressors of flowering were also identified with divergent expression profiles between Sarisha-14 and R-o-18, however they don't have any recorded flowering phenotype or role in flowering time. A straightforward next step for validating their functions in flowering time is using reverse genetics to identify mutant alleles of the candidate genes in R-o-18's TILLING (Targeting Induced Local Lesions In Genomes) population, and comparing flowering time phenotypes in these lines against wild-type R-o-18 and Sarisha-14. Variation in flowering time in TILLING mutant lines would warrant further investigation in their regulatory role in flowering time, otherwise, it is possible that these candidate genes indirectly affect flowering time via other mechanisms, such as *CKL6* which could be accelerating Sarisha-14's cell growth rate, thus accelerating its progression to flowering.

In addition to protein-coding genes, sequence variation within miRNA loci could also contribute to the observed differences in flowering time. Further SNP analysis can be performed at the pri-miRNA loci of the candidate miRNAs from Chapter 3 such as Bra-miR397, Bra-miR398a and Bra-miR408 which are differentially expressed in Sarisha-14 and R-o-18, as well as the ageing pathway miRNA families Bra-miR156 and Bra-miR172. SNPs within the pri-miRNA loci may affect the secondary structure and folding of the precursor transcript, whereas SNPs upstream near the promoter region could influence the transcription rate of the precursor miRNA. Additionally, SNPs may be present within the miRNA complementary binding sites in target transcripts, which would impact binding efficiency and reduce miRNA-mediated repression of these genes. This may explain why some predicted target genes of certain miRNAs were not differentially expressed between the cultivars despite differential miRNA abundance. As pri-miRNA identification was not completed within the scope of this study, there is currently insufficient data to determine whether SNPs are present within pri-miRNA loci.

Sequence divergence in these genes could contribute to earlier expression, but the molecular mechanisms which control the timing of expression are still unknown. A possible approach is investigating the epigenetic profiles of these genes, such as profiling permissive/repressive histone modification marks using ChIP-seq at these candidate loci throughout the floral transition, and comparing it with their expression profiles. This could be paired with ATAC-seq to profile chromatin accessibility at these regions. For example, depletion of repressive H3K27me3 marks at *EMF1* in Sarisha-14 would allow for earlier upregulation than R-o-18. This approach can also identify epigenetic regulatory mechanisms for genes which are differentially expressed between the two cultivars.

By utilising the TILLING population to phenotype the effects of similar candidate loci mutations in R-o-18, performing transcriptome/proteome analysis on these lines to

observe the downstream impact of the gene mutation, and epigenetic analysis would shed more light on the regulatory mechanisms which allow Sarisha-14 to undergo the floral transition earlier than R-o-18.

5.4.6 *Promising breeding targets for agriculture*

Beyond mechanistic validation, these candidate loci may also have practical relevance within Brassica breeding programmes. As discussed in Chapter 1, Brassica crops contribute substantially to oilseed production, vegetable supply chains, and rotational cropping systems, particularly in regions where short growing seasons and climatic variability limit productivity. In this context, genetic variation that reduces, or potentially bypasses, vernalisation requirement could improve crop performance under increasingly warmer winters, where insufficient cold exposure may otherwise delay flowering. Likewise, shorter growth duration, if achieved without compromising oilseed yield, could facilitate the integration of Brassica crops into intercropping systems. Earlier flowering may also reduce dependence on vernalisation in certain environments, although this would need to be balanced against adequate vegetative growth to sustain yield.

Alleles associated with earlier flowering, such as the Sarisha-14 variant of *TFL1* or other loci identified through bulked segregant analysis, could therefore be incorporated into elite germplasm through marker-assisted selection. We could also perform targeted genome editing approaches such as CRISPR/Cas-mediated allelic modification to recreate favourable early-flowering alleles directly in high-performing commercial varieties, while retaining yield, oil quality, and disease resistance traits.

Taken together, the genetic variation identified in this study contributes not only to our fundamental understanding of flowering regulation in polyploid Brassica species, but also to the strategic improvement of short-duration, climate-resilient cultivars.

5.5 CONCLUSION

In this thesis chapter, we utilised bulk segregant analysis and time series transcriptomics to identify potential genomic loci which drive earlier flowering in Sarisha-14. We hypothesised that allelic differences in Sarisha-14 contributes to earlier flowering by differential expression of candidate regulators between the two cultivars, in either expression level or a temporal shift in expression.

We identified three key genomic regions that had strong allelic differences between R-o-18 and Sarisha-14, chrA01 and chrA03 which were associated with earlier flowering, and chrA02 which was associated with later flowering time. Within these regions, we found certain differentially expressed flowering time orthologues namely *TFL1* and

NF-YB1 and which play a key role in flowering time in Arabidopsis. We also identified potential novel *B. rapa* floral regulators from other pathways, such as *CKL6* which acts in anisotropic cell growth, or *AT-HF* which is only characterised in histidine biosynthesis, which are also differentially expressed in Sarisha-14. *EMF1, RD20* and *SUS2* are also flowering time orthologues which are expressed earlier in Sarisha-14, which could drive its earlier flowering time. Overall, we identified 7 known flowering time orthologues, and 51 potential genes with high impact SNPs in Sarisha-14.

SNPs within coding regions would have downstream impact particularly during protein translation, thus time series transcriptomics only captures one half of the story. How the SNPs change protein structure and the molecular regulation of flowering time would be an exciting path for future research. Furthermore, the candidate genes we've identified suggest that there could be other regulatory mechanisms at play at the epigenetic/chromatin level, which hasn't been studied yet in these cultivars.

Now that we've uncovered how Sarisha-14 potentially regulates its flowering time at the genomic level, we could also cross-check our findings with other related rapid-cycling cultivars, such as *B. oleracea* DH1012 or spring type cultivars such as *B. napus* Stellar, to see if these allelic differences are also conserved within rapid-cycling cultivars. Thus, these candidate genes could assist plant breeders to select for earlier flowering cultivars in Brassica species and potentially other crops.

6 DISCUSSION

6.1 CHAPTER SUMMARIES

Going back to the central question of this thesis, I set out to investigate the genetic basis of early flowering in the rapid-cycling *B. rapa* cultivar, Sarisha-14. Using a combination of miRNA timeseries expression profiling, miRNA target gene prediction and bulk segregant analysis, we have identified both conserved regulatory pathways and uncovered Sarisha-14 specific mechanisms that could underlie its earlier flowering time.

6.1.1 *Methodology for microRNA expression profiling*

The thesis begins with the generation of a miRNA expression timeseries dataset for Sarisha-14 and R-o-18 shoot apex and leaf tissue. I began with optimising the total RNA extraction protocol and small RNA sequencing workflow to improve miRNA detection, since miRNAs make up only a small percentage (< 4%) of total RNA extracted from *B. rapa* tissue, while ribosomal RNAs and other non-coding RNAs make up the majority of these libraries. In order to increase miRNA detection, we tested a new sequencing method with RealSeq Biosciences called Sequence Suppressor Probe (SSP) sequencing, which suppresses overrepresented sequences that should enrich miRNA for detection, however, we unexpectedly found that SSP-treated libraries had lower levels of Bra-miR156, which was not ideal for my investigation. After all of the troubleshooting and optimization, I successfully generated a small RNA expression time series dataset for Sarisha-14 and R-o-18 throughout development, which I developed a bioinformatics pipeline to identify canonical *B. rapa* miRNAs expressed in these datasets. Principle Component Analysis of the apex and leaf samples showed tissue specific clusters, and each sampled time point displayed aligned progression throughout development. Thus, this miRNA expression dataset lays the foundation for the analysis in the following chapters.

6.1.2 *Identification of candidate microRNAs involved in flowering*

Here, I explored the role of other miRNAs besides miR156 and miR172 in regulating flowering time. I tested two hypotheses, first, that some miRNAs will display conserved temporal expression dynamics between Sarisha-14 and R-o-18 similar to Bra-miR156 and Bra-miR172, and second, that another group of miRNAs will display cultivar-specific differences in expression pattern, therefore contributing to Sarisha-14's cultivar-specific earlier flowering. I utilised the miRNA timeseries and performed temporal pattern clustering (fuzzy c-means clustering) and curve registration to identify miRNAs with shared temporal dynamics between the two cultivars. Several known flowering associated miRNAs emerged, particularly Bra-miR159, Bra-miR319, Bra-miR394, Bra-miR157 and Bra-miR396 that had similar expression dynamics between Sarisha-14 and R-o-18. These miRNAs form regulatory modules that target flowering genes, growth regulators and stress response genes. On the other hand, Bra-miR398, Bra-miR169, Bra-miR399, Bra-miR397 and Bra-miR408 which mainly target stress response pathways displayed divergent cultivar-specific expression patterns, implying that they could contribute to Sarisha-14's earlier flowering. However, the expression of their target genes do not always reflect the cultivar-specific expression of their miRNAs, leaving a gap for further investigation of these Sarisha-14 specific miRNAs and their role in regulating flowering. This chapter identified several potential miRNAs which could be investigated for their role in Sarisha-14's earlier flowering phenotype.

6.1.3 *Investigating the ageing pathway in rapid-cycling B. rapa*

In chapter 4, I focus on two regulatory modules, miR156-SPL and miR172-AP2 which form the core regulatory network of the ageing pathway which regulates flowering time based on developmental age. I hypothesised that the expression dynamics of Bra-miR156 and Bra-miR172 determine the timing of the floral transition, and that it can be quantitatively measured using the ratio of their expression levels. Using the miRNA timeseries described in the previous chapter, and target gene prediction, I confirm that Bra-miR156-SPL and Bra-miR172-AP2 binding site sequences and expression dynamics are conserved - Bra-miR156 decline with age is reflected by *SPL15.A04* reciprocal upregulation, while Bra-miR172 increase coincides with *SMZ.A09* and *TOE2.A10* repression. The timing of these dynamics coincide with the timing of the floral transition in both cultivars, and I found that when the log10 ratio of miR156 to miR172 expression reached 1.1, the floral transition is initiated in Sarisha-14 and R-o-18 despite their 1 week difference in timing. These findings and the quantitative ratio of the two key miRNAs present an exciting opportunity for plant breeders to remodulate the timing of flowering in crops by manipulating the miR156/miR172 ratio.

6.1.4 *Functional annotation of candidate SNPs linked to earlier flowering*

In the final chapter, I hypothesised that Sarisha-14's earlier flowering phenotype is linked to allele specific differences in flowering associated genomic loci. To investigate this, I performed bulked segregant analysis of a segregating R-o-18 x Sarisha-14 F2 population. Three key genomic regions associated with early flowering were identified within chromosomes A01, A02 and A03. Early flowering alleles inherited from Sarisha-14 were found in chrA01 and chrA03, while in chrA02, alleles from R-o-18 contributed to early flowering. From these regions, I identified SNPs within *TFL1*, a known floral repressor, alongside other flowering gene orthologues of *SUS2, NF-YB1, PBS3, EMF1* and *RD20*. Other candidate genes were also identified, namely *CKL6*, a gene involved in casein biosynthesis, and *AT-HF*, which is characterised in histidine biosynthesis in Arabidopsis. These other genes are not well characterized in *B. rapa*, but their cultivar-specific expression patterns and genomic positions within high Δ SNP-index regions warrants further investigation. Visualisation of these genes' expression dynamics imply that these SNPs could be associated with regulatory divergence in Sarisha-14, but since these SNPs are found within the coding regions of these genes, its highly likely that the SNPs' impacts occur downstream at the protein expression level. This chapter therefore highlights allelic differences in candidate genes that could contribute to earlier flowering, which could explored further to understand how they impact protein-level expression.

6.2 OUTLOOKS AND LIMITATIONS

The findings from this study has provided new insights into the regulation of flowering in rapid-cycling *Brassica rapa*, leading to exciting avenues for future Brassica and flowering time research. Although this thesis has certain limitations, they also present new opportunities for a deeper understanding of the regulatory mechanisms of flowering time.

The *B. rapa* small RNA sequencing timeseries dataset generated in this thesis offers significant research opportunities beyond the scope of this study. While our main focus was on miRNAs, the small RNA libraries also contain abundant non-coding RNAs, which could be explored for their regulatory roles in shoot apex development. Moreover, I found that 10–25% of R-o-18 aligned small RNA reads remain unidentified after filtering other RNA types. These likely include additional novel miRNAs or other regulatory small RNAs. Following that was the high abundance of 24 nt small RNAs, suggesting a strong role for siRNA-mediated regulation in floral development. Ongoing work aims to characterise siRNAs derived from *B. rapa* long non-coding RNAs (lncRNAs) described in Mehraj et al. (2021), and take advantage of 24 nt phasiRNA annotation from the srnaAnno database, which may reveal further regulatory complexity in floral development.

A key limitation in this study is the lack of direct temporal alignment between our small RNA dataset and the RNA-seq timeseries. Ideally, both would be generated from the same biological samples to enable integrated analyses of miRNA–mRNA relationships. Our current design avoids redundancy but reduces resolution in temporal dynamics, particularly since we have limited vegetative stage timepoints in Sarisha-14 gene expression data from Calderwood, Hepworth, et al. (2021). Another limitation arises in leaf sampling strategy. Most vegetative phase change studies measure miR156 expression by sampling the youngest expanding leaf and referencing leaf number (G. Wu and Poethig 2006; J.-W. Wang 2014; Zhao, Doody, and Poethig 2023), whereas I analysed the first true leaf at each timepoint, which ages during development. This choice was partly pragmatic, since rapid-cycling *B. rapa* and *B. oleracea* lack obvious juvenile-to-adult morphological markers such as leaf shape or trichome density, making conventional staging difficult. Nevertheless, this divergence from common practice complicates direct comparison with other ageing studies.

Our data support the hypothesis that changes in miRNA expression underlie age-related regulation of flowering, consistent with preliminary pri-miRNA evidence from Calderwood, Hepworth, et al. (2021). This raises the intriguing possibility that pri-miRNA levels may serve as proxies for mature miRNA abundance. However, this might be on a case-by-case basis, for example, pri-miR172 processing to mature miR172 has been shown to be temperature-dependent, which implies that it may not always be an accurate proxy for mature miRNA abundance (Lauressergues et al. 2015). To test this, improved annotation of pri-miRNA loci in *B. rapa* is essential, and can utilise the existing RNA-seq dataset from Calderwood, Hepworth, et al. (2021) for pri-miRNA expression studies. More broadly, it would be interesting to test our ratio hypothesis across related crop types, such as cabbage vs. broccoli types, which have different flowering requirements (delayed flowering in cabbage, halted floral bud development in broccoli), as well as between vernalisation-dependent and rapid-cycling cultivars. A key question is whether the observed ratio of miR156/miR172 expression (the "1.1 ratio") is unique to *B. rapa* cultivars or represents a generalisable regulatory feature across species. Quantitative reports on their expression profiles also allows computational modelling of this hypothesis, similar to how Calderwood, Lloyd, et al. (2021) modelled *FLC* paralogues in different *B. napus* crop types.

A recurring limitation in this work is the absence of experimental validation of miRNA–target interactions. While target gene expression divergence consistent with miRNA expression divergence suggests regulatory interactions, these remain correlative. Experimental pproaches such as degradome sequencing or RACE (RNA ligase-mediated rapid amplification of cDNA ends) will be essential to confirm target cleavage for specific targets of miR156 and miR172, and should be prioritised in future work.

Exploratory analysis also highlighted additional candidate miRNAs that may contribute to flowering regulation. These could act in parallel with canonical ageing pathways, integrating developmental signals with stress or nutrient responses. However, the same limitation applies: without experimental validation of miRNA-target cleavage, either via degradome sequencing or RACE, these regulatory modules remain putative. Future studies could integrate miRNA–target predictions with gene regulatory network inference, linking miRNA–TF interactions with TF–target modules. This systems-level approach would provide a more holistic view of cultivar-specific regulatory networks controlling flowering.

The BSA approach successfully identified candidate flowering-time genes, which could be potential targets for crop breeding programmes. This demonstrates the strength of BSA in narrowing down complex traits to specific genomic loci. However, more could be done with this analysis, particularly identifying SNP variants associated miRNA regulation. Although SNPs within regulatory regions of miRNA-encoding loci could link Sarisha-14's earlier flowering with miRNA divergent expression, the lack of R-o-18 *B. rapa* pri-miRNA annotation restricts our ability to identify such variants. Calderwood, Hepworth, et al. (2021)'s pri-miRNA annotation was based on a previous *B. rapa* genome annotation (*B. rapa* Chiifu v3 reference genome, L. Zhang et al. (2018)), but analysis is still ongoing to identify miRNA encoding loci using the currently published R-o-18 reference genome, with prediction tools such as miRdeep-P2 and RNAfold.

Similarly, I did not find high-impact SNPs within miRNA target genes, such as *SPL*s or *AP2*-like genes which are target genes from the ageing pathway. This is likely due to methodological constraints: SnpEff is not designed to predict SNP variants in regulatory or miRNA binding-sites as "high impact" SNPs. Thus, identifying SNPs within miRNA target-binding sites would provide further insights on miRNA regulation at the genomic level. Following this train of analysis, integrating protein structure prediction and proteomics would identify how the high impact SNPs alter protein conformation or abundance, providing stronger evidence of allelic differences contributing to differential regulation of flowering time.

6.3 CONCLUDING REMARKS

In this thesis, I explored how flowering time is regulated in rapid-cycling *Brassica rapa*, particularly how one cultivar (Sarisha-14) is able to flower more rapidly than the other (R-o-18). By integrating new small RNA sequencing and genomic variant datasets with published time-series expression data, I uncovered both biological insights and generated resources that is full of potential for aiding flowering research and crop breeding.

This study identified strong candidate flowering time regulators like the floral repressor *TFL1*, supported a ratio-based hypothesis of miR156 and miR172–mediated flowering regulation, and revealed conserved ageing pathways undergoing lineage-specific divergence. These findings highlight how Brassicas are able to finetune their flowering time based on endogenous cues like plant age, and environmental signals like nutrient deficiency or stress response.

Beyond their biological implications, these findings also have potential practical applications in crop breeding. We could alter flowering time by altering the expression of these miRNA regulators, potentially leading to reduced vernalization requirements, reducing the juvenile period in perennial crops, or increased climate resilience against a warming climate. Although more questions remain about how genetic, environmental, and developmental signals integrate to control flowering, this thesis provides both a shortlist of candidate flowering regulators, and a unique hypothesis which can guide future research steps.

A

APPENDIX

A.1 APPENDIX CHAPTER 2

Supplementary Tables for Chapter 2

Table A.1: Nanodrop measurements of DH1012 RNA extraction trial.

Sample ID	Replicate	Extraction kit	Grinding method	RNA conc. (ng/uL)	A260/ A280	A260/ A230	Total RNA (ng)
dz1a	1A	Direct-zol	Steel ball	683.6	2.01	2.32	10254
dz2a	2A	Direct-zol	Steel ball	972.2	2.06	2.35	14583
dz3a	3A	Direct-zol	Steel ball	762.9	2.05	2.34	11443.5
dz7a	7A	Direct-zol	Pestle and drill	363.5	2.04	2.3	5452.5
dz8a	8A	Direct-zol	Pestle and drill	1077.1	2.05	2.36	16156.5
dz9a	9A	Direct-zol	Pestle and drill	730.8	2.03	2.32	10962
en4a	4A (s)	E.Z.N.A	Steel ball	114.1	1.93	1.57	3423
	4A (L)	E.Z.N.A	Steel ball	122.3	2.08	2.3	8561
en5a	5A (s)	E.Z.N.A	Steel ball	84.6	1.95	1.66	2538
	5A (L)	E.Z.N.A	Steel ball	80.6	2.06	2.25	5642
en6a	6A (s)	E.Z.N.A	Steel ball	95.5	1.92	1.58	2865
	6A (L)	E.Z.N.A	Steel ball	163.4	2.08	1.39	11438
en10a	10A (s)	E.Z.N.A	Pestle and drill	89.1	1.91	1.67	2673
	10A (L)	E.Z.N.A	Pestle and drill	73.4	2.03	2.36	5138
en11a	11A (s)	E.Z.N.A	Pestle and drill	97.1	1.81	1.07	2913
	11A (L)	E.Z.N.A	Pestle and drill	154.3	2.05	2.09	10801
en12a	12A (s)	E.Z.N.A	Pestle and drill	67.8	1.87	1.31	2034
	12A (L)	E.Z.N.A	Pestle and drill	155.2	2.08	2.34	10864

Probe name	Sequence
1	ACTGAGATTCAGCCCTTTGTCGCTAAGATTCGA
2	AAAAGCGTGGAGGTTCGAGTCCTCTTCAAGGCACCA
3	AGAGCATCTCGGTTCGAGTCCGAGTAGCGGCACCA
4	GAGGCATCCTAAC\AGACCGGTAGACTTGAAC
5	AGGTCCTGCGGAAAAATAGCTCGACGCCAGGATG
6	TAAAACGACTCTCGGCAACGGATATCT
7	ACCTCCCGGGAAGTCCTCGTGTTGCACCCCT
8	CTGGCCGAGGGCACGTCTGCCTGGGTGTCACA

Table A.2: SSP probe sequences used for trial SSP suppression of overrepresented sequences.

Table A.3: List of *B. rapa* miRNAs used for miRNA identification.

miRNA	Sequence	Length (nt)
Bra-MIR156a	TGACAGAAGAGAGTGAGCAC	20
Bra-MIR156d	TTGACAGAAGAAAGAGAGCAC	21
Bra-MIR156h	GACAGAAGAGAGTGAGCACA	20
Bra-MIR156i	TTGACAGAAGAGAGCGAGCAC	21
Bra-MIR156n	TGACAGAAGAGAGAGAGCAC	20
Bra-MIR157a	TTGACAGAAGATAGAGAGCAC	21
Bra-MIR158a	TTTCCAAATGTAGACAAAGCA	21
Bra-MIR158b	TCCCAAATGTAGACAAAGCA	20
Bra-MIR159a	TTTGGATTGAAGGGAGCTCTA	21
Bra-MIR159c	TTTGGATTGAAGGGAGCTCC	20
Bra-MIR160a	TGCCTGGCTCCCTGTATGCCA	21
Bra-MIR160g	TGCCTGGCTCCCTGTATACCA	21
Bra-MIR162a	TCGATAAACCTCTGCATCCAG	21
Bra-MIR162b	TCGATAAACCTGTGCATCCAG	21
Bra-MIR164a	TGGAGAAGCAGGGCACGTGCA	21
Bra-MIR164b	TGGAGAAGCAGGGCACGTGCG	21
Bra-MIR165a	TCGGACCAGGCTTCATCCCCC	21
Bra-MIR166a	TCGGACCAGGCTTCATTCCCC	21
Bra-MIR167a	TGAAGCTGCCAGCATGATCTA	21
Bra-MIR167c	TGAAGCTGCCAGCATGATCT	20
Bra-MIR168a	TCGCTTGGTGCAGGTCGGGAAC	22
Bra-MIR168b	TCGCTTGGTGCAGGTCGGGAC	21
Bra-MIR169a	CAGCCAAGGATGACTTGCCGG	21
Bra-MIR169d	TGAGCCAAGGATGACTTGCCT	21
Bra-MIR169e	TAGCCAAGGATGACTTGCCTG	21
Bra-MIR169g	TAGCCAAGGATGACTTGCCT	20
Bra-MIR169h	CAGCCAAGGATGACTTGCCG	20
Bra-MIR169i	GTAGCCAAGGATGACTTGCCT	21
Bra-MIR169j	GCAGCCAAGGATGACTTGCCG	21
Bra-MIR169l	CAGCCAAGGATGACTTGCCGA	21
Bra-MIR169q	TGAGCCAAAGATGACTTGCCG	21
Bra-MIR169s	TGAGCCAAGGATGACTTGCCG	21
Bra-MIR170	TATTGGCCTGGTTCACTCAGA	21
Bra-MIR171a	TTGAGCCGTGCCAATATCACG	21
Bra-MIR171e	TGATTGAGCCGCGCCAATATC	21
Bra-MIR171g	TTGAGCCGTGCCAATATCTCT	21

miRNA	Sequence	Length (nt)
Bra-MIR172a	AGAATCTTGATGATGCTGCAT	21
Bra-MIR172c	AGAATCTTGATGATGCTGCAG	21
Bra-MIR172d	GGAATCTTGATGATGCTGCAT	21
Bra-MIR319a	TTGGACTGAAGGGAGCTCCCT	21
Bra-MIR319c	TTGGACTGAAGGGAACTCCCT	21
Bra-MIR319e	TTGGACTGAAGGGAGCTCCTT	21
Bra-MIR319f	TTGGACTGAAGGGAGCTCCTTC	22
Bra-MIR390a	AAGCTCAGGAGGGATAGCGCC	21
Bra-MIR391a	TTCGCAGGAGAGATAGCGCCA	21
Bra-MIR393a	TCCAAAGGGATCGCATTGATCC	22
Bra-MIR393d	TCCAAAGGGATCGCATTGATC	21
Bra-MIR394a	TTGGCATTCTGTCCACCTCC	20
Bra-MIR395a	CTGAAGTGTTTGGGGGAACTC	21
Bra-MIR395g	CTGAAGTGTTTGGGGGGACTC	21
Bra-MIR396a	TTCCACAGCTTTCTTGAACTT	21
Bra-MIR396b	TTCCACAGCTTTCTTGAACTG	21
Bra-MIR397	TCATTGAGTGCAGCGTTGATGT	22
Bra-MIR398a	TGTGTTCTCAGGTCACCCCTG	21
Bra-MIR398c	TGTGTTCTCAGGTCACCCCTT	21
Bra-MIR399a	TGCCAAAGGAGAGTTGCCCTG	21
Bra-MIR399e	TGCCAAAGGAGATTTGCCCGG	21
Bra-MIR399h	TGCCAAAGGAGAGTTGCCCT	20
Bra-MIR400	TATGAGAGTATTATAAGTCAC	21
Bra-MIR403	TTAGATTCACGCACAAACTCG	21
Bra-MIR408	TGCACTGCCTCTTCCCTGGCT	21
Bra-MIR827	TTAGATGACCATCAACAAATA	21
Bra-MIR845	CGGCTCTGATACCAATTGATG	21
Bra-MIR858a	TTCGTTGTCTGTTCGACCTTG	21
Bra-MIR858b	TTTCGTTGTCTGTTCGACCTT	21
Bra-MIR860	TCAATACATTGGACTACATAT	21
Bra-MIR1140	ACAGCCTAAACCAATCGGAGC	21
Bra-MIR1511	AACCTGGCTCTGATACCATGAAGT	24
Bra-MIR2111a	TAATCTGCATCCTGAGGTTTA	21
Bra-MIR2111c	ATCCTCGGGATGCGGATTACC	21
Bra-MIR2111d	ATCCTCGGGACACAGATTACC	21
Bra-MIR2111f	GACCTCAGGATGCGGATTACC	21
Bra-MIR5654a	ATAAATCCCAAGCATCATCCA	21
Bra-MIR5711	TGTTTTGTGGGTTTCTACCGA	21

miRNA	Sequence	Length (nt)
Bra-MIR5712a	AATATTAATATAATTGGTGAG	21
Bra-MIR5712b	AATATTAATATAGTTGGTGAG	21
Bra-MIR5714	TTTCTTGAGGTTGTAGAGTCT	21
Bra-MIR5715	ACGTGATAAGCCTCTGAAGAA	21
Bra-MIR5716	TTGGATAATTGAAGATATAAA	21
Bra-MIR5717	GTTTGGATTGTTTGCCTTGGC	21
Bra-MIR5718	TCAGAACCAAACACAGAACAAG	22
Bra-MIR5722	TGAAATAGAGTCATGTGGAACG	22
Bra-MIR5723	AATGTGCTGCAATATCTCTGC	21
Bra-MIR5724	AACCGCCGGTTTGATAATAGC	21
Bra-MIR5726	TTATTCAAGCAACCTTTGCCT	21
Bra-MIR6029	TGGGGTTGTGATTTCAGGCTT	21
Bra-MIR6030	TCCACCCATACCATACAGACCC	22
Bra-MIR6032	TCTGCTGGTCGTTCCATGTTA	21
Bra-MIR6034	TCTGATGTATATAGCTTTGGG	21
Bra-MIR9552a	CTATCGGTCTACTCGGTCAGC	21
Bra-MIR9552b	CGGTGATGAAGAAGCTAGCGC	21
Bra-MIR9554	TTATATCCAAGTATCATTCCT	21
Bra-MIR9555a	TTCCCGTAAAGCTTAGAACCC	21
Bra-MIR9555b	TAATTGCGGGGTTCTAAGCT	20
Bra-MIR9556	TCTACTTTCACCAATTGGCCT	21
Bra-MIR9557	TTTGCGTTTCAACTCTGTCCT	21
Bra-MIR9558	AGAGATGTCTGGCTTGCAACA	21
Bra-MIR9560a	ACAGGTGGTGGAACAAATATGAGT	24
Bra-MIR9563	AGCGGAATATAAGAACCCGTCTCT	24
Bra-MIRN252	CTGGATTTTTGACAATGGCAAA	22
Bra-MIRN271	TTGGGAAGATTATGAGAGCATG	22
Bra-MIRN316	TGAATGTCTTTCTCTTCATC	20
Bra-MIRN317	TTTTTCCTCGGAATTTCCTCG	21
Bra-MIRN318	AGAGAGATATAGTTAGAGGAC	21
Bra-MIRN319	TTCCGCGATGACTTTCGGCCT	21
Bra-MIRN320	TCGGAATATACCGACGAACT	20
Bra-MIRN321	ATAGAGGAAAAAATAGAGTTC	21
Bra-MIRN322	CATCGAGAAATATTTTTGTGAC	22
Bra-MIRN323	TTGAAACTTTGATCTAGATC	20
Bra-MIRN324a	AGATTTGTAGTTTACATGGCCT	22
Bra-MIRN325	AACTGTAGTTGTTCCTGGCAC	21
Bra-MIRN326	TATAGGATTCCACAAAAATTG	21

miRNA	Sequence	Length (nt)
Bra-MIRN327	GTAAGTCTTCCGGTAGACAAC	21
Bra-MIRN328	TTACTTATTGTCGGTTGACTTT	22
Bra-MIRN329	TTGTTCTTCAATATGTACTTG	21
Bra-MIRN330	TGACTCACTTGAATAAAAGCCT	22
Bra-MIRN331	TGAACGGACATGGGAGGCCCA	21
Bra-MIRN332	TATTGCGTGCTGTTTCATGGCT	22
Bra-MIRN333a	TATTTGCGGGCTAAGGACTTT	21
Bra-MIRN334	TACAAATCCTCCAACTTTTCTT	22
Bra-MIRN335	CTGAAGCTAACTATGTGAAGT	21
Bra-MIRN336	CGACGGCTAGGGTTTTCGGCCT	22
Bra-MIRN337	ACCAAATGATGGTCATTATTTA	22
Bra-MIRN338	TAACAACTCTCCACAAGACAA	21
Bra-MIRN339	CCAGTAGTTCGTTAGGTTTGAC	22
Bra-MIRN340	AATCGGACCAATAAAAAACTC	21
Bra-MIRN341	ATTCCGGACTTTGATGATTTC	21
Bra-MIRN342	GGCAATGTTGGACTTGGACC	20
Bra-MIRN343	TTTCCGATTGTATGAACGAAAT	22
Bra-MIRN344	TCAACCCCCTAGGATGAACCTC	22
Bra-MIRN345	TTGAACCCAAGATCTCACCCT	21
Bra-MIRN346	TGACTAACTCTTGTTACGATC	21
Bra-MIRN347	TTGTCGGAGTTTATGATCTC	20
Bra-MIRN348a	ATTGATGATGATCGTTGCTAA	21
Bra-MIRN348b	ATTGATGATGATTATTGCTAA	21
Bra-MIRN349	TTATCCAGGATTTCTTAGGGC	21
Bra-MIRN350	TTGGTAGTAATGCGTGTTCCCG	22
Bra-MIRN351	GCTTGTGGTTCGGACACTCT	20
Bra-MIRN352	TGTGTGGGAAAAACAGAAGAG	21
Bra-MIRN353	AGCGAGGCGAATCTGTCGGTC	21
Bra-MIRN354	AATGGAGTGGGAAATGGAGTA	21
Bra-MIRN355	AGGAAGACAAGATCATGCTGG	21
Bra-MIRN356	TTCAGTCGGAAGTCTTTGGAGG	22
Bra-MIRN357a	TGCGGAAGGATCATTGTCGTA	21
Bra-MIRN358	TGGAGAGTAGATCTATGGCTG	21
Bra-MIRN359	TTGAAAGTGAAATGAAAGATA	21
Bra-MIRN360	TTGTGTTGTGATGATAATCCG	21
Bra-MIRN361	CGGATATCTTAGGATGAGGTT	21
Bra-MIRN362a	CGTGGACGACTGAATTATAAG	21
Bra-MIRN363	CGAAATCATCGTCCGCGTGTTG	22

miRNA	Sequence	Length (nt)
Bra-MIRN364	GTAGACGGTCTAGGTGTGGAAC	22
Bra-MIRN365	TGGTTGCTGAGAATAGGAGAGT	22
Bra-MIRN366	TTCGTTCTCTATTTCTCCATG	21
Bra-MIRN367	ATCTTGTCGGAGTTTATGATC	21
Bra-MIRN368a	TTGAGGGTTTATTTCACTGGAT	22
Bra-MIRN369	TCATTAGGTTTGGATTTTGGT	21
Bra-MIRN370	ATATGACATGCGTTTTTCGGA	21
Bra-MIRN371	TCGTACTTTGAGAACCATTCGC	22
Bra-MIRN372	CAGCGTCGGGACTCTGATCATC	22
Bra-MIRN373	TCTCGTTCGTCTTCTACTATC	21
Bra-MIRN374a	TGAACCCTTGGATAAATCACA	21
Bra-MIRN375	GTATCCATCTCGTTCGTCTTC	21
Bra-MIRN376	TCCCGGTTTGTGGTATATGTT	21
Bra-MIRN377a	TGGATATGATGAAATGGCATA	21
Bra-MIRN378	TGGCACAATCTGATCTGCATC	21
Bra-MIRN379	ATCTTCTGACTCAACTGCTTG	21
Bra-MIRN380	CGTCCTTAGCTCGGATTTTGGC	22
Bra-MIRN381	AATCGGTACGAATCAGGTAGA	21

Table A.4: Read alignment statistics of full miRNA timeseries.

sample	Trimmed reads	Genome aligned	ncRNA	Repeats	miRNA	coding	unannotated	not ro18
RO_d19_3L	28215567 (100.0%)	22775121 (80.7%)	15250860 (54.1%)	87772 (0.3%)	563166 (2.0%)	1225234 (4.3%)	5648089 (20.0%)	5440446 (19.3%)
RO_d21_3A	29549738 (100.0%)	20180858 (68.3%)	11610463 (39.3%)	133892 (0.5%)	309222 (1.0%)	1524956 (5.2%)	6602325 (22.3%)	9368880 (31.7%)
RO_d11_2A	44587549 (100.0%)	37858575 (84.9%)	28541042 (64.0%)	130571 (0.3%)	279145 (0.6%)	2138753 (4.8%)	6769064 (15.2%)	6728974 (15.1%)
S14_d7_3A	33943042 (100.0%)	29499048 (86.9%)	24234008 (71.4%)	83286 (0.2%)	264607 (0.8%)	1392911 (4.1%)	3524236 (10.4%)	4443994 (13.1%)
S14_d9_3L	33809742 (100.0%)	28539501 (84.4%)	19165899 (56.7%)	148369 (0.4%)	613821 (1.8%)	2257506 (6.7%)	6353906 (18.8%)	5270241 (15.6%)
RO_d15_1L	34369480 (100.0%)	29134548 (84.8%)	21572161 (62.8%)	74937 (0.2%)	464076 (1.4%)	1525866 (4.4%)	5497508 (16.0%)	5234932 (15.2%)
S14_d10_1L	40327746 (100.0%)	33953170 (84.2%)	23119004 (57.3%)	128811 (0.3%)	619471 (1.5%)	2523920 (6.3%)	7561964 (18.8%)	6374576 (15.8%)
S14_d10_1A	37262241 (100.0%)	32156377 (86.3%)	25204735 (67.6%)	147746 (0.4%)	423460 (1.1%)	1813375 (4.9%)	4567061 (12.3%)	5105864 (13.7%)
S14_d13_1L	37756563 (100.0%)	31554053 (83.6%)	23450358 (62.1%)	125650 (0.3%)	613428 (1.6%)	1895349 (5.0%)	5469268 (14.5%)	6202510 (16.4%)
S14_d13_3L	32761267 (100.0%)	27435588 (83.7%)	19299388 (58.9%)	158927 (0.5%)	666046 (2.0%)	1872661 (5.7%)	5438566 (16.6%)	5325679 (16.3%)
S14_d14_2A	39906335 (100.0%)	34623893 (86.8%)	26384449 (66.1%)	117794 (0.3%)	530036 (1.3%)	2042269 (5.1%)	5549345 (13.9%)	5282442 (13.2%)
S14_d24_1L	31370720 (100.0%)	25489725 (81.3%)	17366557 (55.4%)	70392 (0.2%)	823715 (2.6%)	1334799 (4.3%)	5894262 (18.8%)	5880995 (18.7%)
S14_d12_1L	54072055 (100.0%)	44754671 (82.8%)	28637181 (53.0%)	142696 (0.3%)	722615 (1.3%)	3222473 (6.0%)	12029706 (22.2%)	9317384 (17.2%)
S14_d12_3L	42433174 (100.0%)	34982882 (82.4%)	19289077 (45.5%)	195981 (0.5%)	1121200 (2.6%)	3068474 (7.2%)	11308150 (26.6%)	7450292 (17.6%)
S14_d14_2L	27201897 (100.0%)	22636432 (83.2%)	15869763 (58.3%)	120425 (0.4%)	554449 (2.0%)	1353423 (5.0%)	4738372 (17.4%)	4565465 (16.8%)
RO_d21_4L	35583936 (100.0%)	26726569 (75.1%)	17275943 (48.5%)	106034 (0.3%)	652648 (1.8%)	1347190 (3.8%)	7344754 (20.6%)	8857367 (24.9%)
S14_d24_3L	27784798 (100.0%)	22184215 (79.8%)	13819173 (49.7%)	89180 (0.3%)	759895 (2.7%)	1387621 (5.0%)	6128346 (22.1%)	5600583 (20.2%)
S14_d24_2A	20576124 (100.0%)	17485587 (85.0%)	8968453 (43.6%)	91483 (0.4%)	425117 (2.1%)	1662128 (8.1%)	6338406 (30.8%)	3090537 (15.0%)
S14_d12_3A	27590892 (100.0%)	23995294 (87.0%)	18331010 (66.4%)	69048 (0.3%)	306153 (1.1%)	1529322 (5.5%)	3759761 (13.6%)	3595598 (13.0%)
S14_d14_1L	41462780 (100.0%)	35082863 (84.6%)	26100103 (62.9%)	148126 (0.4%)	805901 (1.9%)	1924540 (4.6%)	6104193 (14.7%)	6379917 (15.4%)
S14_d13_2A	27439269 (100.0%)	23357789 (85.1%)	15537007 (56.6%)	110591 (0.4%)	490160 (1.8%)	1921805 (7.0%)	5298226 (19.3%)	4081480 (14.9%)

sample	Trimmed reads	Genome aligned	ncRNA	Repeats	miRNA	coding	unannotated	not ro18
S14_d24_3A	35731029 (100.0%)	28872876 (80.8%)	14512141 (40.6%)	116530 (0.3%)	403001 (1.1%)	2390344 (6.7%)	11450860 (32.0%)	6858153 (19.2%)
S14_d13_3A	35176724 (100.0%)	29964389 (85.2%)	21138487 (60.1%)	82589 (0.2%)	347695 (1.0%)	2057520 (5.8%)	6338098 (18.0%)	5212335 (14.8%)
S14_d12_2A	29602433 (100.0%)	24936587 (84.2%)	15831734 (53.5%)	108044 (0.4%)	598122 (2.0%)	2339412 (7.9%)	6059275 (20.5%)	4665846 (15.8%)
RO_d37_3A	25664474 (100.0%)	17522896 (68.3%)	11455482 (44.6%)	165157 (0.6%)	468886 (1.8%)	1309487 (5.1%)	4123884 (16.1%)	8141578 (31.7%)
RO_d15_1A	32325585 (100.0%)	22876235 (70.8%)	16816207 (52.0%)	153675 (0.5%)	433567 (1.3%)	1293502 (4.0%)	4179284 (12.9%)	9449350 (29.2%)
S14_d7_3L	52899596 (100.0%)	46473948 (87.9%)	41260428 (78.0%)	123486 (0.2%)	247980 (0.5%)	1864184 (3.5%)	2977870 (5.6%)	6425648 (12.1%)
S14_d9_3A	29794824 (100.0%)	25160505 (84.4%)	16718044 (56.1%)	87163 (0.3%)	394735 (1.3%)	1894809 (6.4%)	6065754 (20.4%)	4634319 (15.6%)
RO_d19_3A	31353182 (100.0%)	22376020 (71.4%)	14720874 (47.0%)	172085 (0.5%)	459814 (1.5%)	1589882 (5.1%)	5433365 (17.3%)	8977162 (28.6%)
S14_d10_2A	31412200 (100.0%)	26697250 (85.0%)	20537067 (65.4%)	78292 (0.2%)	331739 (1.1%)	1595633 (5.1%)	4154519 (13.2%)	4714950 (15.0%)
RO_d11_2L	56429438 (100.0%)	47246782 (83.7%)	29659349 (52.6%)	229210 (0.4%)	871116 (1.5%)	3738281 (6.6%)	12748826 (22.6%)	9182656 (16.3%)
RO_d21_3L	33959349 (100.0%)	25381691 (74.7%)	16801682 (49.5%)	81711 (0.2%)	539166 (1.6%)	1226445 (3.6%)	6732687 (19.8%)	8577658 (25.3%)
RO_d19_1A	35369883 (100.0%)	22715550 (64.2%)	15529725 (43.9%)	179083 (0.5%)	437507 (1.2%)	1409572 (4.0%)	5159663 (14.6%)	12654333 (35.8%)
RO_d21_1L	21359689 (100.0%)	14049436 (65.8%)	9824614 (46.0%)	42334 (0.2%)	192867 (0.9%)	540359 (2.5%)	3449262 (16.1%)	7310253 (34.2%)
S14_d7_1L	51567436 (100.0%)	43959608 (85.2%)	38064119 (73.8%)	90447 (0.2%)	177282 (0.3%)	2200407 (4.3%)	3427353 (6.6%)	7607828 (14.8%)
S14_d9_1A	40176461 (100.0%)	33269827 (82.8%)	26904560 (67.0%)	109862 (0.3%)	393412 (1.0%)	1691682 (4.2%)	4170311 (10.4%)	6906634 (17.2%)
S14_d10_2L	39302061 (100.0%)	33342847 (84.8%)	26098030 (66.4%)	146589 (0.4%)	507272 (1.3%)	1889599 (4.8%)	4701357 (12.0%)	5959214 (15.2%)
RO_d15_3A	35119676 (100.0%)	25262663 (71.9%)	16962015 (48.3%)	151961 (0.4%)	478926 (1.4%)	1672635 (4.8%)	5997126 (17.1%)	9857013 (28.1%)
RO_d13_2A	33771886 (100.0%)	28587083 (84.6%)	21954500 (65.0%)	88620 (0.3%)	241648 (0.7%)	1537082 (4.6%)	4765233 (14.1%)	5184803 (15.4%)
S14_d14_1A	40712457 (100.0%)	34529804 (84.8%)	22445099 (55.1%)	149615 (0.4%)	1031354 (2.5%)	2787731 (6.8%)	8116005 (19.9%)	6182653 (15.2%)
S14_d10_3L	31830233 (100.0%)	26519201 (83.3%)	15889290 (49.9%)	187883 (0.6%)	923518 (2.9%)	2407321 (7.6%)	7111189 (22.3%)	5311032 (16.7%)
S14_d24_2L	29116228 (100.0%)	23645745 (81.2%)	15686807 (53.9%)	124790 (0.4%)	1044714 (3.6%)	1331865 (4.6%)	5457569 (18.7%)	5470483 (18.8%)
S14_d13_1A	43785556 (100.0%)	37701374 (86.1%)	27508737 (62.8%)	198719 (0.5%)	726259 (1.7%)	2554943 (5.8%)	6712716 (15.3%)	6084182 (13.9%)
RO_d37_2A	32610667 (100.0%)	22426303 (68.8%)	11997999 (36.8%)	193919 (0.6%)	814229 (2.5%)	2014294 (6.2%)	7405862 (22.7%)	10184364 (31.2%)

sample	Trimmed reads	Genome aligned	ncRNA	Repeats	miRNA	coding	unannotated	not ro18
S14_d13_2L	27656072 (100.0%)	22952070 (83.0%)	17132079 (61.9%)	204087 (0.7%)	459243 (1.7%)	1349524 (4.9%)	3807137 (13.8%)	4704002 (17.0%)
S14_d12_1A	49235343 (100.0%)	42082314 (85.5%)	32925004 (66.9%)	209746 (0.4%)	635072 (1.3%)	2525579 (5.1%)	5786913 (11.8%)	7153029 (14.5%)
RO_d37_3L	46719835 (100.0%)	30067558 (64.4%)	19446876 (41.6%)	90657 (0.2%)	858529 (1.8%)	1548671 (3.3%)	8122825 (17.4%)	16652277 (35.6%)
S14_d12_2L	35374340 (100.0%)	28650755 (81.0%)	17400420 (49.2%)	144481 (0.4%)	703126 (2.0%)	2493317 (7.0%)	7909411 (22.4%)	6723585 (19.0%)
S14_d24_1A	33108399 (100.0%)	27686507 (83.6%)	17112810 (51.7%)	119454 (0.4%)	528891 (1.6%)	2160780 (6.5%)	7764572 (23.5%)	5421892 (16.4%)
S14_d10_3A	35662593 (100.0%)	30697248 (86.1%)	24730594 (69.3%)	72123 (0.2%)	257917 (0.7%)	1516409 (4.3%)	4120205 (11.6%)	4965345 (13.9%)
RO_d21_4A	31262821 (100.0%)	21720640 (69.5%)	13502141 (43.2%)	152968 (0.5%)	387052 (1.2%)	1588279 (5.1%)	6090200 (19.5%)	9542181 (30.5%)
RO_d37_2L	36180363 (100.0%)	25097273 (69.4%)	16821702 (46.5%)	73586 (0.2%)	720482 (2.0%)	1216954 (3.4%)	6264549 (17.3%)	11083090 (30.6%)
RO_d15_3L	40519252 (100.0%)	34336514 (84.7%)	20881486 (51.5%)	144632 (0.4%)	728535 (1.8%)	2079788 (5.1%)	10502073 (25.9%)	6182738 (15.3%)
S14_d7_1A	26778736 (100.0%)	22866815 (85.4%)	18531413 (69.2%)	62786 (0.2%)	246730 (0.9%)	1284278 (4.8%)	2741608 (10.2%)	3911921 (14.6%)
S14_d9_1L	44915246 (100.0%)	37208216 (82.8%)	26688632 (59.4%)	137051 (0.3%)	547504 (1.2%)	2532895 (5.6%)	7302134 (16.3%)	7707030 (17.2%)
RO_d13_2L	26875397 (100.0%)	22406570 (83.4%)	12818599 (47.7%)	187159 (0.7%)	656315 (2.4%)	1913073 (7.1%)	6831424 (25.4%)	4468827 (16.6%)
RO_d19_1L	35568296 (100.0%)	28997259 (81.5%)	20967496 (58.9%)	84878 (0.2%)	579689 (1.6%)	1319461 (3.7%)	6045735 (17.0%)	6571037 (18.5%)
RO_d21_1A	34796470 (100.0%)	24327139 (69.9%)	16372360 (47.1%)	140150 (0.4%)	310635 (0.9%)	1467801 (4.2%)	6036193 (17.3%)	10469331 (30.1%)
RO_d13_1A	47456005 (100.0%)	41173314 (86.8%)	34553991 (72.8%)	108526 (0.2%)	234456 (0.5%)	1620569 (3.4%)	4655772 (9.8%)	6282691 (13.2%)
RO_d13_1L	25805427 (100.0%)	21287895 (82.5%)	10190761 (39.5%)	182160 (0.7%)	874013 (3.4%)	2065038 (8.0%)	7975923 (30.9%)	4517532 (17.5%)
S14_d7_2A	30546305 (100.0%)	25864244 (84.7%)	18794758 (61.5%)	120007 (0.4%)	528385 (1.7%)	1737140 (5.7%)	4683954 (15.3%)	4682061 (15.3%)
S14_d9_2L	24089498 (100.0%)	20693587 (85.9%)	16827156 (69.9%)	104405 (0.4%)	256479 (1.1%)	1122747 (4.7%)	2382800 (9.9%)	3395911 (14.1%)
RO_d11_3A	35141059 (100.0%)	28076430 (79.9%)	17220988 (49.0%)	97678 (0.3%)	284076 (0.8%)	2287625 (6.5%)	8186063 (23.3%)	7064629 (20.1%)
RO_d15_2L	40687833 (100.0%)	34536397 (84.9%)	23709044 (58.3%)	156565 (0.4%)	654517 (1.6%)	1839348 (4.5%)	8176923 (20.1%)	6151436 (15.1%)
RO_d19_2L	35442115 (100.0%)	27896139 (78.7%)	18953760 (53.5%)	152721 (0.4%)	766359 (2.2%)	1422311 (4.0%)	6600988 (18.6%)	7545976 (21.3%)
S14_d14_3L	36507308 (100.0%)	30248894 (82.9%)	20162521 (55.2%)	159090 (0.4%)	875602 (2.4%)	1987126 (5.4%)	7064555 (19.4%)	6258414 (17.1%)
RO_d11_1A	31004010 (100.0%)	25420162 (82.0%)	14830567 (47.8%)	120033 (0.4%)	422064 (1.4%)	2307617 (7.4%)	7739881 (25.0%)	5583848 (18.0%)

sample	Trimmed reads	Genome aligned	ncRNA	Repeats	miRNA	coding	unannotated	not ro18
RO_d37_1L	44223521 (100.0%)	29682815 (67.1%)	22521764 (50.9%)	69613 (0.2%)	611713 (1.4%)	1051693 (2.4%)	5428032 (12.3%)	14540706 (32.9%)
S14_d9_2A	31427981 (100.0%)	27154828 (86.4%)	22493768 (71.6%)	83634 (0.3%)	276832 (0.9%)	1304896 (4.2%)	2995698 (9.5%)	4273153 (13.6%)
RO_d37_1A	30422426 (100.0%)	21065290 (69.2%)	12260745 (40.3%)	139883 (0.5%)	542392 (1.8%)	1698014 (5.6%)	6424256 (21.1%)	9357136 (30.8%)
S14_d14_3A	30634938 (100.0%)	26096311 (85.2%)	16726842 (54.6%)	102781 (0.3%)	452430 (1.5%)	2035110 (6.6%)	6779148 (22.1%)	4538627 (14.8%)
RO_d11_3L	34081793 (100.0%)	29308214 (86.0%)	20538638 (60.3%)	162968 (0.5%)	465656 (1.4%)	2163884 (6.3%)	5977068 (17.5%)	4773579 (14.0%)
RO_d19_2A	35893503 (100.0%)	25738124 (71.7%)	16800160 (46.8%)	185344 (0.5%)	466005 (1.3%)	1745752 (4.9%)	6540863 (18.2%)	10155379 (28.3%)
S14_d7_2L	26852495 (100.0%)	23076262 (85.9%)	19751725 (73.6%)	151603 (0.6%)	186451 (0.7%)	804076 (3.0%)	2182407 (8.1%)	3776233 (14.1%)
RO_d13_3A	34845166 (100.0%)	23418427 (67.2%)	13175319 (37.8%)	145166 (0.4%)	431278 (1.2%)	1902041 (5.5%)	7764623 (22.3%)	11426739 (32.8%)
RO_d15_2A	35791554 (100.0%)	26306144 (73.5%)	19839064 (55.4%)	136019 (0.4%)	321666 (0.9%)	1484143 (4.1%)	4525252 (12.6%)	9485410 (26.5%)
RO_d11_1L	39899028 (100.0%)	33669375 (84.4%)	20492674 (51.4%)	174267 (0.4%)	671030 (1.7%)	2712020 (6.8%)	9619384 (24.1%)	6229653 (15.6%)
RO_d13_3L	35166942 (100.0%)	27409723 (77.9%)	17541845 (49.9%)	89868 (0.3%)	265659 (0.8%)	1705830 (4.9%)	7806521 (22.2%)	7757219 (22.1%)

A.2 APPENDIX CHAPTER 3

Supplementary Figures and Tables for Chapter 3

Table A.5: **Mfuzz soft clustering membership scores for Sarisha-14 miRNAs across 5 expression clusters.** Each miRNA's membership score (0-1) indicates similarity to cluster expression patterns (described in header).Cluster 1 (low confidence/noise), Cluster 2 (early downregulation), Cluster 3 (late upregulation), Cluster 4 (late downregulation), Cluster 5 (early upregulation). The rightmost column indicates the cluster the miRNA is most highly associated with.

miRNA	Cluster 1 (low conf.)	Cluster 2 (down early)	Cluster 3 (up late)	Cluster 4 (down late)	Cluster 5 (up early)	Primary Cluster
Bra-miR1140	0.333	0.048	0.469	0.048	0.101	3
Bra-miR1511	0.455	0.02	0.448	0.02	0.057	1
Bra-miR156a	0.008	0.093	0.006	0.873	0.021	4
Bra-miR156d	0.01	0.036	0.006	0.902	0.046	4
Bra-miR156h	0.016	0.173	0.014	0.758	0.039	4
Bra-miR156i	0.034	0.688	0.044	0.197	0.037	2
Bra-miR156n	0.113	0.112	0.062	0.343	0.371	5
Bra-miR157a	0.006	0.054	0.005	0.915	0.021	4
Bra-miR158a	0.128	0.043	0.049	0.131	0.648	5
Bra-miR158b	0.039	0.796	0.064	0.073	0.029	2
Bra-miR159a	0.638	0.015	0.266	0.018	0.063	1
Bra-miR159c	0.448	0.019	0.459	0.02	0.054	3
Bra-miR160a	0.051	0.036	0.022	0.254	0.638	5
Bra-miR160g	0.137	0.357	0.322	0.118	0.066	2
Bra-miR162a	0.34	0.099	0.298	0.112	0.151	1
Bra-miR164a	0.075	0.063	0.034	0.389	0.439	5
Bra-miR164b	0.114	0.102	0.073	0.337	0.374	5
Bra-miR165a	0.21	0.017	0.731	0.014	0.028	3
Bra-miR166a	0.174	0.192	0.145	0.311	0.178	4
Bra-miR167a	0.005	0.948	0.006	0.036	0.006	2
Bra-miR167c	0.013	0.812	0.014	0.142	0.019	2
Bra-miR168a	0.041	0.046	0.021	0.562	0.33	4
Bra-miR168b	0.407	0.029	0.129	0.064	0.371	1
Bra-miR169a	0.004	0.926	0.004	0.06	0.006	2
Bra-miR169d	0.12	0.27	0.094	0.342	0.174	4
Bra-miR169e	0.01	0.893	0.011	0.072	0.013	2
Bra-miR169g	0.095	0.482	0.106	0.211	0.106	2
Bra-miR169h	0.07	0.436	0.058	0.331	0.106	2
Bra-miR169i	0.026	0.541	0.024	0.364	0.046	2

miRNA	Cluster 1 (low conf.)	Cluster 2 (down early)	Cluster 3 (up late)	Cluster 4 (down late)	Cluster 5 (up early)	Primary Cluster
Bra-miR169j	0.322	0.096	0.426	0.063	0.094	3
Bra-miR169l	0.038	0.733	0.04	0.149	0.041	2
Bra-miR169q	0.006	0.959	0.007	0.023	0.005	2
Bra-miR170	0.016	0.85	0.02	0.096	0.019	2
Bra-miR171a	0.185	0.139	0.143	0.261	0.272	5
Bra-miR171e	0.046	0.748	0.059	0.107	0.04	2
Bra-miR171g	0.119	0.244	0.088	0.369	0.181	4
Bra-miR172a	0.159	0.004	0.826	0.003	0.008	3
Bra-miR172c	0.079	0.004	0.909	0.003	0.006	3
Bra-miR172d	0.265	0.005	0.713	0.005	0.013	3
Bra-miR2111a	0.404	0.092	0.3	0.068	0.136	1
Bra-miR2111c	0.705	0.014	0.104	0.022	0.155	1
Bra-miR2111d	0.704	0.016	0.152	0.022	0.106	1
Bra-miR2111f	0.63	0.014	0.3	0.014	0.043	1
Bra-miR319a	0.625	0.009	0.321	0.01	0.035	1
Bra-miR319c	0.411	0.007	0.555	0.007	0.02	3
Bra-miR319e	0.092	0.013	0.878	0.007	0.011	3
Bra-miR319f	0.102	0.013	0.866	0.007	0.011	3
Bra-miR390a	0.043	0.095	0.027	0.667	0.168	4
Bra-miR391a	0.028	0.412	0.028	0.479	0.053	4
Bra-miR393a	0.23	0.095	0.113	0.169	0.392	5
Bra-miR393d	0.239	0.238	0.194	0.154	0.174	1
Bra-miR394a	0.083	0.004	0.903	0.003	0.007	3
Bra-miR395a	0.476	0.033	0.338	0.043	0.11	1
Bra-miR395g	0.504	0.034	0.216	0.056	0.189	1
Bra-miR396a	0.013	0.317	0.011	0.63	0.03	4
Bra-miR397	0.816	0.006	0.127	0.008	0.042	1
Bra-miR398a	0.1	0.011	0.025	0.036	0.828	5
Bra-miR398c	0.459	0.017	0.081	0.037	0.407	1
Bra-miR399a	0.007	0.943	0.009	0.033	0.007	2
Bra-miR399e	0.046	0.756	0.059	0.103	0.036	2
Bra-miR399h	0.134	0.443	0.226	0.129	0.067	2
Bra-miR400	0.109	0.023	0.036	0.063	0.77	5
Bra-miR403	0.484	0.041	0.288	0.045	0.141	1
Bra-miR408	0.191	0.017	0.044	0.046	0.702	5
Bra-miR5654a	0.023	0.123	0.016	0.756	0.081	4
Bra-miR5711	0.156	0.126	0.628	0.048	0.042	3

miRNA	Cluster 1 (low conf.)	Cluster 2 (down early)	Cluster 3 (up late)	Cluster 4 (down late)	Cluster 5 (up early)	Primary Cluster
Bra-miR5712b	0.605	0.026	0.169	0.032	0.169	1
Bra-miR5718	0.006	0.938	0.007	0.042	0.007	2
Bra-miR5722	0.008	0.849	0.009	0.122	0.012	2
Bra-miR5723	0.277	0.059	0.568	0.036	0.059	3
Bra-miR5724	0.086	0.007	0.896	0.004	0.007	3
Bra-miR6030	0.087	0.194	0.071	0.497	0.151	4
Bra-miR6032	0.069	0.674	0.148	0.073	0.036	2
Bra-miR827	0.359	0.061	0.372	0.063	0.145	3
Bra-miR858a	0.005	0.953	0.006	0.031	0.006	2
Bra-miR858b	0.088	0.013	0.88	0.007	0.011	3
Bra-miR860	0.532	0.026	0.126	0.041	0.275	1
Bra-miR9552a	0.008	0.027	0.005	0.925	0.035	4
Bra-miR9554	0.274	0.091	0.515	0.049	0.071	3
Bra-miR9555a	0.311	0.127	0.171	0.131	0.26	1
Bra-miR9556	0.313	0.137	0.232	0.108	0.21	1
Bra-miR9557	0.319	0.04	0.565	0.026	0.049	3
Bra-miR9558	0.067	0.042	0.03	0.178	0.683	5
Bra-miR9560a	0.549	0.026	0.212	0.04	0.173	1
Bra-miR9563	0.194	0.211	0.216	0.199	0.181	3
Bra-miRN271	0.277	0.113	0.165	0.168	0.277	1
Bra-miRN317	0.611	0.013	0.311	0.014	0.051	1
Bra-miRN318	0.434	0.016	0.075	0.033	0.442	5
Bra-miRN328	0.036	0.011	0.013	0.053	0.887	5
Bra-miRN330	0.203	0.212	0.336	0.146	0.102	3
Bra-miRN331	0.134	0.325	0.143	0.252	0.146	2
Bra-miRN334	0.322	0.076	0.47	0.047	0.085	3
Bra-miRN335	0.085	0.014	0.883	0.007	0.012	3
Bra-miRN337	0.104	0.034	0.045	0.143	0.674	5
Bra-miRN338	0.053	0.719	0.074	0.112	0.042	2
Bra-miRN339	0.022	0.757	0.026	0.168	0.027	2
Bra-miRN340	0.079	0.011	0.022	0.037	0.85	5
Bra-miRN343	0.318	0.135	0.236	0.11	0.2	1
Bra-miRN344	0.028	0.526	0.025	0.376	0.045	2
Bra-miRN346	0.23	0.027	0.682	0.023	0.039	3
Bra-miRN347	0.009	0.031	0.006	0.899	0.055	4
Bra-miRN348a	0.045	0.02	0.018	0.118	0.798	5
Bra-miRN348b	0.203	0.063	0.105	0.159	0.47	5

miRNA	Cluster 1 (low conf.)	Cluster 2 (down early)	Cluster 3 (up late)	Cluster 4 (down late)	Cluster 5 (up early)	Primary Cluster
Bra-miRN349	0.337	0.019	0.583	0.017	0.043	3
Bra-miRN350	0.005	0.948	0.006	0.036	0.006	2
Bra-miRN351	0.004	0.942	0.005	0.042	0.006	2
Bra-miRN354	0.107	0.112	0.066	0.424	0.291	4
Bra-miRN355	0.135	0.323	0.176	0.241	0.125	2
Bra-miRN356	0.056	0.003	0.934	0.002	0.004	3
Bra-miRN358	0.06	0.563	0.057	0.249	0.071	2
Bra-miRN360	0.343	0.019	0.067	0.04	0.531	5
Bra-miRN361	0.243	0.211	0.275	0.131	0.139	3
Bra-miRN362a	0.212	0.055	0.642	0.039	0.052	3
Bra-miRN363	0.044	0.078	0.025	0.583	0.271	4
Bra-miRN366	0.004	0.969	0.005	0.019	0.004	2
Bra-miRN367	0.012	0.038	0.007	0.88	0.063	4
Bra-miRN368a	0.044	0.454	0.039	0.379	0.082	2
Bra-miRN369	0.03	0.012	0.011	0.063	0.884	5
Bra-miRN372	0.248	0.159	0.452	0.064	0.078	3
Bra-miRN376	0.131	0.441	0.205	0.139	0.084	2
Bra-miRN377a	0.102	0.131	0.072	0.474	0.22	4
Bra-miRN378	0.172	0.215	0.461	0.09	0.062	3
Bra-miRN380	0.194	0.171	0.176	0.254	0.205	4

Table A.6: **Mfuzz soft clustering membership scores for R-o-18**. miRNAs across 5 expression clusters. Each miRNA's membership score (0-1) indicates similarity to cluster expression patterns: Cluster 1 (early downregulation), Cluster 2 (peak abundance), Cluster 3 (late downregulation), Cluster 4 (early upregulation), Cluster 5 (low confidence/noise). The rightmost column indicates the cluster the miRNA is most highly associated with.

miRNA	Cluster 1 (down early)	Cluster 2 (peak)	Cluster 3 (down late)	Cluster 4 (up early)	Cluster 5 (low conf.)	Primary cluster
Bra-miR1140	0.079	0.052	0.399	0.175	0.295	3
Bra-miR1511	0.087	0.082	0.388	0.131	0.313	3
Bra-miR156a	0.062	0.806	0.04	0.015	0.077	2
Bra-miR156d	0.03	0.085	0.071	0.018	0.796	5
Bra-miR156h	0.095	0.797	0.042	0.018	0.047	2
Bra-miR156i	0.596	0.249	0.059	0.041	0.054	1
Bra-miR157a	0.021	0.934	0.015	0.005	0.025	2
Bra-miR158a	0.091	0.136	0.501	0.1	0.171	3
Bra-miR158b	0.213	0.549	0.095	0.045	0.099	2
Bra-miR159a	0.032	0.011	0.06	0.88	0.018	4
Bra-miR159c	0.057	0.017	0.076	0.818	0.031	4
Bra-miR160a	0.103	0.59	0.105	0.037	0.165	2
Bra-miR160g	0.084	0.061	0.52	0.237	0.098	3
Bra-miR162a	0.257	0.136	0.302	0.205	0.1	3
Bra-miR164a	0.041	0.909	0.021	0.008	0.021	2
Bra-miR164b	0.129	0.064	0.31	0.418	0.079	4
Bra-miR165a	0.062	0.02	0.163	0.714	0.042	4
Bra-miR166a	0.224	0.149	0.31	0.224	0.093	3
Bra-miR167a	0.108	0.143	0.124	0.047	0.578	5
Bra-miR167c	0.102	0.084	0.215	0.084	0.516	5
Bra-miR168a	0.29	0.345	0.191	0.089	0.085	2
Bra-miR168b	0.148	0.032	0.132	0.648	0.039	4
Bra-miR169a	0.128	0.787	0.034	0.017	0.034	2
Bra-miR169e	0.273	0.597	0.053	0.029	0.049	2
Bra-miR169g	0.614	0.098	0.086	0.134	0.067	1
Bra-miR169h	0.175	0.112	0.229	0.112	0.373	5
Bra-miR169i	0.603	0.205	0.083	0.061	0.048	1
Bra-miR169j	0.247	0.297	0.242	0.103	0.11	2
Bra-miR169l	0.15	0.264	0.189	0.08	0.317	5
Bra-miR169q	0.638	0.125	0.079	0.093	0.065	1
Bra-miR170	0.115	0.144	0.22	0.068	0.453	5
Bra-miR171a	0.109	0.458	0.158	0.046	0.229	2
Bra-miR171e	0.062	0.079	0.375	0.064	0.421	5

miRNA	Cluster 1 (down early)	Cluster 2 (peak)	Cluster 3 (down late)	Cluster 4 (up early)	Cluster 5 (low conf.)	Primary cluster
Bra-miR171g	0.175	0.157	0.391	0.139	0.138	3
Bra-miR172a	0.014	0.005	0.028	0.945	0.009	4
Bra-miR172c	0.049	0.021	0.162	0.707	0.062	4
Bra-miR172d	0.026	0.009	0.064	0.878	0.022	4
Bra-miR2111a	0.634	0.141	0.087	0.089	0.049	1
Bra-miR2111c	0.267	0.19	0.23	0.184	0.128	1
Bra-miR2111d	0.225	0.598	0.07	0.038	0.068	2
Bra-miR2111f	0.126	0.051	0.324	0.367	0.132	4
Bra-miR319a	0.029	0.008	0.039	0.912	0.012	4
Bra-miR319c	0.021	0.006	0.028	0.937	0.009	4
Bra-miR319e	0.061	0.018	0.092	0.801	0.027	4
Bra-miR319f	0.059	0.018	0.088	0.81	0.026	4
Bra-miR390a	0.222	0.514	0.08	0.038	0.145	2
Bra-miR391a	0.136	0.042	0.179	0.532	0.112	4
Bra-miR393a	0.178	0.717	0.04	0.022	0.042	2
Bra-miR393d	0.303	0.528	0.074	0.043	0.052	2
Bra-miR394a	0.021	0.006	0.034	0.929	0.009	4
Bra-miR395a	0.668	0.112	0.086	0.071	0.063	1
Bra-miR395g	0.554	0.08	0.128	0.16	0.077	1
Bra-miR396a	0.026	0.933	0.014	0.005	0.021	2
Bra-miR397	0.04	0.07	0.256	0.042	0.591	5
Bra-miR398a	0.032	0.045	0.133	0.03	0.76	5
Bra-miR398c	0.037	0.042	0.15	0.039	0.731	5
Bra-miR399a	0.295	0.102	0.232	0.185	0.186	1
Bra-miR399e	0.562	0.054	0.087	0.255	0.042	1
Bra-miR399h	0.195	0.226	0.176	0.074	0.328	5
Bra-miR400	0.083	0.13	0.26	0.083	0.445	5
Bra-miR403	0.149	0.168	0.389	0.15	0.143	3
Bra-miR408	0.027	0.053	0.096	0.021	0.804	5
Bra-miR5654a	0.083	0.171	0.245	0.076	0.425	5
Bra-miR5711	0.059	0.017	0.093	0.81	0.021	4
Bra-miR5712b	0.098	0.139	0.148	0.07	0.544	5
Bra-miR5714	0.069	0.039	0.335	0.368	0.189	4
Bra-miR5718	0.03	0.899	0.025	0.008	0.038	2
Bra-miR5722	0.505	0.266	0.084	0.065	0.08	1
Bra-miR5723	0.018	0.005	0.03	0.938	0.009	4
Bra-miR5724	0.096	0.158	0.473	0.091	0.182	3

miRNA	Cluster 1 (down early)	Cluster 2 (peak)	Cluster 3 (down late)	Cluster 4 (up early)	Cluster 5 (low conf.)	Primary cluster
Bra-miR6030	0.249	0.34	0.105	0.056	0.25	2
Bra-miR6032	0.757	0.071	0.063	0.074	0.036	1
Bra-miR827	0.159	0.694	0.061	0.033	0.054	2
Bra-miR858a	0.05	0.051	0.419	0.105	0.375	3
Bra-miR858b	0.055	0.025	0.236	0.598	0.087	4
Bra-miR860	0.075	0.036	0.231	0.586	0.072	4
Bra-miR9552a	0.08	0.266	0.228	0.058	0.367	5
Bra-miR9552b	0.058	0.029	0.382	0.465	0.066	4
Bra-miR9554	0.046	0.016	0.092	0.817	0.028	4
Bra-miR9555a	0.152	0.105	0.358	0.122	0.262	3
Bra-miR9557	0.153	0.149	0.413	0.181	0.105	3
Bra-miR9558	0.036	0.062	0.095	0.029	0.778	5
Bra-miR9560a	0.052	0.02	0.176	0.712	0.04	4
Bra-miR9563	0.07	0.053	0.463	0.188	0.227	3
Bra-miRN252	0.132	0.044	0.241	0.523	0.061	4
Bra-miRN271	0.111	0.02	0.069	0.775	0.025	4
Bra-miRN317	0.109	0.035	0.251	0.527	0.078	4
Bra-miRN318	0.141	0.103	0.219	0.17	0.367	5
Bra-miRN319	0.111	0.285	0.169	0.071	0.364	5
Bra-miRN324a	0.104	0.081	0.317	0.234	0.264	3
Bra-miRN328	0.105	0.448	0.192	0.058	0.197	2
Bra-miRN330	0.584	0.118	0.102	0.133	0.063	1
Bra-miRN331	0.255	0.152	0.193	0.236	0.164	1
Bra-miRN333a	0.097	0.071	0.427	0.153	0.252	3
Bra-miRN334	0.033	0.084	0.115	0.026	0.743	5
Bra-miRN335	0.22	0.064	0.197	0.385	0.134	4
Bra-miRN337	0.11	0.128	0.245	0.115	0.401	5
Bra-miRN338	0.239	0.147	0.158	0.095	0.362	5
Bra-miRN339	0.221	0.306	0.247	0.106	0.121	2
Bra-miRN340	0.049	0.785	0.047	0.016	0.104	2
Bra-miRN341	0.159	0.399	0.194	0.09	0.158	2
Bra-miRN343	0.119	0.041	0.213	0.573	0.054	4
Bra-miRN344	0.04	0.129	0.099	0.023	0.709	5
Bra-miRN346	0.256	0.039	0.112	0.556	0.037	4
Bra-miRN347	0.081	0.516	0.145	0.041	0.217	2
Bra-miRN348a	0.162	0.251	0.141	0.069	0.377	5
Bra-miRN348b	0.09	0.108	0.145	0.065	0.592	5

miRNA	Cluster 1 (down early)	Cluster 2 (peak)	Cluster 3 (down late)	Cluster 4 (up early)	Cluster 5 (low conf.)	Primary cluster
Bra-miRN349	0.099	0.036	0.155	0.66	0.05	4
Bra-miRN350	0.041	0.852	0.037	0.012	0.058	2
Bra-miRN351	0.204	0.054	0.188	0.445	0.109	4
Bra-miRN354	0.174	0.092	0.24	0.196	0.298	5
Bra-miRN355	0.39	0.095	0.138	0.297	0.08	1
Bra-miRN358	0.143	0.036	0.17	0.581	0.07	4
Bra-miRN360	0.057	0.042	0.535	0.223	0.143	3
Bra-miRN362a	0.188	0.102	0.336	0.264	0.11	3
Bra-miRN363	0.098	0.105	0.46	0.103	0.234	3
Bra-miRN366	0.035	0.912	0.018	0.007	0.028	2
Bra-miRN367	0.059	0.704	0.08	0.023	0.133	2
Bra-miRN368a	0.092	0.154	0.193	0.076	0.485	5
Bra-miRN369	0.033	0.903	0.026	0.009	0.029	2
Bra-miRN370	0.033	0.008	0.035	0.913	0.011	4
Bra-miRN372	0.225	0.159	0.309	0.215	0.093	3
Bra-miRN373	0.096	0.037	0.228	0.594	0.046	4
Bra-miRN375	0.025	0.007	0.043	0.914	0.012	4
Bra-miRN376	0.109	0.097	0.264	0.137	0.393	5
Bra-miRN378	0.729	0.079	0.068	0.078	0.046	1
Bra-miRN380	0.074	0.063	0.411	0.161	0.291	3
Bra-miRN381	0.058	0.063	0.133	0.051	0.696	5

Table A.7: miRNAs with highest cluster membership (>0.6) in Sarisha-14 (S14) and R-o-18 (RO) Mfuzz clusters. Each column lists the miRNAs most strongly associated with that cluster (membership score >0.6). RO Cluster 5 was not included due to no miRNAs found above this threshold.

S14 Cluster 1	S14 Cluster 2	S14 Cluster 3	S14 Cluster 4	S14 Cluster 5	RO Cluster 1	RO Cluster 2	RO Cluster 3	RO Cluster 4
Bra-miR2111a	Bra-miR156i	Bra-miR159a	Bra-miR168b	Bra-miR156a	Bra-miR159a	Bra-miR156d	Bra-miR156a	Bra-miR169q
Bra-miR9554	Bra-miR158b	Bra-miR159c	Bra-miR398a	Bra-miR156d	Bra-miR159c	Bra-miR398a	Bra-miR156h	Bra-miR169t
Bra-miRN372	Bra-miR167a	Bra-miR165a	Bra-miR398c	Bra-miR156h	Bra-miR165a	Bra-miR398c	Bra-miR157a	Bra-miR169u
	Bra-miR167c	Bra-miR172a	Bra-miR400	Bra-miR157a	Bra-miR168b	Bra-miR408	Bra-miR164a	Bra-miR2111a
	Bra-miR169a	Bra-miR172c	Bra-miR408	Bra-miR168a	Bra-miR172a	Bra-miR9558	Bra-miR169a	Bra-miR395a
	Bra-miR169e	Bra-miR172d	Bra-miRN318	Bra-miR390a	Bra-miR172c	Bra-miRN334	Bra-miR393a	Bra-miR6032
	Bra-miR169i	Bra-miR2111f	Bra-miRN328	Bra-miR5654a	Bra-miR172d	Bra-miRN344	Bra-miR396a	Bra-miRN378
	Bra-miR169l	Bra-miR319a	Bra-miRN337	Bra-miR9552a	Bra-miR319a	Bra-miRN381	Bra-miR5718	
	Bra-miR169q	Bra-miR319c	Bra-miRN340	Bra-miRN347	Bra-miR319c	Bra-miR827		
	Bra-miR170	Bra-miR319e	Bra-miRN348a	Bra-miRN363	Bra-miR319e	Bra-miRN340		
	Bra-miR171e	Bra-miR319f	Bra-miRN360	Bra-miRN367	Bra-miR319f	Bra-miRN350		
	Bra-miR399a	Bra-miR394a	Bra-miRN369	Bra-miR394a	Bra-miRN366			
	Bra-miR399e	Bra-miR5724	Bra-miR5711	Bra-miRN367				
	Bra-miR5718	Bra-miR858b	Bra-miR5723	Bra-miRN369				
	Bra-miR5722	Bra-miRN317	Bra-miR9554					
	Bra-miR858a	Bra-miRN335	Bra-miR9560a					
	Bra-miRN338	Bra-miRN346	Bra-miRN271					
	Bra-miRN339	Bra-miRN349	Bra-miRN349					
	Bra-miRN350	Bra-miRN356	Bra-maiRN370					
	Bra-miRN351	Bra-miRN362a	Bra-miRN375					
	Bra-miRN366							

Table A.8: **TargetFinder predictions for target genes of candidate flowering-associated miRNAs**. The table lists all predicted target genes (brapa_id) for *B. rapa* miRNAs discussed in Chapter 3, with their *A. thaliana* orthologues (ath_orth), targetFinder scores (lower = more complementary bases), target binding site sequence, and base pairing patterns. Colons (:) indicate Watson-Crick matches, dots (.) mismatches, and spaces indicate gaps in the miRNA–target duplex.

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR156a	A07p029860.1_BraROA	SPL15_A07	1	GUGCUCUCUCUCUUCUGUCA	:::::: ::::::::::::
Bra-miR156a	A04p037800.1_BraROA	SPL9_A04	1	GUGCUCUCUCUCUUCUGUCA	:::::: ::::::::::::
Bra-miR156a	A05p004000.1_BraROA	SPL9_A05	1	GUGCUCUCUCUCUUCUGUCA	:::::: ::::::::::::
Bra-miR156a	A02p022860.1_BraROA	AT1G69170_A02	1	GUGCUCUCUCUCUUCUGUCA	:::::: ::::::::::::
Bra-miR156a	A06p056880.1_BraROA	SPL2_A06	1	GUGCUCUCUCUCUUCUGUCA	:::::: ::::::::::::
Bra-miR156a	A09p023610.1_BraROA	SPL2_A09	1	GUGCUCUCUCUCUUCUGUCA	:::::: ::::::::::::
Bra-miR156a	A04p003490.1_BraROA	SPL15_A04	1.5	GUGUUCUCUCUCUUCUGUCA	:::.:: ::::::::::::
Bra-miR156a	A10p018910.1_BraROA	AT5G55830_A10	2	GUGCUCUCUUUCUUCUGUCA	:::::: ::.::::::::::
Bra-miR156a	A03p017260.1_BraROA	SPL13B_A03	2	GCGCUCUCUCUCUUCUGUCA	: :::: ::::::::::::
Bra-miR156a	A03p017260.1_BraROA	SPL13A_A03	2	GCGCUCUCUCUCUUCUGUCA	: :::: ::::::::::::
Bra-miR156a	A06p026580.1_BraROA	AT3G43240_A06	3.5	GUGACUUACUCUCUUCUGACA	::: ::.:::::::::: ::
Bra-miR156a	A07p052440.1_BraROA	CBSX4_A07	3.5	CUGCUCAUGUCUCUUCUGUCA	::::::. ::::::::::::
Bra-miR156a	A06p038110.1_BraROA	AT5G64170_A06	4	GUGCUGGUUCUCUUCUGCUCA	::::: ..::::::::: :::
Bra-miR156a	A03p023440.1_BraROA	AT2G40460_A03	4	GUGAACACUCUCUUC-GUCA	::: ::::::::: ::::
Bra-miR156a	A06p037740.1_BraROA	AAP4_A06	4	GUUCUUGUUCUCUUUUGUCA	:: ::...::::::.:::::
Bra-miR156a	A08p041500.1_BraROA	AT1G12340_A08	4	UUUCUCACUAUCUUCUGUCA	: :::::: :::::::::
Bra-miR156a	A08p041500.1_BraROA	AT1G12390_A08	4	UUUCUCACUAUCUUCUGUCA	: :::::: :::::::::
Bra-miR156a	A09p082350.1_BraROA	PCS2_A09	4	GUGUUCACUGCUCUUCUGUUG	:::.::::: :::::::..
Bra-miR156d	A10p018910.1_BraROA	AT5G55830_A10	0	GUGCUCUCUUUCUUCUGUCAA	::::::::::::::::::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR156d	A07p029860.1_BraROA	SPL15_A07	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR156d	A04p037800.1_BraROA	SPL9_A04	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR156d	A05p004000.1_BraROA	SPL9_A05	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR156d	A06p056880.1_BraROA	SPL2_A06	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR156d	A09p023610.1_BraROA	SPL2_A09	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR156d	A04p003490.1_BraROA	SPL15_A04	2.5	GUGUUCUCUCUCUUCUGUCAA	:::.::::: ::::::::::
Bra-miR156d	A02p022860.1_BraROA	AT1G69170_A02	2.5	GUGCUCUCUCUCUUCUGUCAG	::::::::: :::::::::.
Bra-miR156d	A01p059560.1_BraROA	AT3G01175_A01	3.5	UUGUUCUCUUUCUU-UGUCAA	::.::::::::: ::::::
Bra-miR156d	A03p041470.1_BraROA	SPL5_A03	4	UCGCUGCUCUUUCUUCUGUCAC	::: ::::::::::::::
Bra-miR156d	A02p029100.1_BraROA	OPR1_A02	4	GUGGUGUCUU-CUUCUGUCAA	::: : :::: ::::::::::
Bra-miR156d	A02p029100.1_BraROA	OPR2_A02	4	GUGGUGUCUU-CUUCUGUCAA	::: : :::: ::::::::::
Bra-miR156d	A03p017260.1_BraROA	SPL13B_A03	4	GCGCUCUCUCUCUUCUGUCAU	: ::::::: :::::::::
Bra-miR156d	A03p017260.1_BraROA	SPL13A_A03	4	GCGCUCUCUCUCUUCUGUCAU	: ::::::: :::::::::
Bra-miR156d	A05p038890.1_BraROA	AT1G12672_A05	4	GUGCUUCCCUUUCUUCUCUCAA	::::: : :::::::: ::::
Bra-miR156d	A08p015190.1_BraROA	XGD1_A08	4	AUGCUCUCUUUCU-CUGUUAA	:::::::::::: :::::.::
Bra-miR156d	A09p071710.1_BraROA	AT2G21630_A09	4	CAGCUCUCUUUCUUC-GUCAA	:::::::::::: :::::
Bra-miR156h	A07p029860.1_BraROA	SPL15_A07	2	UGUGCUCUCUCUCUUCUGUC	::::::: ::::::::::::
Bra-miR156h	A04p037800.1_BraROA	SPL9_A04	2	UGUGCUCUCUCUCUUCUGUC	::::::: ::::::::::::
Bra-miR156h	A05p004000.1_BraROA	SPL9_A05	2	UGUGCUCUCUCUCUUCUGUC	::::::: ::::::::::::
Bra-miR156h	A02p022860.1_BraROA	AT1G69170_A02	2	UGUGCUCUCUCUCUUCUGUC	::::::: ::::::::::::
Bra-miR156h	A04p003490.1_BraROA	SPL15_A04	2.5	UGUGUUCUCUCUCUUCUGUC	::::.:: ::::::::::::
Bra-miR156h	A10p018910.1_BraROA	AT5G55830_A10	3	UGUGCUCUCUUUCUUCUGUC	::::::: ::.:::::::::
Bra-miR156h	A03p017260.1_BraROA	SPL13B_A03	3	UGCGCUCUCUCUCUUCUGUC	:: :::: ::::::::::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR156h	A03p017260.1_BraROA	SPL13A_A03	3	UGCGCUCUCUCUCUUCUGUC	:: :::: ::::::::::::
Bra-miR156h	A06p056880.1_BraROA	SPL2_A06	3	GGUGCUCUCUCUCUUCUGUC	:::::: ::::::::::::
Bra-miR156h	A09p023610.1_BraROA	SPL2_A09	3	GGUGCUCUCUCUCUUCUGUC	:::::: ::::::::::::
Bra-miR156h	A09p082350.1_BraROA	PCS2_A09	3	UGUGUUCACUGCUCUUCUGUU	::::.:::::: :::::::::.
Bra-miR156h	A08p034660.1_BraROA	SPL10_A08	4	UGCUCUCUCUCUCUUCUGUC	:: ::: ::::::::::::
Bra-miR156i	A07p029860.1_BraROA	SPL15_A07	1	GUGCUCUCUCUCUUCUGUCAA	:::::: ::::::::::::::
Bra-miR156i	A04p037800.1_BraROA	SPL9_A04	1	GUGCUCUCUCUCUUCUGUCAA	:::::: ::::::::::::::
Bra-miR156i	A05p004000.1_BraROA	SPL9_A05	1	GUGCUCUCUCUCUUCUGUCAA	:::::: ::::::::::::::
Bra-miR156i	A06p056880.1_BraROA	SPL2_A06	1	GUGCUCUCUCUCUUCUGUCAA	:::::: ::::::::::::::
Bra-miR156i	A09p023610.1_BraROA	SPL2_A09	1	GUGCUCUCUCUCUUCUGUCAA	:::::: ::::::::::::::
Bra-miR156i	A04p003490.1_BraROA	SPL15_A04	1.5	GUGUUCUCUCUCUUCUGUCAA	:::.:: ::::::::::::::
Bra-miR156i	A02p022860.1_BraROA	AT1G69170_A02	1.5	GUGCUCUCUCUCUUCUGUCAG	:::::: :::::::::::::.
Bra-miR156i	A10p018910.1_BraROA	AT5G55830_A10	2	GUGCUCUCUUUCUUCUGUCAA	:::::: ::.::::::::::::
Bra-miR156i	A03p017260.1_BraROA	SPL13B_A03	3	GCGCUCUCUCUCUUCUGUCAU	: :::: ::::::::::::::
Bra-miR156i	A03p017260.1_BraROA	SPL13A_A03	3	GCGCUCUCUCUCUUCUGUCAU	: :::: ::::::::::::::
Bra-miR156i	A06p038110.1_BraROA	AT5G64170_A06	4	GUGCUGGUUCUCUUCUGCUCAG	::::: :.::::::::: :::.
Bra-miR156i	A06p037740.1_BraROA	AAP4_A06	4	GUUCUUGUUCUCUUUUGUCAC	:: ::.:.::::::.:::::
Bra-miR157a	A07p029860.1_BraROA	SPL15_A07	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR157a	A04p037800.1_BraROA	SPL9_A04	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR157a	A05p004000.1_BraROA	SPL9_A05	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR157a	A10p018910.1_BraROA	AT5G55830_A10	2	GUGCUCUCUUUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR157a	A06p056880.1_BraROA	SPL2_A06	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::
Bra-miR157a	A09p023610.1_BraROA	SPL2_A09	2	GUGCUCUCUCUCUUCUGUCAA	::::::::: ::::::::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR157a	A04p003490.1_BraROA	SPL15_A04	2.5	GUGUUCUCUCUCUUCUGUCAA	:::.::::: ::::::::::
Bra-miR157a	A02p022860.1_BraROA	AT1G69170_A02	2.5	GUGCUCUCUCUCUUCUGUCAG	::::::::: :::::::::.
Bra-miR157a	A04p036770.1_BraROA	AT2G40630_A04	3.5	CUGC-CGUUAUCUUCUGUCAA	::: : .::::::::::::::
Bra-miR157a	A05p039460.1_BraROA	AT5G51000_A05	3.5	GUGAU-UCUAUCUUCUGUUAG	::: : ::::::::::::.:.
Bra-miR157a	A02p029100.1_BraROA	OPR1_A02	4	GUGGUGUCU-UCUUCUGUCAA	::: : ::: ::::::::::
Bra-miR157a	A02p029100.1_BraROA	OPR2_A02	4	GUGGUGUCU-UCUUCUGUCAA	::: : ::: ::::::::::
Bra-miR157a	A03p017260.1_BraROA	SPL13B_A03	4	GCGCUCUCUCUCUUCUGUCAU	: ::::::: :::::::::
Bra-miR157a	A03p017260.1_BraROA	SPL13A_A03	4	GCGCUCUCUCUCUUCUGUCAU	: ::::::: :::::::::
Bra-miR157a	A05p039200.1_BraROA	AT3G20690_A05	4	GUG-UCUCUAUCUUCGGUUAA	::: :::::::::: ::.::
Bra-miR157a	A05p039200.1_BraROA	AT3G20700_A05	4	GUG-UCUCUAUCUUCGGUUAA	::: :::::::::: ::.::
Bra-miR157a	A06g500400.1_BraROA	AT1G48090_A06	4	GAGCUCUGGAUCUUCUGUCAA	: ::::: ::::::::::::
Bra-miR157a	A08p041500.1_BraROA	AT1G12340_A08	4	UUUCUCACUAUCUUCUGUCAU	: ::: ::::::::::::
Bra-miR157a	A08p041500.1_BraROA	AT1G12390_A08	4	UUUCUCACUAUCUUCUGUCAU	: ::: ::::::::::::
Bra-miR159c	A03p018660.1_BraROA	MYB101_A03	2.5	AGAGCUUCCUUCAGUCCAAA	:::::.::::::.::::::
Bra-miR159c	A03p018660.1_BraROA	MYB101_A03	2.5	AGAGCUUCCUUCAGUCCAAA	:::::.::::::.::::::
Bra-miR159c	A01p053820.1_BraROA	MYB65_A01	3	GGAGCUCCCUUCAUUCCAAU	:::::::::::: :::::
Bra-miR159c	A01p053820.1_BraROA	MYB65_A01	3	GGAGCUCCCUUCAUUCCAAU	:::::::::::: :::::
Bra-miR159c	A02p015870.1_BraROA	MYB97_A02	3	AGAGCUCCCUUCAAACCAAA	:::::::::::: :::::
Bra-miR159c	A02p015870.1_BraROA	MYB97_A02	3	AGAGCUCCCUUCAAACCAAA	:::::::::::: :::::
Bra-miR159c	A06p032910.1_BraROA	ACT12_A06	3	GAAGCUCCCCUCAAUCCAAA	: ::::::: :::::::::
Bra-miR159c	A06p032910.1_BraROA	ACT12_A06	3	GAAGCUCCCCUCAAUCCAAA	: ::::::: :::::::::
Bra-miR159c	A04p030440.1_BraROA	MYB101_A04	3.5	AGAGCUUCCUUCAAACCAAA	:::::.::::::: :::::
Bra-miR159c	A04p030440.1_BraROA	MYB101_A04	3.5	AGAGCUUCCUUCAAACCAAA	:::::.::::::: :::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR159c	A01p020080.1_BraROA	SPL_A01	4	UGAGCUCUCUUCAAUCCCAA	:::::::::::::::::: ::
Bra-miR159c	A01p020080.1_BraROA	SPL_A01	4	UGAGCUCUCUUCAAUCCCAA	:::::::::::::::::: ::
Bra-miR159c	A03p014830.1_BraROA	MYB120_A03	4	AGACCUCCCUUCAAACCAAA	:: :::::::::: :::::
Bra-miR159c	A03p014830.1_BraROA	MYB120_A03	4	AGACCUCCCUUCAAACCAAA	:: :::::::::: :::::
Bra-miR159c	A03p019540.1_BraROA	AT2G34010_A03	4	AGAGCCCCCUUCAAACCAAA	:::: :::::::: :::::
Bra-miR159c	A03p019540.1_BraROA	AT2G34010_A03	4	AGAGCCCCCUUCAAACCAAA	:::: :::::::: :::::
Bra-miR159c	A03p040620.1_BraROA	AT3G13800_A03	4	CAAGCUUCCCUUCAAUCCAAU	:::: ::::::::::::
Bra-miR159c	A03p040620.1_BraROA	AT3G13800_A03	4	CAAGCUUCCCUUCAAUCCAAU	:::: ::::::::::::
Bra-miR159c	A06g503660.1_BraROA	ACT12_A06	4	GAAGCGCCCCUCAAUCCAAA	: ::: ::: :::::::::
Bra-miR159c	A06g503660.1_BraROA	ACT12_A06	4	GAAGCGCCCCUCAAUCCAAA	: ::: ::: :::::::::
Bra-miR159c	A06p023690.1_BraROA	ACT12_A06	4	GAAGCGCCCCUCAAUCCAAA	: ::: ::: :::::::::
Bra-miR159c	A06p023690.1_BraROA	ACT12_A06	4	GAAGCGCCCCUCAAUCCAAA	: ::: ::: :::::::::
Bra-miR159c	A09p001230.1_BraROA	AT4G01570_A09	4	CCGGCCUCCUUCAAUCCAAA	.:: .::::::::::::
Bra-miR159c	A09p001230.1_BraROA	AT4G01570_A09	4	CCGGCCUCCUUCAAUCCAAA	.:: .::::::::::::
Bra-miR169e	A03p047750.1_BraROA	AT2G05760_A03	4	CAUACAAAUCAUUCUUGGCUA	:: ::: ::::.:::::::
Bra-miR169g	A03p047750.1_BraROA	AT2G05760_A03	4	AUACAAAUCAUUCUUGGCUA	: ::: ::::.:::::::
Bra-miR169i	A09p012140.1_BraROA	AT1G72590_A09	3.5	GGAUGAG-CAUCCUUGGCUAC	.: ..:: ::::::::::::
Bra-miR169i	A09p012140.1_BraROA	AT2G16530_A09	3.5	GGAUGAG-CAUCCUUGGCUAC	.: ..:: ::::::::::::
Bra-miR169q	A03p025620.1_BraROA	RAB6A_A03	3	CGCCAAUGCAUCUUUGGCUCA	:: ::: ::::::::::::
Bra-miR169q	A03p050500.1_BraROA	ATG3_A03	3.5	CGGCAACUUAUCUUGUGGCUCA	:::::: :.::::: :::::::
Bra-miR169q	A03p062520.1_BraROA	DGK6_A03	3.5	CUGCGGAGUCAUUUUUGGCUCA	: ::. ::::::.::::::::
Bra-miR169q	A02p007040.1_BraROA	PAM1_A02	4	GAGCAAGU-AUCUUUGGCUCA	:::::: ::::::::::::
Bra-miR172a	A10p023550.1_BraROA	TOE2_A10	1	AUGCAGCAUCAUCAGGAUUCU	::::::::::::::.::::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR172a	A02p012040.1_BraROA	TOE2_A02	1	AUGCAGCAUCAUCAGGAUUCU	::::::::::::::.::::::
Bra-miR172a	A03p071180.1_BraROA	AP2_A03	2	CUGCAGCAUCAUCAGGAUUCU	::::::::::::::.::::::
Bra-miR172a	A07p022340.1_BraROA	TOE3_A07	2.5	AUGGCGGCAUCAUCAGGAUUCU	::: :.:::::::::.::::::
Bra-miR172a	A05g500900.1_BraROA	RCA_A05	2.5	UUGUAACAUCAUCAAGAUUCU	::.: ::::::::::::::
Bra-miR172a	A09p060690.1_BraROA	SMZ_A09	3	UUGCAGCAUCAUCAGGAUUCC	::::::::::::.:::::
Bra-miR172a	A03p027400.1_BraROA	RAP2.7_A03	3	CAGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::
Bra-miR172a	A07p024870.1_BraROA	RAP2.7_A07	3	CAGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::
Bra-miR172a	A02p053020.1_BraROA	CYP706A3_A02	3.5	AUUCAAGUCUCAUCAAGAUUCU	:: :: :. ::::::::::::
Bra-miR172a	A03p033760.1_BraROA	SK32_A03	3.5	GUGGAG-AUUAUCAAGAUUCU	.:: :: ::.::::::::::
Bra-miR172a	A04p006670.1_BraROA	emb1967_A04	3.5	AUGCUGU-UCAUCAAGAUUCC	:::: :. ::::::::::::
Bra-miR172a	A04p029170.1_BraROA	SKdZeta_A04	3.5	GUGGAG-AUCAUCAAGGUUCU	.:: :: :::::::::.::::
Bra-miR172a	A05p016790.1_BraROA	GSK1_A05	3.5	GUGGAG-AUCAUCAAGGUUCU	.:: :: :::::::::.::::
Bra-miR172a	A09p000580.1_BraROA	SK32_A09	3.5	GUGGAG-AUUAUCAAGAUUCU	.:: :: ::.::::::::::
Bra-miR172a	A01p008270.1_BraROA	AT4G30130_A01	4	CUUCCUCAUCAUCAAGAUUCU	: : ::::::::::::::
Bra-miR172a	A01p026970.1_BraROA	CKA2_A01	4	AAGU-GCGUUAUCAAGAUUCU	: :. ::.:.::::::::::
Bra-miR172a	A03p011980.1_BraROA	EXO70H7_A03	4	CUACAG-AUCAUCAAGAUUCG	: ::: ::::::::::::
Bra-miR172a	A03p015630.1_BraROA	CB5-E_A03	4	ACGAUUCAUCAUCAAGAUUCU	: : ::::::::::::::
Bra-miR172a	A06p033260.1_BraROA	HSD3_A06	4	AAGCAGCCUUAUUAAGAUUCU	: ::::: :.::.:::::::::
Bra-miR172a	A09p025990.1_BraROA	AT5G46160_A09	4	AAGAAGCAUC-UCAAGAUUCU	: : :::::: ::::::::::
Bra-miR172c	A03p071180.1_BraROA	AP2_A03	1	CUGCAGCAUCAUCAGGAUUCU	::::::::::::::.::::::
Bra-miR172c	A03p027400.1_BraROA	RAP2.7_A03	2	CAGCAGCAUCAUCAGGAUUCU	: ::::::::::::.::::::
Bra-miR172c	A07p024870.1_BraROA	RAP2.7_A07	2	CAGCAGCAUCAUCAGGAUUCU	: ::::::::::::.::::::
Bra-miR172c	A10p023550.1_BraROA	TOE2_A10	2	AUGCAGCAUCAUCAGGAUUCU	::::::::::::::.::::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR172c	A02p012040.1_BraROA	TOE2_A02	2	AUGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::
Bra-miR172c	A05g500900.1_BraROA	RCA_A05	2	UUGUAACAUCAUCAAGAUUCU	.::.: ::::::::::::::
Bra-miR172c	A09p060690.1_BraROA	SMZ_A09	2.5	UUGCAGCAUCAUCAGGAUUCC	.::::::::::::.::::::
Bra-miR172c	A07p022340.1_BraROA	TOE3_A07	3	UGGCGGCAUCAUCAGGAUUCU	. ::.:::::::::.::::::
Bra-miR172c	A01p008270.1_BraROA	AT4G30130_A01	3	CUUCCUCAUCAUCAAGAUUCU	:: : ::::::::::::::
Bra-miR172c	A03p011980.1_BraROA	EXO70H7_A03	3	CUACAG-AUCAUCAAGAUUCG	:: ::: ::::::::::::
Bra-miR172c	A01p044280.1_BraROA	CRK10_A01	3.5	CUGUACCUUCAUCAAGAUUCA	::::.: : ::::::::::::
Bra-miR172c	A02p025330.1_BraROA	TFIIIA_A02	4	CUGCAGAAGCAUCAGGAUUCU	:::::: : :::::.::::::
Bra-miR172c	A02p031310.1_BraROA	GAPCP-1_A02	4	CUUCA-CAUCGUCAAGAUUUU	:: :: ::::.::::::::.:
Bra-miR172c	A02p053020.1_BraROA	CYP706A3_A02	4	UUCAAGUCUCAUCAAGAUUCU	.: ::. ::::::::::::
Bra-miR172c	A03p033760.1_BraROA	SK32_A03	4	GUGGAG-AUUAUCAAGAUUCU	:: :: ::.::::::::::
Bra-miR172c	A04p029170.1_BraROA	SKdZeta_A04	4	GUGGAG-AUCAUCAAGGUUCU	:: :: :::::::::.::::
Bra-miR172c	A05p016790.1_BraROA	GSK1_A05	4	GUGGAG-AUCAUCAAGGUUCU	:: :: :::::::::.::::
Bra-miR172c	A06p009240.1_BraROA	AT1G11410_A06	4	UUGUACCUUCAUCAAGAUUCA	.::.: : ::::::::::::
Bra-miR172c	A08p043680.1_BraROA	AT1G08800_A08	4	CUUCAACAUCAUCA-GAUUCU	:: :: :::::::: ::::::
Bra-miR172c	A09p000580.1_BraROA	SK32_A09	4	GUGGAG-AUUAUCAAGAUUCU	:: :: ::.::::::::::
Bra-miR172d	A10p023550.1_BraROA	TOE2_A10	1.5	AUGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::.
Bra-miR172d	A02p012040.1_BraROA	TOE2_A02	1.5	AUGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::.
Bra-miR172d	A09p060690.1_BraROA	SMZ_A09	2	UUGCAGCAUCAUCAGGAUUCC	::::::::::::.::::::
Bra-miR172d	A03p071180.1_BraROA	AP2_A03	2.5	CUGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::.
Bra-miR172d	A04p006670.1_BraROA	emb1967_A04	2.5	AUGCUGU-UCAUCAAGAUUCC	:::: :. ::::::::::::
Bra-miR172d	A07p022340.1_BraROA	TOE3_A07	3	AUGGCGGCAUCAUCAGGAUUCU	::: :.:::::::::.::::::.
Bra-miR172d	A05g500900.1_BraROA	RCA_A05	3	UUGUAACAUCAUCAAGAUUCU	::.: ::::::::::::::.

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR172d	A03p027400.1_BraROA	RAP2.7_A03	3.5	CAGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::.
Bra-miR172d	A07p024870.1_BraROA	RAP2.7_A07	3.5	CAGCAGCAUCAUCAGGAUUCU	::::::::::::.::::::.
Bra-miR172d	A09p055230.1_BraROA	AT3G49970_A09	3.5	AU-CAGCGUCAUCAAGAUCCC	:: ::::.:::::::::: ::
Bra-miR172d	A02p025920.1_BraROA	AT1G72700_A02	4	AUGGAGGAUC-UCAAGAUUCC	::: :: :: ::::::::::
Bra-miR172d	A02p053020.1_BraROA	CYP706A3_A02	4	AUUCAAGUCUCAUCAAGAUUCU	:: :: :. :::::::::::.
Bra-miR172d	A03p010160.1_BraROA	AT5G20090_A03	4	AUGGCAACAGCAUCAAGAUUCC	::: :: :: ::::::::::::
Bra-miR172d	A03p011980.1_BraROA	EXO70H7_A03	4	CUACAG-AUCAUCAAGAUUCG	: ::: ::::::::::::
Bra-miR172d	A03p033760.1_BraROA	SK32_A03	4	GUGGAG-AUUAUCAAGAUUCU	.:: :: ::.:::::::::.
Bra-miR172d	A03p050520.1_BraROA	AT5G61540_A03	4	AUGGAGCAACAUUCAAGAUUCC	::: :::: ::: :::::::::
Bra-miR172d	A04p019170.1_BraROA	AT5G41130_A04	4	UUGCAACUAUCAUCAAGAUUUC	:::: : ::::::::::::.:
Bra-miR172d	A04p029170.1_BraROA	SKdZeta_A04	4	GUGGAG-AUCAUCAAGGUUCU	.:: :: :::::::::.:::.
Bra-miR172d	A05p016790.1_BraROA	GSK1_A05	4	GUGGAG-AUCAUCAAGGUUCU	.:: :: :::::::::.:::.
Bra-miR172d	A09p000580.1_BraROA	SK32_A09	4	GUGGAG-AUUAUCAAGAUUCU	.:: :: ::.:::::::::.
Bra-miR319c	A01p010940.1_BraROA	TCP2_A01	2.5	AGGGGGA-CCCUUCAGUCCAA	:::::: ::::::::::::
Bra-miR319c	A01p010940.1_BraROA	TCP2_A01	2.5	AGGGGGA-CCCUUCAGUCCAA	:::::: ::::::::::::
Bra-miR319c	A01p050870.1_BraROA	TCP4_A01	3	AGAGGGGUCCCCUUCAGUCCAG	:: ::.:: ::::::::::::.
Bra-miR319c	A01p050870.1_BraROA	TCP4_A01	3	AGAGGGGUCCCCUUCAGUCCAG	:: ::.:: ::::::::::::.
Bra-miR319c	A03p041320.1_BraROA	TCP4_A03	3	AGAGGGGUCCCCUUCAGUCCAG	:: ::.:: ::::::::::::.
Bra-miR319c	A03p041320.1_BraROA	TCP4_A03	3	AGAGGGGUCCCCUUCAGUCCAG	:: ::.:: ::::::::::::.
Bra-miR319c	A05p016730.1_BraROA	TCP10_A05	3	ACGGGGUACCCUUCAGUCCAG	: ::.:: ::::::::::::.
Bra-miR319c	A05p016730.1_BraROA	TCP10_A05	3	ACGGGGUACCCUUCAGUCCAG	: ::.:: ::::::::::::.
Bra-miR319c	A05p044300.1_BraROA	TCP4_A05	3	AGAGGGGUCCCCUUCAGUCCAG	:: ::.:: ::::::::::::.
Bra-miR319c	A05p044300.1_BraROA	TCP4_A05	3	AGAGGGGUCCCCUUCAGUCCAG	:: ::.:: ::::::::::::.

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR319c	A08p002160.1_BraROA	TCP3_A08	3.5	AGAGGGGUCCCCUUCAGUCCAU	:: ::.:: ::::::::::::
Bra-miR319c	A08p002160.1_BraROA	TCP3_A08	3.5	AGAGGGGUCCCCUUCAGUCCAU	:: ::.:: ::::::::::::
Bra-miR319c	A04p027800.1_BraROA	BGLU14_A04	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A04p027800.1_BraROA	BGLU14_A04	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A04p027800.1_BraROA	BGLU12_A04	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A04p027800.1_BraROA	BGLU12_A04	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A04p027800.1_BraROA	BGLU13_A04	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A04p027800.1_BraROA	BGLU13_A04	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A05p001950.1_BraROA	ASY3_A05	4	AGGAAGUUCCCUUGAGUUCAA	::: ::::::::: :::.:::
Bra-miR319c	A05p001950.1_BraROA	ASY3_A05	4	AGGAAGUUCCCUUGAGUUCAA	::: ::::::::: :::.:::
Bra-miR319c	A05g500560.1_BraROA	BGLU14_A05	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A05g500560.1_BraROA	BGLU14_A05	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A05g500560.1_BraROA	BGLU12_A05	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A05g500560.1_BraROA	BGLU12_A05	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A05g500560.1_BraROA	BGLU13_A05	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR319c	A05g500560.1_BraROA	BGLU13_A05	4	AAGGAGUUCCCAUCGGUCCAA	: ::::::::: ::.:::::::
Bra-miR390a	A02p058880.1_BraROA	DiT2.2_A02	3.5	AGCUUUA-CCCUCCUGAGCUU	:: .:: ::::::::::::
Bra-miR390a	A03p037830.1_BraROA	RECA3_A03	3.5	GGUGAUUGUUCCUCCUGAGCUU	::.: .:.:.::::::::::::
Bra-miR390a	A06p011970.1_BraROA	MIOX1_A06	3.5	GACGAGGUCCCUCCUGAGCUU	: :: .::::::::::::::
Bra-miR390a	A08p040520.1_BraROA	AL7_A08	3.5	GACGAGGUCCCUCCUGAGCUU	: :: .::::::::::::::
Bra-miR390a	A06p005640.1_BraROA	AT1G07650_A06	4	GGCAUUGUCCCUCCUGAGUUC	::: .:.:::::::::::.:
Bra-miR390a	A09p082560.1_BraROA	LIN2_A09	4	GGUGUUAUGCCUCCUGAAGCUU	::.:.::: :::::::: ::::
Bra-miR394a	A07p016910.1_BraROA	LCR_A07	1	GGAGGUUGACAGAAUGCCAA	:::::: ::::::::::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR394a	A09p050010.1_BraROA	LCR_A09	1	GGAGGUUGACAGAAUGCCAA	:::::: ::::::::::::
Bra-miR394a	A10p000600.1_BraROA	AT1G01320_A10	4	GGAGGUGGAAAGAAUCCCAA	::::::::: ::::: ::::
Bra-miR394a	A02p007190.1_BraROA	SYN4_A02	4	GGAGGUGGA-AGGAUGCCAU	::::::::: ::.::::::
Bra-miR394a	A05p018660.1_BraROA	HCF109_A05	4	AAAGGAGGACAGAAUGUCAA	::: ::::::::::.:::
Bra-miR394a	A06p038550.1_BraROA	AT5G64550_A06	4	GGAGGUGGGCAGAGAUGUCAA	::::::::.:::: :::.:::
Bra-miR394a	A07p049860.1_BraROA	AT1G77260_A07	4	GGAGAUGGACAGGAUGCUGA	:::: :::::::.:::..:
Bra-miR394a	A09p009530.1_BraROA	AT5G64550_A09	4	GGAGGUGGGCAGAGAUGUCAA	::::::::.:::: :::.:::
Bra-miR394a	A09p065020.1_BraROA	AT3G59570_A09	4	GGAGCAGCGACAGAAUGUCAA	:::: : :::::::::.:::
Bra-miR396a	A08p042870.1_BraROA	AT1G10120_A08	3	AGGAACAAGAAAGCUGUGGAG	:.: ::::::::::::::.
Bra-miR396a	A03p038710.1_BraROA	APEM9_A03	3	GAGUUCAUGGAAGCUGUGGAG	.:::::: :.:::::::::.
Bra-miR396a	A04p039940.1_BraROA	GRF9_A04	3	ACGUUCAAGAAAGCUUGUGGAA	: :::::::::::: ::::::
Bra-miR396a	A08p036810.1_BraROA	RABG3B_A08	3	AAGUACAAGAA-GCUGUGGAA	:::: :::::: :::::::::
Bra-miR396a	A01p002620.1_BraROA	ROT3_A01	3.5	UAGUU-AAGAAAGUUGUGGAG	:::: :::::::.:::::::.
Bra-miR396a	A03p054900.1_BraROA	AT4G18530_A03	3.5	ACGUUGGAAGAAAGCUGUGGAG	: ::: ::::::::::::::.
Bra-miR396a	A04p025740.1_BraROA	ARSK1_A04	3.5	GAAUGCAAAGAAAGCUGUGGAA	.: : ::: ::::::::::::
Bra-miR396a	A05p009490.1_BraROA	CAC3_A05	3.5	AGGUUUCAGGAAGCUGUGGAG	:.:::. ::.:::::::::.
Bra-miR396a	A05p043610.1_BraROA	AT3G15605_A05	3.5	GAGAU-AAGAAAGCUGUGGGA	.:: : ::::::::::::.:
Bra-miR396a	A06p009480.1_BraROA	AT1G11710_A06	3.5	AAGGUGGAGAAAGCUGUGGAU	::: : .::::::::::::
Bra-miR396a	A06p036350.1_BraROA	AT5G62620_A06	3.5	AAGUUGGUAGAAAGCUGUGGGA	::::: . :::::::::::.:
Bra-miR396a	A06p045630.1_BraROA	AT5G27230_A06	3.5	CAGUUCGCGAAAGUUGUGGAA	::::::. :::::.:::::::
Bra-miR396a	A07p040190.1_BraROA	AT1G65780_A07	3.5	GAGUUCAAGAAAUCUGUGGGA	.:::::::::: :::::::.:
Bra-miR396a	A08p026400.1_BraROA	AT4G30790_A08	3.5	AACCUCAAGAAGGCUGUGGAG	:: :::::::.::::::::.
Bra-miR396a	A09p079160.1_BraROA	AT1G10120_A09	4	CGGAACAAGAAAGCUGUGGAG	.: ::::::::::::::.

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR396a	A10p038080.1_BraROA	CHIL_A10	4	GAGCUCGAGAAAGUUGUGGGA	.:: ::.:::::::.::::::.:
Bra-miR396a	A01p001470.1_BraROA	GRF2_A01	4	CCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A01p006020.1_BraROA	AT4G32590_A01	4	GAGGUGGAGGAAGCUGUGGAA	.:: : .::.::::::::::::
Bra-miR396a	A01p016480.1_BraROA	GRF8_A01	4	UCGUUCAAGAAAGCAUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A02p016990.1_BraROA	GRF7_A02	4	UCGUUCAAGAAAGCAUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A03p020710.1_BraROA	GRF3_A03	4	CCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A03p028580.1_BraROA	GRF1_A03	4	UCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A03p059340.1_BraROA	GRF8_A03	4	UCGUUCAAGAAAGCAUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A03p071810.1_BraROA	GRF2_A03	4	CCGUUCAAGAAAGCUUGUGGAA	:::::::::::: ::::::
Bra-miR396a	A04p008130.1_BraROA	GRF4_A04	4	CCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A04p019640.1_BraROA	AT5G41620_A04	4	GAGCUCAAGAAACCUGUGGAG	.:: :::::::: :::::::
Bra-miR396a	A04p033540.1_BraROA	GRF3_A04	4	CCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A05p008470.1_BraROA	JAL22_A05	4	GAGUACGAGAAAGCUG-GGAA	.::: :.::::::::: ::::
Bra-miR396a	A05p008470.1_BraROA	JAL23_A05	4	GAGUACGAGAAAGCUG-GGAA	.::: :.::::::::: ::::
Bra-miR396a	A05p011040.1_BraROA	GRF3_A05	4	CCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A05p042930.1_BraROA	PAB6_A05	4	GAAUGCAAAGAAAGCUGUGGAG	.: : ::: ::::::::::::.
Bra-miR396a	A06p007340.1_BraROA	AT1G09600_A06	4	UGGUUGAGGGAAGCUGUGGAA	.::: :.:.::::::::::::
Bra-miR396a	A06p002830.1_BraROA	EXO70G2_A06	4	AUGUUAAAGAAAGCUGGGGAA	: ::: ::::::::: ::::
Bra-miR396a	A06p038170.1_BraROA	SC35_A06	4	AAGCUCAC-AAAGCUGUGGAA	::: ::: ::::::::::::
Bra-miR396a	A06p046010.1_BraROA	AT5G28740_A06	4	GAGUCCACGAGAGCUGUGGAG	.::: :: ::.:::::::::.
Bra-miR396a	A07p027320.1_BraROA	GRF4_A07	4	CCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR396a	A09p000880.1_BraROA	AT4G01030_A09	4	AAGUGGGAGAAGGCUGUGGAG	:::: .::::.::::::::.
Bra-miR396a	A09p021390.1_BraROA	HEMA1_A09	4	AGGUUGAGGAAAGCUAUGGAA	:.::: :.:::::::: :::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR396a	A09p057270.1_BraROA	BAG4_A09	4	AAGUUGCAGGAAGCUGUGGAU	::::: ::.::::::::::
Bra-miR396a	A09p058490.1_BraROA	GRF4_A09	4	CCGUUCAAGAAAGCCUGUGGAA	:::::::::::: :::::::
Bra-miR397	A02p012100.1_BraROA	LAC17_A02	2.5	AUAAUCAACGCUGCACUUAAUGA	:.: ::::::::::::.:::::
Bra-miR397	A02p000890.1_BraROA	LAC11_A02	3	UCAUCAACGCUGCACUAAAUGA	:::::::::::::: :::::
Bra-miR397	A04p034830.1_BraROA	IRX12_A04	3	AUAGUCAAUGCUGCACUUAAUGA	:.: ::::.::::::::.:::::
Bra-miR397	A05p018330.1_BraROA	LAC2_A05	3.5	UGAUCAAUGCUGCACUCGAUGA	:::::.::::::::.:::::
Bra-miR397	A05p009480.1_BraROA	IRX12_A05	4	UAGUCAAUGCUGCACUUAAUGA	.:::::.::::::::.:::::
Bra-miR398a	A01p054070.1_BraROA	DREB2B_A01	4	AAGGAG-GACCUGAGAACACU	::: : ::::::::::::
Bra-miR399a	A10g505860.1_BraROA	UGT76E2_A10	4	UAGAG-AGCUCUUCUUUGGCA	.:: : :.::::.::::::::
Bra-miR399a	A10g505860.1_BraROA	UGT76E1_A10	4	UAGAG-AGCUCUUCUUUGGCA	.:: : :.::::.::::::::
Bra-miR399a	A04p023120.1_BraROA	AT3G29255_A04	4	CAGUGCAGAUCUCCUUUGGCG	::: :::. :::::::::::.
Bra-miR399a	A04p023120.1_BraROA	PEN3_A04	4	CAGUGCAGAUCUCCUUUGGCG	::: :::. :::::::::::.
Bra-miR399e	A04p023120.1_BraROA	AT3G29255_A04	3	CAGUGCAGAUCUCCUUUGGCG	: : :::.::::::::::::.
Bra-miR399e	A04p023120.1_BraROA	PEN3_A04	3	CAGUGCAGAUCUCCUUUGGCG	: : :::.::::::::::::.
Bra-miR399e	A01p024030.1_BraROA	UGT84A3_A01	4	UCGUUUAUAUCUCCUUUGGCA	.:: .: ::::::::::::
Bra-miR399e	A09p078200.1_BraROA	AT1G11800_A09	4	ACGGG-AAGUUUCCUUUGGCA	:::: ::.:.::::::::::
Bra-miR408	A02p047610.1_BraROA	ARPN_A02	1.5	GGCCAAGGGAAGAGGCAGUGCA	.:::: :::::::::::::::
Bra-miR408	A02p010530.1_BraROA	AT5G21100_A02	3	ACGAAGGGAAGAGGCAGUGCA	: :::::::::::::::::
Bra-miR408	A02p026900.1_BraROA	AT1G73930_A02	3.5	AAUCAGG-AGGAGGCAGUGCA	: .:::: :.::::::::::
Bra-miR408	A10p030150.1_BraROA	LNG1_A10	4	AGCCAGGGAAGAG-CAGUGUC	:::::::::::: :::::.
Bra-miR408	A01p059390.1_BraROA	RPK2_A01	4	AGCUAAGGAAGAGGCUGUGCG	:::.: :::::::: :::::.
Bra-miR408	A03p004860.1_BraROA	AT5G10710_A03	4	CACCAGCUGAAGAGGCAGUGCA	:::: ::::::::::::::
Bra-miR408	A04p028430.1_BraROA	LAC3_A04	4	AACCAGUGAAGAGGCUGUGCA	: :::: :::::::: :::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miR408	A05p017350.1_BraROA	LAC3_A05	4	AACCAGUGAAGAGGCUGUGCA	: :::: :::::::: :::::
Bra-miRN347	A07p012850.1_BraROA	IPT7_A07	0	GAGAUCAUAAACUCCGACAA	::::::::::::::::::::
Bra-miRN347	A04p000350.1_BraROA	IPT3_A04	1	GAGAUCAUAAACUCCGAUAA	::::::::::::::::.::
Bra-miRN347	A10p027050.1_BraROA	IPT5_A10	1.5	GAGAUCGUAAACUCUGACAA	::::::.::::::::.:::::
Bra-miRN347	A02p009000.1_BraROA	IPT5_A02	1.5	GAGAUCGUAAAUUCCGACAA	::::::.:::::.::::::::
Bra-miRN347	A01p047010.1_BraROA	IPT8_A01	3	GAGAUCAUCAACUCUGACAA	:::::::: :::::.:::::
Bra-miRN347	A03p043720.1_BraROA	IPT8_A03	3	GAGAUCAUCAACUCUGACAA	:::::::: :::::.:::::
Bra-miRN347	A07p039650.1_BraROA	REV3_A07	3.5	GAGGACAUAAACUCCGAUGA	:::. ::::::::::::..:
Bra-miRN347	A01p002610.1_BraROA	AT4G36390_A01	4	AAGAUCCUAGAUUCCGACAA	::::: ::.:.::::::::
Bra-miRN347	A02p036960.1_BraROA	AT4G10060_A02	4	GAGAUCAUGAGCUCCGGCGA	::::::::.:.:::::.:.:
Bra-miRN347	A03p054920.1_BraROA	AT4G18540_A03	4	GCGA-CUUAGACUCCGACAA	: :: : ::.::::::::::
Bra-miRN347	A04p026570.1_BraROA	IPT2_A04	4	GAGAUCAUAAACGCCGACGC	:::::::::::: :::::
Bra-miRN347	A09p058700.1_BraROA	UGT73D1_A09	4	GAGUUGCAUAAUCUCCGACAA	::: : ::::: ::::::::
Bra-miRN367	A07p012850.1_BraROA	IPT7_A07	0	GAUCAUAAACUCCGACAAGAU	::::::::::::::::::::
Bra-miRN367	A04p000350.1_BraROA	IPT3_A04	1	GAUCAUAAACUCCGAUAAGAU	::::::::::::::.:::::
Bra-miRN367	A10p027050.1_BraROA	IPT5_A10	1.5	GAUCGUAAACUCUGACAAGAU	::::.:::::::.::::::::
Bra-miRN367	A02p009000.1_BraROA	IPT5_A02	1.5	GAUCGUAAAUUCCGACAAGAU	::::.:::::.::::::::::
Bra-miRN367	A01p059700.1_BraROA	VDAC1_A01	3	GACCA-AAACUCCGGCAAGAU	:: :: ::::::::.::::::
Bra-miRN367	A03p034350.1_BraROA	VDAC1_A03	3	GACCA-AAACUCCGGCAAGAU	:: :: ::::::::.::::::
Bra-miRN367	A06p009640.1_BraROA	AHP2_A06	3.5	GAU-AUGUACUCUGACAAGAU	::: ::. ::::.::::::::
Bra-miRN367	A06p013830.1_BraROA	CYP79F1_A06	3.5	CGUCAUAAACUCUGACGAGAU	.:::::::::.:::.:::::
Bra-miRN367	A06p013830.1_BraROA	CYP79F2_A06	3.5	CGUCAUAAACUCUGACGAGAU	.:::::::::.:::.:::::
Bra-miRN367	A01p020010.1_BraROA	WES1_A01	4	GAGUAUAGAUUCCGACAAGAC	:: .:::.:.::::::::::

MiRNA	B. rapa ID	Ath orth.	Score	Target binding site	Basepairs
Bra-miRN367	A01p047010.1_BraROA	IPT8_A01	4	GAUCAUCAACUCUGACAAAAU	:::::: ::::.::::: ::
Bra-miRN367	A03p043720.1_BraROA	IPT8_A03	4	GAUCAUCAACUCUGACAAAAU	:::::: ::::.::::: ::
Bra-miRN367	A05p055140.1_BraROA	VDAC1_A05	4	GACCA-AAACUCUGGCAAGAU	:: :: ::::::.:.::::::
Bra-miRN367	A08p003610.1_BraROA	DABB1_A08	4	GA-CGCUGACUCCGACAAGAU	:: :. .::::::::::::
Bra-miRN367	A08p045990.1_BraROA	POR C_A08	4	AAUCUUAAA-UCCGACAAGAU	::: :::: ::::::::::
Bra-miRN367	A09p008250.1_BraROA	CXE20_A09	4	GAUCAGAAACUCCGACGAGGA	::::: ::::::::::.::.

A.3 APPENDIX CHAPTER 4

Supplementary Figures and Tables for Chapter 4

Figure A.1 (continued)

(Top panel) Expression pattern of Bra-miR156 members during development. Each point represents the expression level in each apex replicate per timepoint, while the line is the average expression in each genotype. Vertical lines represent the first day of the floral transition in Sarisha-14 (day 12, blue line) and R-o-18 (day 19, red line). Standard deviation is represented in the shaded region.

(Bottom panel) Expression profiles of all expressed Bra-miR156 target genes predicted by TargetFinder. Only genes which were expressed above 1TMMC in at least one genotype for more than 8 timepoints are displayed.

Figure A.2 (continued)

(Top panel) Expression pattern of Bra-miR172 members during development. Each point represents the expression level in each apex replicate per timepoint, while the line is the average expression in each genotype. Vertical lines represent the first day of the floral transition in Sarisha-14 (day 12, blue line) and R-o-18 (day 19, red line). Standard deviation is represented in the shaded region.

(Bottom panel) Expression profiles of all expressed Bra-miR172 target genes predicted by TargetFinder. Only genes which were expressed above 1TMMC in at least one genotype for more than 8 timepoints are displayed.

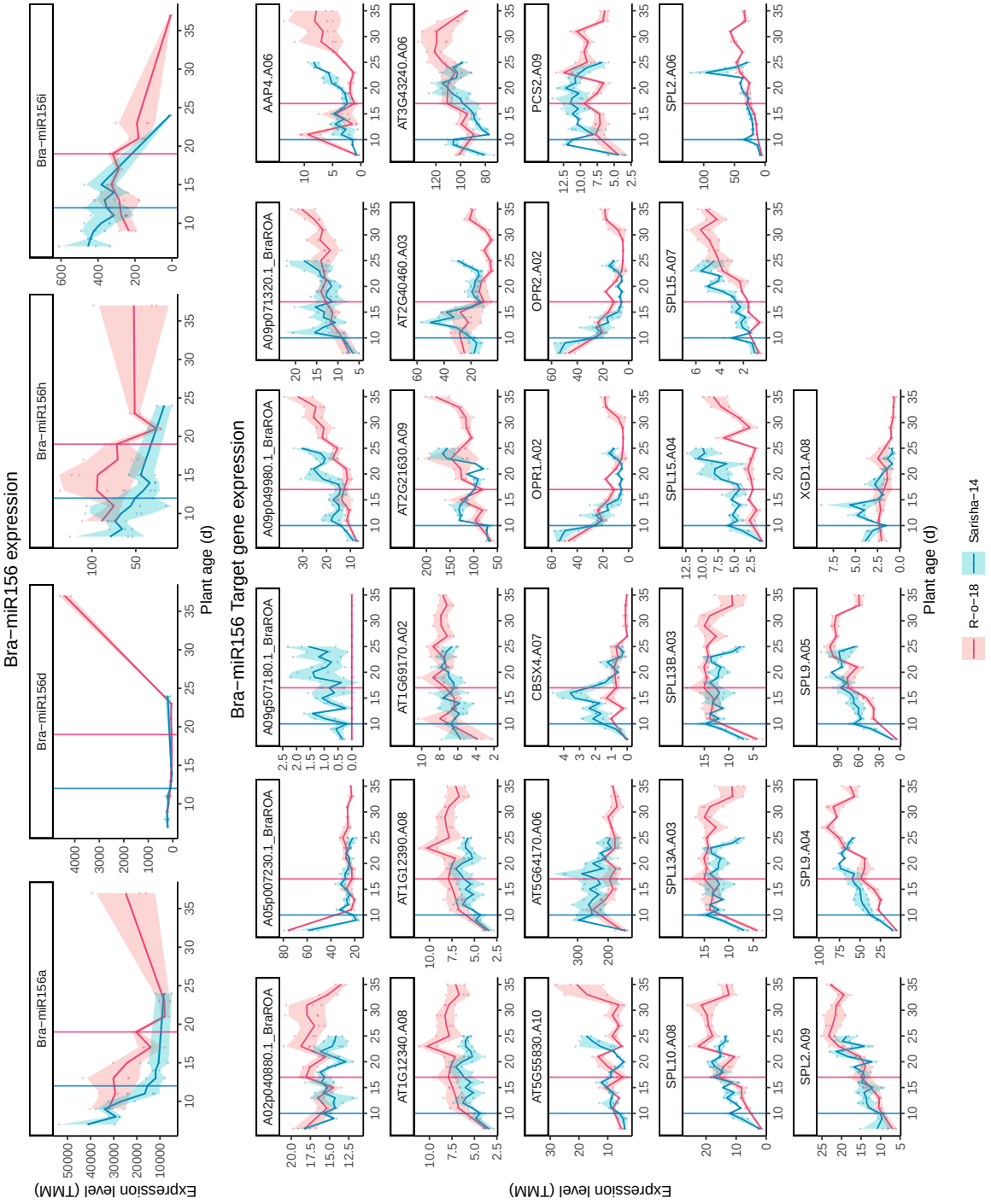


Figure A.1: **Expression pattern of Bra-miR156 predicted target genes.** (caption continued in above page)

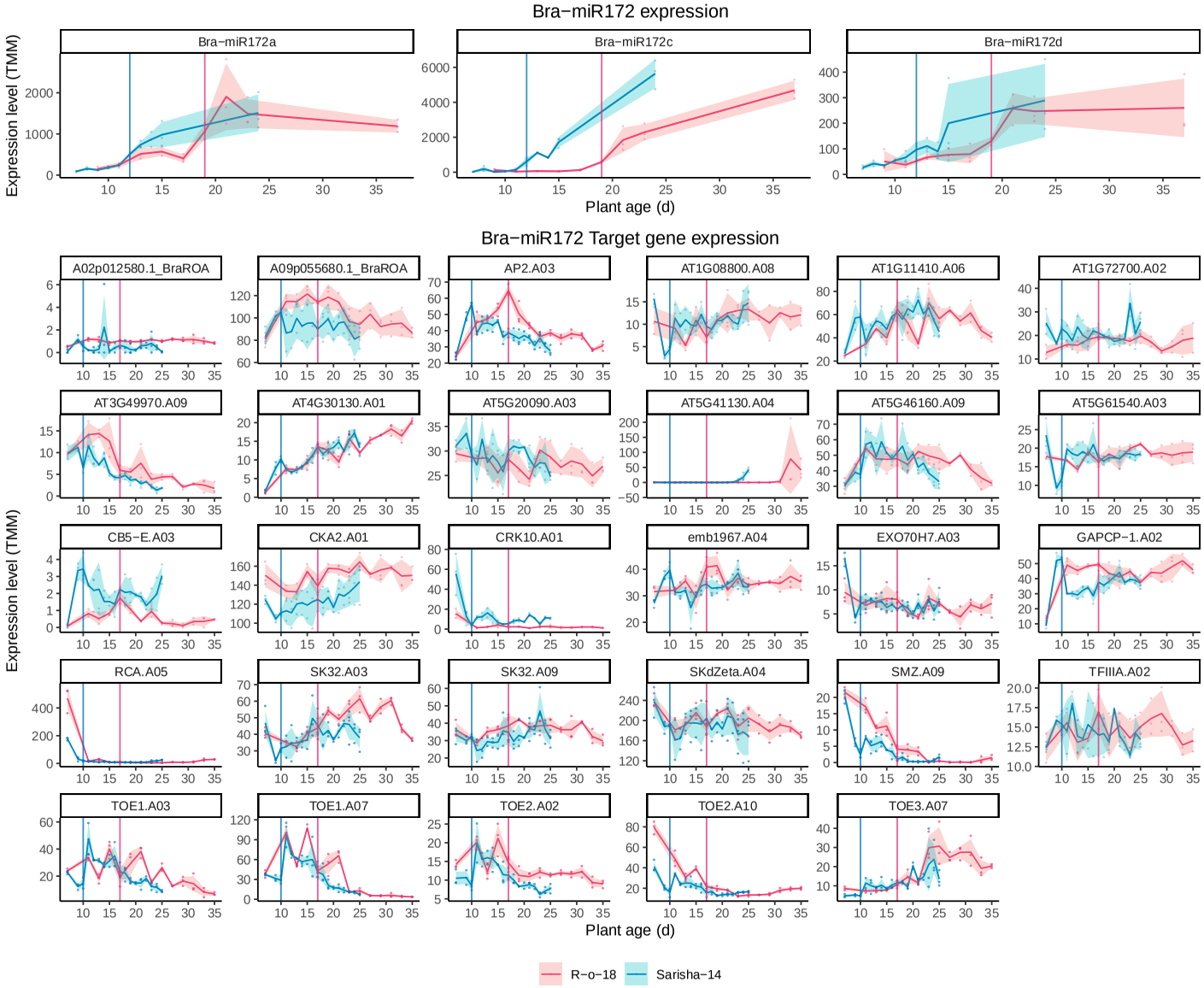


Figure A.2: **Expression pattern of Bra-miR172 predicted target genes.** (caption continued in above page)

REFERENCES

Ahmed, Waqas et al. (Mar. 2020). "Comparative Analysis of miRNA Expression Profiles between Heat-Tolerant and Heat-Sensitive Genotypes of Flowering Chinese Cabbage Under Heat Stress Using High-Throughput Sequencing." en. In: *Genes* 11.3. Publisher: Multidisciplinary Digital Publishing Institute, p. 264. ISSN: 2073-4425. DOI: 10.3390/genes11030264. URL: https://www.mdpi.com/2073-4425/11/3/264 (visited on 09/28/2025).

Aparicio-Puerta, Ernesto et al. (July 2022). "sRNAbench and sRNAtoolbox 2022 update: accurate miRNA and sncRNA profiling for model and non-model organisms." In: *Nucleic Acids Research* 50.W1, W710–W717. ISSN: 0305-1048. DOI: 10.1093/nar/gkac363. URL: https://doi.org/10.1093/nar/gkac363 (visited on 08/27/2025).

Aukerman, Milo J. and Hajime Sakai (Nov. 2003). "Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes." In: *The Plant Cell* 15.11, pp. 2730–2741. ISSN: 1040-4651. DOI: 10.1105/tpc.016238. URL: https://doi.org/10.1105/tpc.016238 (visited on 07/23/2025).

Bagheri, H. et al. (2013). "Identification of seed-related QTL in Brassica rapa." eng. In: *Spanish journal of agricultural research* 11.4. Publisher: Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) Section: Spanish journal of agricultural research, pp. 1085–1093. ISSN: 2171-9292, 1695-971X. URL: https://dialnet.unirioja.es/servlet/articulo?codigo=4861430 (visited on 02/14/2026).

Bangladesh Agricultural Research Institute (2006). *BARI Sarisa-14*. https://dhcrop.bsmrau.net/bari-sarisa-14/. Digital Herbarium of Crop Plants, Bangabandhu Sheikh Mujibur Rahman Agricultural University. Accessed: 2025-09-24.

Ben-Naim, Orna et al. (2006). "The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA." en. In: *The Plant Journal* 46.3. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-313X.2006.02706.x, pp. 462–476. ISSN: 1365-313X. DOI: 10.1111/j.1365-313X.2006.02706.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2006.02706.x (visited on 08/07/2025).

Ben-Nissan, Gili et al. (Dec. 2008). "Arabidopsis Casein Kinase 1-Like 6 Contains a Microtubule-Binding Domain and Affects the Organization of Cortical Microtubules." In: *Plant Physiology* 148.4, pp. 1897–1907. ISSN: 0032-0889. DOI: 10.1104/pp.108.129346. URL: https://doi.org/10.1104/pp.108.129346 (visited on 08/01/2025).

Bernardi, Yanel et al. (June 2022). "MicroRNA miR394 regulates flowering time in Arabidopsis thaliana." en. In: *Plant Cell Reports* 41.6, pp. 1375–1388. ISSN: 1432-203X.

DOI: 10.1007/s00299-022-02863-0. URL: https://doi.org/10.1007/s00299-022-02863-0 (visited on 09/17/2025).

Blée, Elizabeth et al. (Aug. 2014). "The Reductase Activity of the Arabidopsis Caleosin RESPONSIVE TO DESSICATION20 Mediates Gibberellin-Dependent Flowering Time, Abscisic Acid Sensitivity, and Tolerance to Oxidative Stress." en. In: *Plant Physiology* 166.1, pp. 109–124. ISSN: 1532-2548. DOI: 10.1104/pp.114.245316. URL: https://academic.oup.com/plphys/article/166/1/109/6113196 (visited on 08/11/2025).

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." In: *Bioinformatics* 30.15, pp. 2114–2120.

Bonfield, James K et al. (2021). "HTSlib: C library for reading/writing high-throughput sequencing data." In: *Gigascience* 10.2, giab007.

Brown, James K. M., Rebecca Beeby, and Steven Penfield (May 2019). "Yield instability of winter oilseed rape modulated by early winter temperature." en. In: *Scientific Reports* 9.1. Publisher: Nature Publishing Group, p. 6953. ISSN: 2045-2322. DOI: 10.1038/s41598-019-43461-7. URL: https://www.nature.com/articles/s41598-019-43461-7 (visited on 02/21/2026).

Bustos-Sanmamed, Pilar et al. (Oct. 2013). "Overexpression of miR160 affects root growth and nitrogen-fixing nodule number in Medicago truncatula." In: *Functional Plant Biology* 40.12, pp. 1208–1220. ISSN: 1445-4408. DOI: 10.1071/FP13123. URL: https://doi.org/10.1071/FP13123 (visited on 02/24/2026).

Cai, Xiaoning et al. (Sept. 2007). "A Putative CCAAT-Binding Transcription Factor Is a Regulator of Flowering Timing in Arabidopsis." In: *Plant Physiology* 145.1, pp. 98–105. ISSN: 0032-0889. DOI: 10.1104/pp.107.102079. URL: https://doi.org/10.1104/pp.107.102079 (visited on 08/08/2025).

Calderwood, Alexander, Jo Hepworth, et al. (Jan. 2021). "Comparative transcriptomics reveals desynchronisation of gene expression during the floral transition between Arabidopsis and Brassica rapa cultivars." en. In: *Quantitative Plant Biology* 2, e4. ISSN: 2632-8828. DOI: 10.1017/qpb.2021.6. URL: https://www.cambridge.org/core/journals/quantitative-plant-biology/article/comparative-transcriptomics-reveals-desynchronisation-of-gene-expression-during-the-floral-transition-between-arabidopsis-and-brassica-rapa-cultivars (visited on 07/24/2025).

Calderwood, Alexander, Andrew Lloyd, et al. (2021). "Total FLC transcript dynamics from divergent paralogue expression explains flowering diversity in Brassica napus." In: *New Phytologist* 229.6, pp. 3534–3548.

carringtonlab/TargetFinder (Dec. 2024). original-date: 2014-02-05T23:56:30Z. URL: https://github.com/carringtonlab/TargetFinder (visited on 07/24/2025).

Cheng, Hui et al. (Mar. 2004). "Gibberellin regulates Arabidopsis floral development via suppression of DELLA protein function." In: *Development* 131.5, pp. 1055–1064. ISSN: 0950-1991. DOI: 10.1242/dev.00992. URL: https://doi.org/10.1242/dev.00992 (visited on 09/25/2025).

Chuck, George et al. (Dec. 2007). "The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1." en. In: *Nature Genetics* 39.12. Publisher: Nature Publishing Group, pp. 1517–1521. ISSN: 1546-1718. DOI: 10.1038/ng.2007.20. URL: https://www.nature.com/articles/ng.2007.20 (visited on 02/21/2026).

Cingolani, P. et al. (2012). "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3." In: *Fly* 6.2, pp. 80–92.

Craft, Judith et al. (2005). "New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis." en. In: *The Plant Journal* 41.6. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-313X.2005.02342.x, pp. 899–918. ISSN: 1365-313X. DOI: 10.1111/j.1365-313X.2005.02342.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2005.02342.x (visited on 02/23/2026).

Danecek, Petr et al. (2021). "Twelve years of SAMtools and BCFtools." In: *Gigascience* 10.2, giab008.

Debernardi, Juan M., Ramiro E. Rodriguez, et al. (Jan. 2012). "Functional Specialization of the Plant miR396 Regulatory Network through Distinct MicroRNA–Target Interactions." en. In: *PLOS Genetics* 8.1. Publisher: Public Library of Science, e1002419. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1002419. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002419 (visited on 02/21/2026).

Debernardi, Juan M., Daniel P. Woods, et al. (Apr. 2022). "MiR172-APETALA2-like genes integrate vernalization and plant age to control flowering time in wheat." In: *PLOS Genetics* 18.4, e1010157. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1010157. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9037917/ (visited on 09/23/2025).

Dong, Qingkun, Binbin Hu, and Cui Zhang (Feb. 2022). "microRNAs and Their Roles in Plant Development." English. In: *Frontiers in Plant Science* 13. Publisher: Frontiers. ISSN: 1664-462X. DOI: 10.3389/fpls.2022.824240. URL: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.824240/full (visited on 02/20/2026).

Ewels, Philip et al. (Oct. 2016). "MultiQC: summarize analysis results for multiple tools and samples in a single report." In: *Bioinformatics* 32.19, pp. 3047–3048. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btw354. URL: https://doi.org/10.1093/bioinformatics/btw354 (visited on 08/27/2025).

Fahlgren, Noah and James C. Carrington (2010). "miRNA Target Prediction in Plants." In: *Plant MicroRNAs: Methods and Protocols*. Ed. by Blake C. Meyers and Pamela J. Green. Totowa, NJ: Humana Press, pp. 51–57. ISBN: 978-1-60327-005-2. DOI: 10.1007/978-1-60327-005-2_4. URL: https://doi.org/10.1007/978-1-60327-005-2_4.

Fahlgren, Noah, Taiowa A. Montgomery, et al. (May 2006). "Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA Affects Developmental Timing and Patterning in Arabidopsis." English. In: *Current Biology* 16.9. Publisher: Elsevier, pp. 939–944. ISSN: 0960-9822. DOI: 10.1016/j.cub.2006.03.065. URL: https://www.cell.com/current-biology/abstract/S0960-9822(06)01397-2 (visited on 09/17/2025).

Fang, Yujie et al. (2021). "Roles of miR319-regulated TCPs in plant development and response to abiotic stress." In: *The Crop Journal* 9.1, pp. 17–28. ISSN: 2214-5141. DOI: https://doi.org/10.1016/j.cj.2020.07.007. URL: https://www.sciencedirect.com/science/article/pii/S2214514120301197.

Feng, Jing et al. (Nov. 2016). "ZRF1 Chromatin Regulators Have Polycomb Silencing and Independent Roles in Development." In: *Plant Physiology* 172.3, pp. 1746–1759. ISSN: 0032-0889. DOI: 10.1104/pp.16.00193. URL: https://doi.org/10.1104/pp.16.00193 (visited on 08/01/2025).

Fischer, David S, Fabian J Theis, and Nir Yosef (Nov. 2018). "Impulse model-based differential expression analysis of time course sequencing data." In: *Nucleic Acids Research* 46.20, e119. ISSN: 0305-1048. DOI: 10.1093/nar/gky675. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC6237758/ (visited on 09/17/2025).

Franco-Zorrilla, José Manuel et al. (Aug. 2007). "Target mimicry provides a new mechanism for regulation of microRNA activity." en. In: *Nature Genetics* 39.8. Publisher: Nature Publishing Group, pp. 1033–1037. ISSN: 1546-1718. DOI: 10.1038/ng2079. URL: https://www.nature.com/articles/ng2079 (visited on 02/23/2026).

Futschik, Matthias E. and Bronwyn Carlisle (Aug. 2005). "NOISE-ROBUST SOFT CLUSTERING OF GENE EXPRESSION TIME-COURSE DATA." en. In: *Journal of Bioinformatics and Computational Biology* 03.04, pp. 965–988. ISSN: 0219-7200, 1757-6334. DOI: 10.1142/S0219720005001375. URL: https://www.worldscientific.com/doi/abs/10.1142/S0219720005001375 (visited on 09/19/2025).

Gandikota, Madhuri et al. (2007). "The miRNA156/157 recognition element in the 3 UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings." en. In: *The Plant Journal* 49.4, pp. 683–693. ISSN: 1365-313X. DOI: 10.1111/j.1365-313X.2006.02983.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2006.02983.x (visited on 07/25/2025).

Gao, Yu et al. (2022). "The evolution and functional roles of miR408 and its targets in plants." In: *International Journal of Molecular Sciences* 23.1, p. 530.

Garg, Vanika and Rajeev K. Varshney (2022). "Analysis of Small RNA Sequencing Data in Plants." en. In: *Plant Bioinformatics: Methods and Protocols*. Ed. by David Edwards. New York, NY: Springer US, pp. 497–509. ISBN: 978-1-07-162067-0. DOI: 10.1007/978-1-0716-2067-0_26. URL: https://doi.org/10.1007/978-1-0716-2067-0_26 (visited on 08/27/2025).

Garrison, Erik and Gabor Marth (2012). *Haplotype-based variant detection from short-read sequencing*. arXiv: 1207.3907 [q-bio.GN]. URL: https://arxiv.org/abs/1207.3907.

German, Marcelo A. et al. (Aug. 2008). "Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends." en. In: *Nature Biotechnology* 26.8. Publisher: Nature Publishing Group, pp. 941–946. ISSN: 1546-1696. DOI: 10.1038/nbt1417. URL: https://www.nature.com/articles/nbt1417 (visited on 01/23/2026).

Giraud, Estelle et al. (2010). "TCP transcription factors link the regulation of genes encoding mitochondrial proteins with the circadian clock in Arabidopsis thaliana." In: *The Plant Cell* 22.12, pp. 3921–3934.

Guo, Zhonglong, Zheng Kuang, and Xiaozeng Yang (2025). "The Usage of PmiREN (Plant miRNA ENcyclopedia)." en. In: *Plant MicroRNAs: Methods and Protocols*. Ed. by Stefan de Folter. New York, NY: Springer US, pp. 83–90. ISBN: 978-1-07-164398-3. DOI: 10.1007/978-1-0716-4398-3_4. URL: https://doi.org/10.1007/978-1-0716-4398-3_4 (visited on 08/27/2025).

Hanano, Shigeru and Koji Goto (Sept. 2011). "*Arabidopsis* TERMINAL FLOWER1 Is Involved in the Regulation of Flowering Time and Inflorescence Development through Transcriptional Repression." en. In: *The Plant Cell* 23.9, pp. 3172–3184. ISSN: 1532-298X, 1040-4651. DOI: 10.1105/tpc.111.088641. URL: https://academic.oup.com/plcell/article/23/9/3172/6097598 (visited on 08/05/2025).

Huang, Shili et al. (2020). "Plant miR397 and its functions." In: *Functional Plant Biology* 48.4, pp. 361–370.

Jeong, Dong-Hoon et al. (July 2013). "Comprehensive Investigation of MicroRNAs Enhanced by Analysis of Sequence Variants, Expression Patterns, ARGONAUTE Loading, and Target Cleavage." In: *Plant Physiology* 162.3, pp. 1225–1245. ISSN: 0032-0889. DOI: 10.1104/pp.113.219873. URL: https://doi.org/10.1104/pp.113.219873 (visited on 01/23/2026).

Jiang, Jianxia et al. (Nov. 2021). "Identification of miRNAs and their target genes in genic male sterility lines in Brassica napus by small RNA sequencing." In: *BMC Plant Biology* 21.1, p. 520. ISSN: 1471-2229. DOI: 10.1186/s12870-021-03306-w. URL: https://doi.org/10.1186/s12870-021-03306-w (visited on 09/28/2025).

Jones, Ashley and Benjamin Schwessinger (2021). *Sorbitol washing complex homogenate for improved DNA extractions — protocols.io*. https://www.protocols.io/view/sorbitol-washing-complex-homogenate-for-improved-d-ewov1875pgr2/v1. [Accessed 02-07-2025]. DOI: 10.17504/protocols.io.beuvjew6.

Jones, Ashley, Cynthia Torkel, et al. (July 2021). "High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing." In: *PLOS ONE* 16.7, pp. 1–6. DOI: 10.1371/journal.pone.0253830. URL: https://doi.org/10.1371/journal.pone.0253830.

Jones-Rhoades, Matthew W and David P Bartel (2004). "Computational identification of plant microRNAs and their targets, including a stress-induced miRNA." In: *Molecular cell* 14.6, pp. 787–799.

Jung, Jae-Hoon et al. (Sept. 2007). "The GIGANTEA-Regulated MicroRNA172 Mediates Photoperiodic Flowering Independent of CONSTANS in Arabidopsis." In: *The Plant Cell* 19.9, pp. 2736–2748. ISSN: 1040-4651. DOI: 10.1105/tpc.107.054528. URL: https://doi.org/10.1105/tpc.107.054528 (visited on 07/23/2025).

Kalvari, Ioanna et al. (2021). "Rfam 14: expanded coverage of metagenomic, viral and microRNA families." In: *Nucleic acids research* 49.D1, pp. D192–D200.

Kardailsky, Igor et al. (Dec. 1999). "Activation Tagging of the Floral Inducer FT." In: *Science* 286.5446. Publisher: American Association for the Advancement of Science, pp. 1962–1965. DOI: 10.1126/science.286.5446.1962. URL: https://www.science.org/doi/10.1126/science.286.5446.1962 (visited on 07/31/2025).

Kim, Bumjin et al. (Nov. 2012). "Identification and profiling of novel microRNAs in the Brassica rapagenome based on small RNA deep sequencing." en. In: *BMC Plant Biology* 12.1, p. 218. ISSN: 1471-2229. DOI: 10.1186/1471-2229-12-218. URL: https://doi.org/10.1186/1471-2229-12-218 (visited on 01/22/2026).

Kim, Daehwan et al. (2019). "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype." In: *Nature biotechnology* 37.8, pp. 907–915.

Kim, Jeong Hoe and Byung Ha Lee (2006). "GROWTH-REGULATING FACTOR4 of Arabidopsis thaliana is required for development of leaves, cotyledons, and shoot apical meristem." In: *Journal of Plant Biology* 49.6, pp. 463–468.

Kim, Wanhui et al. (2011). "The role of the miR399-PHO2 module in the regulation of flowering time in response to different ambient temperatures in Arabidopsis thaliana." In: *Molecules and cells* 32.1, pp. 83–88.

Kinoshita, Atsuko et al. (Dec. 2020). "Regulation of shoot meristem shape by photoperiodic signaling and phytohormones during floral induction of Arabidopsis." In: *eLife* 9. Ed. by Hao Yu, Christian S Hardtke, and Yuling Jiao. Publisher: eLife Sciences Publications, Ltd, e60661. ISSN: 2050-084X. DOI: 10.7554/eLife.60661. URL: https://doi.org/10.7554/eLife.60661 (visited on 09/19/2025).

Kobayashi, Yasushi et al. (1999). "A pair of related genes with antagonistic roles in mediating flowering signals." In: *Science* 286.5446, pp. 1960–1962.

Kolde, Raivo and Maintainer Raivo Kolde (2015). "Package 'pheatmap'." In: *R package* 1.7, p. 790.

Kong, Congcong et al. (Jan. 2022). "Global DNA Methylation and mRNA-miRNA Variations Activated by Heat Shock Boost Early Microspore Embryogenesis in Cabbage (Brassica oleracea)." en. In: *International Journal of Molecular Sciences* 23.9. Publisher: Multidisciplinary Digital Publishing Institute, p. 5147. ISSN: 1422-0067. DOI: 10.3390/ijms23095147. URL: https://www.mdpi.com/1422-0067/23/9/5147 (visited on 09/28/2025).

Kristianingsih, Ruth (2024). *greatR: Gene Registration from Expression and Time-Courses in R*. R package version 2.0.0.9000, https://github.com/ruthkr/greatR/. URL: https://ruthkr.github.io/greatR/.

Kuang, Zheng et al. (2019). "miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants." In: *Bioinformatics* 35.14, pp. 2521–2522.

Kumar, Lokesh and Matthias E Futschik (2007). "Mfuzz: a software package for soft clustering of microarray data." In: *Bioinformation* 2.1, p. 5.

Langmead, Ben et al. (Mar. 2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." In: *Genome Biology* 10.3, R25. ISSN: 1474-760X. DOI: 10.1186/gb-2009-10-3-r25. URL: https://doi.org/10.1186/gb-2009-10-3-r25 (visited on 08/27/2025).

Langridge, J (1957). "Effect of day-length and gibberellic acid on the flowering of Arabidopsis." In: *Nature* 180.4575, pp. 36–37.

Laufs, Patrick et al. (Sept. 2004). "MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems." In: *Development* 131.17, pp. 4311–4322. ISSN: 0950-1991. DOI: 10.1242/dev.01320. URL: https://doi.org/10.1242/dev.01320 (visited on 02/21/2026).

Lauressergues, Dominique et al. (Apr. 2015). "Primary transcripts of microRNAs encode regulatory peptides." en. In: *Nature* 520.7545. Publisher: Nature Publishing Group, pp. 90–93. ISSN: 1476-4687. DOI: 10.1038/nature14346. URL: https://www.nature.com/articles/nature14346 (visited on 07/23/2025).

Lauter, Nick et al. (2005). "microRNA172 down-regulates glossy15 to promote vegetative phase change in maize." In: *Proceedings of the National Academy of Sciences* 102.26, pp. 9412–9417.

Li, Heng (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM." In: *arXiv preprint arXiv:1303.3997*.

Li, Hui et al. (Mar. 2017). "Small RNA Sequencing Reveals Differential miRNA Expression in the Early Development of Broccoli (Brassica oleracea var. italica) Pollen." English. In: *Frontiers in Plant Science* 8. Publisher: Frontiers. ISSN: 1664-462X. DOI: 10.3389/fpls.2017.00404. URL: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2017.00404/full (visited on 01/22/2026).

Li, Qinglian et al. (2021). "Evaluation and application of tools for the identification of known microRNAs in plants." en. In: *Applications in Plant Sciences* 9.3. _eprint: https://bsapubs.onlinelibrary.wiley.com/doi/pdf/10.1002/aps3.11414, e11414. ISSN: 2168-0450. DOI: 10.1002/aps3.11414. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/aps3.11414 (visited on 08/27/2025).

Liang, Xiao et al. (Dec. 2018). "Family-Wide Characterization of Histone Binding Abilities of PHD Domains of AL Proteins in Arabidopsis thaliana." en. In: *The Protein Journal* 37.6, pp. 531–538. ISSN: 1875-8355. DOI: 10.1007/s10930-018-9796-4. URL: https://doi.org/10.1007/s10930-018-9796-4 (visited on 09/30/2025).

Liebsch, Daniela and Javier F Palatnik (Feb. 2020). "MicroRNA miR396, GRF transcription factors and GIF co-regulators: a conserved plant growth regulatory module with potential for breeding and biotechnology." In: *Current Opinion in Plant Biology*. Growth and development 53, pp. 31–42. ISSN: 1369-5266. DOI: 10.1016/j.pbi.2019.09.008. URL: https://www.sciencedirect.com/science/article/pii/S1369526619300779 (visited on 09/30/2025).

Lim, Mi-Hye et al. (Mar. 2004). "A New Arabidopsis Gene, FLK, Encodes an RNA Binding Protein with K Homology Motifs and Regulates Flowering Time via FLOWERING LOCUS C [W]." In: *The Plant Cell* 16.3, pp. 731–740. ISSN: 1040-4651. DOI: 10.1105/tpc.019331. URL: https://doi.org/10.1105/tpc.019331 (visited on 09/25/2025).

Liu, Yong-Xin, Meng Wang, and Xiu-Jie Wang (Apr. 2014). "Endogenous Small RNA Clusters in Plants." In: *Genomics, Proteomics & Bioinformatics* 12.2, pp. 64–71. ISSN: 1672-0229. DOI: 10.1016/j.gpb.2014.04.003. URL: https://doi.org/10.1016/j.gpb.2014.04.003 (visited on 01/22/2026).

Llave, Cesar et al. (Sept. 2002). "Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA." In: *Science* 297.5589. Publisher: American Association for the Advancement of Science, pp. 2053–2056. DOI: 10.1126/science.1076311. URL: https://www.science.org/doi/full/10.1126/science.1076311 (visited on 09/30/2025).

Llave, César et al. (2011). "Target Validation of Plant microRNAs." en. In: *MicroRNAs in Development: Methods and Protocols*. Ed. by Tamas Dalmay. Totowa, NJ: Humana Press, pp. 187–208. ISBN: 978-1-61779-083-6. DOI: 10.1007/978-1-61779-083-6_14. URL: https://doi.org/10.1007/978-1-61779-083-6_14 (visited on 09/29/2025).

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." In: *Genome biology* 15.12, p. 550.

Lowe, Andrew et al. (May 2004). "Efficient large-scale development of microsatellites for marker and mapping applications in Brassica crop species." In: *TAG. Theoretical*

and applied genetics. Theoretische und angewandte Genetik 108, pp. 1103–12. DOI: 10.1007/s00122-003-1522-7.

Lu, Cheng et al. (Oct. 2006). "MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant." In: *Genome Research* 16.10, pp. 1276–1288. ISSN: 1088-9051. DOI: 10.1101/gr.5530106. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC1581437/ (visited on 02/20/2026).

Lu, Xiang et al. (Sept. 2022). "Winter warming post floral initiation delays flowering via bud dormancy activation and affects yield in a winter annual crop." In: *Proceedings of the National Academy of Sciences* 119.39. Publisher: Proceedings of the National Academy of Sciences, e2204355119. DOI: 10.1073/pnas.2204355119. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2204355119 (visited on 02/21/2026).

Macknight, Richard et al. (May 1997). "FCA, a Gene Controlling Flowering Time in Arabidopsis, Encodes a Protein Containing RNA-Binding Domains." English. In: *Cell* 89.5. Publisher: Elsevier, pp. 737–745. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/S0092-8674(00)80256-1. URL: https://www.cell.com/cell/abstract/S0092-8674(00)80256-1 (visited on 09/25/2025).

Maple, Robert et al. (2024). "Flowering time: from physiology, through genetics to mechanism." In: *Plant Physiology* 195.1, pp. 190–212.

Marquardt, S et al. (Oct. 2006). "Additional targets of the Arabidopsis autonomous pathway members, FCA and FY." In: *Journal of Experimental Botany* 57.13, pp. 3379–3386. ISSN: 0022-0957. DOI: 10.1093/jxb/erl073. URL: https://doi.org/10.1093/jxb/erl073 (visited on 09/25/2025).

Martin, Marcel (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." In: *EMBnet.journal* 17.1, pp. 10–12. ISSN: 2226-6089. DOI: 10.14806/ej.17.1.200. URL: https://journal.embnet.org/index.php/embnetjournal/article/view/200.

Mathieu, Johannes et al. (July 2009). "Repression of Flowering by the miR172 Target SMZ." en. In: *PLOS Biology* 7.7. Publisher: Public Library of Science, e1000148. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1000148. URL: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000148 (visited on 07/24/2025).

Mehraj, Hasan et al. (Mar. 2021). "Genome-wide analysis of long noncoding RNAs, 24-nt siRNAs, DNA methylation and H3K27me3 marks in Brassica rapa." en. In: *PLOS ONE* 16.3. Publisher: Public Library of Science, e0242530. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0242530. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242530 (visited on 09/28/2025).

Meier, Uwe et al. (Feb. 2009). "The BBCH system to coding the phenological growth stages of plants – history and publications –." en. In: *Journal of Cultivated Plants*. Artwork Size: 41-52 Pages Publisher: Journal of Cultivated Plants, 41–52 Pages. DOI: 10.5073/JFK.2009.02.01. URL: https://ojs.openagrar.de/index.php/Kulturpflanzenjournal/article/view/12142 (visited on 09/19/2025).

Michaels, Scott D and Richard M Amasino (1999). "FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering." In: *The Plant Cell* 11.5, pp. 949–956.

Molitor, Anne Marie et al. (Jan. 2014). "Arabidopsis AL PHD-PRC1 Complexes Promote Seed Germination through H3K4me3-to-H3K27me3 Chromatin State Switch in Repression of Seed Developmental Genes." In: *PLOS Genetics* 10.1. Publisher: Public Library of Science, pp. 1–12. DOI: 10.1371/journal.pgen.1004091. URL: https://doi.org/10.1371/journal.pgen.1004091.

Muralla, Rosanna et al. (June 2007). "Genetic Dissection of Histidine Biosynthesis in Arabidopsis." In: *Plant Physiology* 144.2, pp. 890–903. ISSN: 0032-0889. DOI: 10.1104/pp.107.096511. URL: https://doi.org/10.1104/pp.107.096511 (visited on 08/07/2025).

Nagaharu, UJJJB, NJJJB Nagaharu, et al. (1935). "Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization." In: *Jpn J Bot* 7.7, pp. 389–452.

Nguyen, Hai Ngoc et al. (2021). "Isopentenyltransferases as master regulators of crop performance: their function, manipulation, and genetic potential for stress adaptation and yield improvement." In: *Plant Biotechnology Journal* 19.7, pp. 1297–1313. DOI: https://doi.org/10.1111/pbi.13603. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/pbi.13603. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/pbi.13603.

Parry, G. et al. (Dec. 2009). "Complex regulation of the TIR1/AFB family of auxin receptors." In: *Proceedings of the National Academy of Sciences* 106.52. Publisher: Proceedings of the National Academy of Sciences, pp. 22540–22545. DOI: 10.1073/pnas.0911967106. URL: https://www.pnas.org/doi/full/10.1073/pnas.0911967106 (visited on 02/21/2026).

Payá-Milans, Miriam et al. (Dec. 2019). "Genome-wide analysis of the H3K27me3 epigenome and transcriptome in Brassica rapa." In: *GigaScience* 8.12, giz147. ISSN: 2047-217X. DOI: 10.1093/gigascience/giz147. URL: https://doi.org/10.1093/gigascience/giz147 (visited on 07/29/2025).

Picó, Sara et al. (Aug. 2015). "Deciphering the Role of POLYCOMB REPRESSIVE COMPLEX1 Variants in Regulating the Acquisition of Flowering Competence in Arabidopsis." In: *Plant Physiology* 168.4, pp. 1286–1297. ISSN: 0032-0889. DOI: 10.1104/pp.15.00073. URL: https://doi.org/10.1104/pp.15.00073 (visited on 07/29/2025).

Poethig, R. Scott and Jim Fouracre (Jan. 2024). "Temporal regulation of vegetative phase change in plants." English. In: *Developmental Cell* 59.1. Publisher: Elsevier, pp. 4–19. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2023.11.010. URL: https://www.cell.com/developmental-cell/abstract/S1534-5807(23)00608-1 (visited on 09/23/2025).

Pohlschröder, Mechthild et al. (Nov. 1997). "Protein Translocation in the Three Domains of Life: Variations on a Theme." English. In: *Cell* 91.5. Publisher: Elsevier, pp. 563–566. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/S0092-8674(00)80443-2. URL: https://www.cell.com/cell/abstract/S0092-8674(00)80443-2 (visited on 08/10/2025).

Putterill, Joanna et al. (1995). "The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors." In: *cell* 80.6, pp. 847–857.

Qin, Liuyu, Peng Xu, and Yuannian Jiao (Mar. 2025). "Evolution of Plant Conserved microRNAs After Whole-Genome Duplications." In: *Genome Biology and Evolution*

17.3, evaf045. ISSN: 1759-6653. DOI: 10.1093/gbe/evaf045. URL: https://doi.org/10.1093/gbe/evaf045 (visited on 09/21/2025).

Rajagopalan, Ramya et al. (Dec. 2006). "A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana." In: *Genes & Development* 20.24, pp. 3407–3425. ISSN: 0890-9369. DOI: 10.1101/gad.1476406. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC1698448/ (visited on 02/20/2026).

Ratcliffe, Oliver J. et al. (May 1998). "A common mechanism controls the life cycle and architecture of plants." In: *Development* 125.9, pp. 1609–1615. ISSN: 0950-1991. DOI: 10.1242/dev.125.9.1609. URL: https://doi.org/10.1242/dev.125.9.1609 (visited on 08/05/2025).

Ritchie, Matthew E et al. (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." In: *Nucleic acids research* 43.7, e47–e47.

"RNA Quality Assessment with the Agilent Automated Electrophoresis Systems" (n.d.). en. In: ().

Schmid, Markus et al. (Dec. 2003). "Dissection of floral induction pathways using global expression analysis." In: *Development* 130.24, pp. 6001–6012. ISSN: 0950-1991. DOI: 10.1242/dev.00842. URL: https://doi.org/10.1242/dev.00842 (visited on 07/24/2025).

Schomburg, Fritz M. et al. (June 2001). "FPA, a Gene Involved in Floral Induction in Arabidopsis, Encodes a Protein Containing RNA-Recognition Motifs." In: *The Plant Cell* 13.6, pp. 1427–1436. ISSN: 1040-4651. DOI: 10.1105/TPC.010017. URL: https://doi.org/10.1105/TPC.010017 (visited on 09/25/2025).

Schwartz, Brian W., Edward C. Yeung, and David W. Meinke (Nov. 1994). "Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of Arabidopsis." In: *Development* 120.11, pp. 3235–3245. ISSN: 0950-1991. DOI: 10.1242/dev.120.11.3235. URL: https://doi.org/10.1242/dev.120.11.3235 (visited on 08/11/2025).

Searle, Iain et al. (2006). "The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis." In: *Genes & development* 20.7, pp. 898–912.

Shumate, Alaina et al. (2022). "Improved transcriptome assembly using a hybrid of long and short reads with StringTie." In: *PLoS computational biology* 18.6, e1009730.

Siriwardana, Chamindika L. et al. (Dec. 2016). "NUCLEAR FACTOR Y, Subunit A (NF-YA) Proteins Positively Regulate Flowering and Act Through FLOWERING LOCUS T." en. In: *PLOS Genetics* 12.12. Publisher: Public Library of Science, e1006496. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1006496. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006496 (visited on 08/07/2025).

Smid, Marcel et al. (2018). "Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons." In: *BMC bioinformatics* 19.1, p. 236.

Song, Chieun et al. (Apr. 2022). "Overexpression of DREB2C Delays Flowering in Arabidopsis thaliana via the Activation of FLC." en. In: *Journal of Plant Biology* 65.2, pp. 133–143. ISSN: 1867-0725. DOI: 10.1007/s12374-021-09337-y. URL: https://doi.org/10.1007/s12374-021-09337-y (visited on 09/17/2025).

Sorin, Céline et al. (2014). "A miR169 isoform regulates specific NF-YA targets and root architecture in Arabidopsis." en. In: *New Phytologist* 202.4. _eprint: https://nph.onlinelibrary.wiley.com/doi/

pp. 1197–1211. ISSN: 1469-8137. DOI: 10.1111/nph.12735. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.12735 (visited on 02/21/2026).

Spies, Daniel et al. (Oct. 2017). "Comparative analysis of differential gene expression tools for RNA sequencing time course data." In: *Briefings in Bioinformatics* 20.1, pp. 288–298. ISSN: 1467-5463. DOI: 10.1093/bib/bbx115. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC6357553/ (visited on 09/17/2025).

Stephenson, Pauline et al. (Apr. 2010). "A rich TILLING resource for studying gene function in Brassica rapa." In: *BMC Plant Biology* 10.1, p. 62. ISSN: 1471-2229. DOI: 10.1186/1471-2229-10-62. URL: https://doi.org/10.1186/1471-2229-10-62 (visited on 09/24/2025).

Suárez-López, Paula et al. (2001). "CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis." In: *Nature* 410.6832, pp. 1116–1120.

Sultana, Fahmida, Jamilur Rahman, and Md Mahmudul Hasan (2020). "Genetic variability and character association of Bangladeshi popular varieties of mustard (*Brassica rapa L.*)" en. In: *Journal of Bangladesh Academy of Sciences* 44.2, pp. 95–107. ISSN: 2224-7270. DOI: 10.3329/jbas.v44i2.51454. URL: https://www.banglajol.info/index.php/JBAS/article/view/51454 (visited on 02/14/2026).

Sun, Chao et al. (Nov. 2015). "Impacts of Whole-Genome Triplication on MIRNA Evolution in Brassica rapa." In: *Genome Biology and Evolution* 7.11, pp. 3085–3096. ISSN: 1759-6653. DOI: 10.1093/gbe/evv206. URL: https://doi.org/10.1093/gbe/evv206 (visited on 09/21/2025).

Sung, Z. R. et al. (Dec. 1992). "EMF, an Arabidopsis Gene Required for Vegetative Shoot Development." In: *Science* 258.5088. Publisher: American Association for the Advancement of Science, pp. 1645–1647. DOI: 10.1126/science.258.5088.1645. URL: https://www.science.org/doi/10.1126/science.258.5088.1645 (visited on 08/11/2025).

Takahashi, Seiji et al. (July 2000). "An Arabidopsis Gene Encoding a Ca2+-Binding Protein is Induced by Abscisic Acid during Dehydration." In: *Plant and Cell Physiology* 41.7, pp. 898–903. ISSN: 0032-0781. DOI: 10.1093/pcp/pcd010. URL: https://doi.org/10.1093/pcp/pcd010 (visited on 08/11/2025).

Teotia, Sachin and Guiliang Tang (Mar. 2015). "To Bloom or Not to Bloom: Role of MicroRNAs in Plant Flowering." English. In: *Molecular Plant* 8.3. Publisher: Elsevier, pp. 359–377. ISSN: 1674-2052. DOI: 10.1016/j.molp.2014.12.018. URL: https://www.cell.com/molecular-plant/abstract/S1674-2052(14)00082-3 (visited on 07/23/2025).

Trindade, Inês et al. (Feb. 2010). "miR398 and miR408 are up-regulated in response to water deficit in Medicago truncatula." en. In: *Planta* 231.3, pp. 705–716. ISSN: 1432-2048. DOI: 10.1007/s00425-009-1078-0. URL: https://doi.org/10.1007/s00425-009-1078-0 (visited on 09/17/2025).

Waheed, Saquib and Lihui Zeng (Mar. 2020). "The Critical Role of miRNAs in Regulation of Flowering Time and Flower Development." In: *Genes* 11.3, p. 319. ISSN: 2073-4425. DOI: 10.3390/genes11030319. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7140873/ (visited on 09/19/2025).

Wang, Guan-Feng et al. (July 2011). "Multiple Roles of WIN3 in Regulating Disease Resistance, Cell Death, and Flowering Time in Arabidopsis." In: *Plant Physiology*

156.3, pp. 1508–1519. ISSN: 0032-0889. DOI: 10.1104/pp.111.176776. URL: https://doi.org/10.1104/pp.111.176776 (visited on 08/10/2025).

Wang, Jia-Wei (Sept. 2014). "Regulation of flowering time by the miR156-mediated age pathway." In: *Journal of Experimental Botany* 65.17, pp. 4723–4730. ISSN: 0022-0957. DOI: 10.1093/jxb/eru246. URL: https://doi.org/10.1093/jxb/eru246 (visited on 07/23/2025).

Wang, Jia-Wei, Benjamin Czech, and Detlef Weigel (Aug. 2009). "miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis thaliana." English. In: *Cell* 138.4. Publisher: Elsevier, pp. 738–749. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2009.06.014. URL: https://www.cell.com/cell/abstract/S0092-8674(09)00713-2 (visited on 07/23/2025).

Wang, Jia-Wei, Mee Yeon Park, et al. (Feb. 2011). "MiRNA Control of Vegetative Phase Change in Trees." en. In: *PLOS Genetics* 7.2. Publisher: Public Library of Science, e1002012. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1002012. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002012 (visited on 07/23/2025).

Wenkel, Stephan et al. (Nov. 2006). "CONSTANS and the CCAAT Box Binding Complex Share a Functionally Important Domain and Interact to Regulate Flowering of Arabidopsis." In: *The Plant Cell* 18.11, pp. 2971–2984. ISSN: 1040-4651. DOI: 10.1105/tpc.106.043299. URL: https://doi.org/10.1105/tpc.106.043299 (visited on 08/07/2025).

Wickham, Hadley (2011). "ggplot2." In: *Wiley interdisciplinary reviews: computational statistics* 3.2, pp. 180–185.

Wickland, Daniel P. and Yoshie Hanzawa (July 2015). "The *FLOWERING LOCUS T/TERMINAL FLOWER 1* Gene Family: Functional Evolution and Molecular Mechanisms." In: *Molecular Plant* 8.7, pp. 983–997. ISSN: 1674-2052. DOI: 10.1016/j.molp.2015.01.007. URL: https://www.sciencedirect.com/science/article/pii/S1674205215000945 (visited on 08/05/2025).

Williams, Leor et al. (Aug. 2005). "Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes." In: *Development* 132.16, pp. 3657–3668. ISSN: 0950-1991. DOI: 10.1242/dev.01942. URL: https://doi.org/10.1242/dev.01942 (visited on 02/21/2026).

Wilson, Ruth N., John W. Heckman, and Chris R. Somerville (Sept. 1992). "Gibberellin Is Required for Flowering in Arabidopsis thaliana under Short Days 1." In: *Plant Physiology* 100.1, pp. 403–408. ISSN: 0032-0889. DOI: 10.1104/pp.100.1.403. URL: https://doi.org/10.1104/pp.100.1.403 (visited on 09/25/2025).

Winkler, Malcolm E. and Smirla Ramos-Montañez (Sept. 2009). "Biosynthesis of Histidine." In: *EcoSal Plus* 3.2. Publisher: American Society for Microbiology, 10.1128/ecosalplus.3.6.1.9. DOI: 10.1128/ecosalplus.3.6.1.9. URL: https://journals.asm.org/doi/full/10.1128/ecosalplus.3.6.1.9 (visited on 08/07/2025).

Woodhouse, Shannon et al. (July 2021). "Validation of a novel associative transcriptomics pipeline in Brassica oleracea: identifying candidates for vernalisation response." In: *BMC Genomics* 22.1, p. 539. ISSN: 1471-2164. DOI: 10.1186/s12864-021-07805-w. URL: https://doi.org/10.1186/s12864-021-07805-w.

Wu, Gang, Mee Yeon Park, et al. (Aug. 2009). "The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis." English. In: *Cell* 138.4.

Publisher: Elsevier, pp. 750–759. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2009.06.031. URL: https://www.cell.com/cell/abstract/S0092-8674(09)00778-8 (visited on 07/23/2025).

Wu, Gang and R. Scott Poethig (Sept. 2006). "Temporal regulation of shoot development in Arabidopsis thalianaby miR156 and its target SPL3." In: *Development* 133.18, pp. 3539–3547. ISSN: 0950-1991. DOI: 10.1242/dev.02521. URL: https://doi.org/10.1242/dev.02521 (visited on 07/23/2025).

Wu, Miin-Feng, Qing Tian, and Jason W. Reed (Nov. 2006). "Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction." In: *Development* 133.21, pp. 4211–4218. ISSN: 0950-1991. DOI: 10.1242/dev.02602. URL: https://doi.org/10.1242/dev.02602 (visited on 02/24/2026).

Wu, Ying et al. (Aug. 2022). "Genome-wide analysis of the DREB family genes and functional identification of the involvement of BrDREB2B in abiotic stress in wucai (Brassica campestris L.)" en. In: *BMC Genomics* 23.1, p. 598. ISSN: 1471-2164. DOI: 10.1186/s12864-022-08812-1. URL: https://doi.org/10.1186/s12864-022-08812-1 (visited on 09/17/2025).

Xie, Kabin, Congqing Wu, and Lizhong Xiong (2006). "Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice." In: *Plant physiology* 142.1, pp. 280–293.

Xing, Shuping et al. (Dec. 2010). "miR156-Targeted and Nontargeted SBP-Box Transcription Factors Act in Concert to Secure Male Fertility in Arabidopsis." In: *The Plant Cell* 22.12, pp. 3935–3950. ISSN: 1040-4651. DOI: 10.1105/tpc.110.079343. URL: https://doi.org/10.1105/tpc.110.079343 (visited on 09/21/2025).

Xu, Jingya et al. (2021). "VDAC1 Negatively Regulates Floral Transition in Arabidopsis thaliana." In: *International Journal of Molecular Sciences* 22.21. ISSN: 1422-0067. DOI: 10.3390/ijms222111603. URL: https://www.mdpi.com/1422-0067/22/21/11603.

Xu, Miao Yun et al. (Dec. 2013). "Stress-induced early flowering is mediated by miR169 in Arabidopsis thaliana." In: *Journal of Experimental Botany* 65.1, pp. 89–101. ISSN: 0022-0957. DOI: 10.1093/jxb/ert353. URL: https://doi.org/10.1093/jxb/ert353.

Xu, Mingli, Tieqiang Hu, Michael R. Smith, et al. (Jan. 2016). "Epigenetic Regulation of Vegetative Phase Change in Arabidopsis." In: *The Plant Cell* 28.1, pp. 28–41. ISSN: 1040-4651. DOI: 10.1105/tpc.15.00854. URL: https://doi.org/10.1105/tpc.15.00854 (visited on 07/29/2025).

Xu, Mingli, Tieqiang Hu, Jianfei Zhao, et al. (Aug. 2016). "Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana." en. In: *PLOS Genetics* 12.8. Publisher: Public Library of Science, e1006263. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1006263. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006263 (visited on 07/23/2025).

Xu, Mingli, Aaron R. Leichty, et al. (Jan. 2018). "H2A.Z promotes the transcription of MIR156A and MIR156C in Arabidopsis by facilitating the deposition of H3K4me3." In: *Development* 145.2, dev152868. ISSN: 0950-1991. DOI: 10.1242/dev.152868. URL: https://doi.org/10.1242/dev.152868 (visited on 07/29/2025).

Xu, Yunmin et al. (Dec. 2016). "Regulation of Vegetative Phase Change by SWI2/SNF2 Chromatin Remodeling ATPase BRAHMA." In: *Plant Physiology* 172.4, pp. 2416–2428. ISSN: 0032-0889. DOI: 10.1104/pp.16.01588. URL: https://doi.org/10.1104/pp.16.01588 (visited on 07/29/2025).

Yang, Li et al. (Mar. 2013). "Sugar promotes vegetative phase change in Arabidopsis thaliana by repressing the expression of MIR156A and MIR156C." In: *eLife* 2. Ed. by Richard Amasino. Publisher: eLife Sciences Publications, Ltd, e00260. ISSN: 2050-084X. DOI: 10.7554/eLife.00260. URL: https://doi.org/10.7554/eLife.00260 (visited on 07/28/2025).

Yu, Sha et al. (Mar. 2013). "Sugar is an endogenous cue for juvenile-to-adult phase transition in plants." In: *eLife* 2. Ed. by Richard Amasino. Publisher: eLife Sciences Publications, Ltd, e00269. ISSN: 2050-084X. DOI: 10.7554/eLife.00269. URL: https://doi.org/10.7554/eLife.00269 (visited on 07/28/2025).

Zhang, Lei et al. (Jan. 2018). "Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies." In: *Horticulture Research* 5, p. 50. ISSN: 2052-7276. DOI: 10.1038/s41438-018-0071-9. URL: https://doi.org/10.1038/s41438-018-0071-9 (visited on 09/29/2025).

Zhang, Shuxin, Yuhui Liu, and Bin Yu (2015). "New insights into pri-miRNA processing and accumulation in plants." en. In: *WIREs RNA* 6.5, pp. 533–545. ISSN: 1757-7012. DOI: 10.1002/wrna.1292. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1292 (visited on 07/23/2025).

Zhao, Jianfei, Erin Doody, and R. Scott Poethig (Feb. 2023). "Reproductive competence is regulated independently of vegetative phase change in Arabidopsis thaliana." English. In: *Current Biology* 33.3. Publisher: Elsevier, 487–497.e2. ISSN: 0960-9822. DOI: 10.1016/j.cub.2022.12.029. URL: https://www.cell.com/current-biology/abstract/S0960-9822(22)01966-2 (visited on 09/21/2025).

Zhou, Chuan-Miao et al. (May 2013). "Molecular Basis of Age-Dependent Vernalization in Cardamine flexuosa." In: *Science* 340.6136. Publisher: American Association for the Advancement of Science, pp. 1097–1100. DOI: 10.1126/science.1234340. URL: https://www.science.org/doi/full/10.1126/science.1234340 (visited on 07/24/2025).